Understanding Data, One Visualization at a Time:

Forms, Tools, and Skills for Big-Data Visual Exploration and Communication

Benjamin Bach, PhD
Habilitation à diriger des recherches (HDR)
Université de Bordeaux

Presented on December, 13th 2024

Jury

- Prof. Miriah Meyer, *Linköping University, Sweden* (Rapporteur)
- Dr. Vidya Setlur, *Tableau Research, USA* (Rapporteur)
- Prof. Niklas Elmqvist, *Aarhus University, Denmark* (Rapporteur)
- Dr. Petra Isenberg, *Inria, France* (Examinateur)
- Prof. Min Chen, *Oxford University, UK* (Examinateur)
- Prof. David Auber, *University of Bordeaux, France* (Examinateur)

Abstract

This thesis presents novel forms and tools to communicating data and creating visualizations. With the proliferation of data-driven practices, data visualizations are essential for the analysis of data as well as the communication to diverse stakeholders. Yet, the design of visualization and communication of data are made challenging by the complexity of the real-world phenomena represented by the data, the datasets themselves, and the expertise of the analysts and communicators. Hence, in this thesis, I seek answers to the questions *How to support people present and explain complex data?* and *How to help them create and use visualization effectively?*. I summarize some of my research addressing three respective key challenges within this context: (i) the creation and study of **forms for data-driven storytelling and communication** (data comics, dashboards, visualization atlases, visualization articles, dataGIFs); (ii) the **design and implementation of tools to design, create and use visualization** (authoring interactive data comics, semi-automatic data tours, a network visualization grammar); and (iii) the **education in skills** related to visualization literacy and design (challenges experienced in exploring networks, challenges in visualization education, automatic explanation of visual patterns). I conclude with a discussion of future research directions around novel forms of storytelling, visualization authoring and design, and visualization education.

Publications and supervisions

Publications

The following selection of my publications from the past 7 years are, in strongly abbreviated form, included this dissertation. The complete list of my publications can be found on my Google Scholar profile: `https://scholar.google.co.id/citations?user=dXbz4FgAAAAJ&hl=en`.

Chapter 2: Forms for communication and presentation

1. **Bach, B.**, Kerracher, N., Hall, K.W., Carpendale, S., Kennedy, J. and Henry Riche, N., 2016, May. *Telling stories about dynamic networks with graph comics.* In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 3670-3682).
2. **Bach, B.**, Wang, Z., Farinella, M., Murray-Rust, D. and Henry Riche, N., 2018, April. *Design patterns for data comics.* In Proceedings of the 2018 chi conference on human factors in computing systems (pp. 1-12).
3. Wang, Z., Wang, S., Farinella, M., Murray-Rust, D., Henry Riche, N. and **Bach, B.**, 2019, May. *Comparing effectiveness and engagement of data comics and infographics*. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-12).
4. **Bach, B.**, Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y. and Chen, M., 2022. *Dashboard design patterns*. IEEE Transactions on Visualization and Computer Graphics, 29(1), pp.342-352.
5. Wang J., **Bach B.**, Hinrichs U., 2024, *Visualization Atlases: Explaining and Exploring Complex Topics through Data, Visualization, and Narration*. IEEE Transactions on Visaulizations and Computer Graphics, to appear.

Chapter 3: Tools for data-driven storytelling and visualization design

1. Wang, Z., Romat, H., Chevalier, F., Riche, N.H., Murray-Rust, D. and **Bach, B.**, 2021. *Interactive data comics*. IEEE Transactions on Visualization and Computer Graphics, 28(1), pp.944-954.
2. Li, W., Schöttler, S., Scott-Brown, J., Wang, Y., Chen, S., Qu, H. and **Bach, B.**, 2023, April. *Networknarratives: Data tours for visual network exploration and analysis*. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-15).
3. Scott-Brown, J., Pister A., and **Bach B.**, 2024, *NetworkPanorama: A Declarative Grammar for Network Construction, Transformation, and Interactive Visualization*, Arxiv.

Chapter 4: Skills for visualization, data-driven storytelling and visualization design.

1. AlKadi, M., Serrano, V., Scott-Brown, J., Plaisant, C., Fekete, J.D., Hinrichs, U. and **Bach, B.**, 2022. *Understanding barriers to network exploration with visualization: A report from the trenches*. IEEE Transactions on Visualization and Computer Graphics, 29(1), pp.907-917.
2. **Bach, B.**, Keck, M., Rajabiyazdi, F., Losev, T., Meirelles, I., Dykes, J., Laramee, R.S., AlKadi, M., Stoiber, C., Huron, S. and Perin, C., 2023. *Challenges and opportunities in data visualization education: A call to action*. IEEE Transactions on visualization and computer graphics.
3. Shu, X., Pister, A., Tang, J., Chevalier, F. and **Bach, B.**, 2024. *Does This Have a Particular Meaning? Interactive Pattern Explanation for Network Visualizations*. arXiv preprint arXiv:2408.01272.

People

Over the past 7 years, since joining the University of Edinburgh as a faculty member in 2017, I have supervised the following postdocs, engineers, and PhD students as part of my lab. Many of the publications in this dissertation are done in collaboration with them.

Postdocs and Engineers:

1. James-Scott Brown (postdoc)
2. Xinhuan Shu (postdoc)
3. Alexis Pister (engineer)
4. Tomas Vancisin (engineer)
5. Phillip Heslop (postdoc)
6. Devanjan Bhattacharya (postdoc)
7. Aaran Ridley (postdoc)
8. Aba-Sha Dadzie (postdoc / teacher)
9. Iain Carson (teacher)

PhD students

1. Zezhong Wang (graduated 2020, main supervisor)
2. Kun Ting Chen (graduated 2022, assistant supervisor)
3. Ryan Bowler (graduated 2022, main supervisor)
4. Lucy Havens (graduated 2023, assistant supervisor))
5. Mashael AlKadi (graduated 2024, main supervisor)
6. Tobias Kauer (graduaging 2024, main supervisor)
7. Sarah Schöttler (graduating 2024, main supervisor)
8. Sarah Dunn (main supervisor)
9. Jinrui Wang (assistant supervisor)
10. Madgalena Boucher (assistant supervisor)
11. Rea Michalopoulou (assistant supervisor)

Thanks

This thesis is the results of 10 years of research in data visualization, data-driven storytelling, human-computer interaction and visualization design. Much of the research is motivated by hands-on experience from designing visualizations, building visualization tools, teaching visualization, and discussions and observations from users, scientists and artists. Most of this research involved the work of master and doctoral students, postdocs and research engineers, as well as colleagues. To all of them, I owe my un-expressible thank for their time, patience, critical feedback, and co-ideation.

As always, I am deeply grateful to **Jean-Daniel Fekete**, **Nathalie Henry Riche**, **Emmanuel Pietriga**, **Tim Dwyer** and **Hanspeter Pister** who directly mentored me throughout my academic trajectory. They are the best mentors I could ever have, full of passion for their work, sharp pragmatism, and creative love for data visualization and human-computer interaction.

A lot of the work in this thesis is done with and through my postdocs and doctoral students at the University of Edinburgh, who are such a talented and inspiring team: **Zezhong Wang**, who took on the task to explore and popularize data comics; **Tobias Kauer**, who is making visualizations discursive; **Mashael AlKadi**, who managed to understand the barriers network analysts are facing and build her own interventions; **Kun-Ting Wang**, how is drawing visualizations on all sorts of topologies such as torus and shperes, **Ryan Bowler**, how is making dealing with uncertainty in scheduleing possible; **Sarah Schöttler**, who is streamlining and automating responsiveness of map visualizations; **Jinrui Wang**, who is exploring and building visualization atlases; **Magdalena Boucher**, who explores (data) comics for visualization education; **Xinhuan Shu**, who teaches visualization patterns; **James Scott-Brown**, who creates interactive network visualization.

My great thanks also go to dear colleague and friend **Uta Hinrichs** who is co-creating our VisHub research group at the University of Edinburgh for her thoughtful and critical input into our collaborations and for co-organizing the Information+ conference in Edinburgh in 2023. Likewise, Alexis Pister, in his role as data visualization engineer has become an invaluable resource of advise, ideas, and support for the whole VisHub research group and wider university. I want to thank, in particular, **Prof. Min Chen** for being such a kind mentor right from the start of me arriving in the UK without knowing many people there. Min has involved me into his project to help visualization during the Corona pandemic and which lead to many follow-on works.

I also want to thank my HDR committee jury: Prof. Miriah Meyer, Dr. Vidya Setlur, Prof. Niklas Elmqvist, Dr. Petra Isenberg, Prof. Min Chen, and Prof. David Auber.

Eventually, I want to list many of those I had the pleasure to work with during these past 10 years and who have inspired me, participated in discussions, helped me think through ideas, make difficult decisions, and become the researcher I am, in no particular order: Wolfgang Aigner, Marian Dörk, Fanny Chevalier, Isabell Meireilles, Mandy Keck, Huamin Qu and all the others who inspired my work and thinking; Sheelagh Carpendale, Tamara Munzner, Pierre Dragicevic, Yvonne Jansen, and Jessica Hullman.

How to read this document?

This habilitation, to which I refer in its more generic form *thesis*, is a compilation of 11 papers, representing the main problems and contributions of my past work within the story of this thesis. While all papers are collaborative projects, I selected papers that either *a*) I first-authored myself, or *b*) were first-authored by a student, research associate, or visiting researcher whom I directly supervised as main supervisor. The introduction to each chapter gives pointers to my other work in the respective area not included explicitly in this thesis.

Each paper is found in abbreviated form within this thesis, as one section or subsection. A the beginning of each section, the full citation of the original paper is cited on the page margin. The abbreviated versions in this thesis focus on motivation, contribution, examples of the work done, its main findings and its discussion. Technical and methodological details are mostly omitted for brevity. Where appropriate, I included further details and explanations within the context of this thesis. At the end of each section, a gray box summarizes the main findings and discussion points from each paper. The reader familiar with a particular paper may refer to these summaries to refresh their mind or skip to the chapter discussion instead. The reader unfamilar with a particular work is invited to explore the details. I used bullet-point lists wherever possible to facilitate overview and keep the sections short.

All papers in this thesis, except for Section 3.3 (*NetPanorama: A Declarative Grammar for Network Construction, Transformation, and Interactive Visualization*), have already been published in peer-reviewed journals and conferences—either the IEEE VIS conference or the ACM CHI conference. Introductions and discussions to each chapter, as well as the thesis introduction (Chapter 1) are written for the purpose of this thesis. No Large Language Models were involved in the writing and editing of this thesis. Where I refer to specifically personal motives and thoughts, I used the form "I", otherwise I use the collective "we".

Contents

Chapter 1

Introduction

1.1 Context

Wicked problems—as opposed to "tame" problems—are problems that lack clear problem definitions and that have no silver-bullet solution [Rittel and Webber, 1973]. Instead, they involve a large range of options, decisions, and people with different opinions and priorities. For these very reasons, many of today's global problems are wicked problems; the creation of economic equality and prosperity; solving the environmental crisis [Ludwig, 2001, Levin et al., 2012]; deciding on public spending and planning [Yarnes, 2013]; solving humanities' violent conflicts; providing individual healthcare, manage the application and development of artificial intelligence, and many more. Wicked problems comprise complex dynamic and non-linear systems that are only partially understood by modern science, such as the human body, the brain, the economy, the climate, black-box artificial intelligence. Moreover, over time, such systems and problems change due to internal of external expected or unexpected events. As research better understands these systems, the understanding of the problems change, too. This results in refined problem definitions that can require new approaches and solutions making the entire wicked problem a highly-dynamic complex long-term endeavor with implications on the health, well-being, and future of large portions of the population and those to come [MacAskill, 2022].

Addressing wicked problems is part of the problem itself. Often we are lacking appropriate methodologies to their solution as existing ways of thinking can partially be to blame for the problem in the first place [Ludwig, 2001]). Or, solutions can lead to rebound effects if the problem or approaches are not fully understood. For example, international intervention in armed conflicts can escalate, medication can lead to side effects, and climate mitigation to solve one problem (e.g., efficient electric batteries) can cause problems elsewhere to cope with the resulting increased demand (e.g., mining the respective minerals, switching back to non-renewable energy resources). For those reasons, addressing wicked problems requires a range of ingredients. They require a wide range of orchestrated approaches to work hand-in hand. These approaches must be overseen, resources allocated, stakeholders coordinated, decisions being made. They also require creative thinking and the experimentation with new ideas; they require collaboration between experts from many domains. They also require buy-in from large fractions of a populations to define the problem, seek solutions, monitor progress, and follow instructions. In this context, local knowledge is as crucial as expert knowledge as it involves those directly affected by the problem and who can provide both first-hand experiences of the problem, ideas for solutions as well as play an important role in mitigation. Solutions to wicked problems hence needs to be carefully planned, managed, and debated and in order to do so. Hence, while wicked problems *"cannot be separated from issues of values, equity, and social justice"* [Ludwig, 2001] every step of the process—from defining the initial problem, to making decisions, monitoring progress, and adapting to evolving problems and approaches—needs evidence in the form of *data*.

Data about wicked problems

Data and analysis are especially important in contexts where (multiple) opinions are involved, where situations evolve and that require re-evaluation and tracking of progress to adjust decisions. Data can serve as the basis for understanding the complex systems in charge (human bodies, the climate, the economy, etc.) and help simulate potential decisions upfront; it can refer to evidence in collected measures, inform model parameters, and track success measures (key performance indicators).

Data about wicked problems can come from many sources including real-world data collection as well as generation through computational models (e.g. [Sudmant et al., 2024]). In fact, a growing number of initiatives collect comprehensive data about wicked problems with the focus on presenting and making them accessible to diverse audiences. For example, the International Panel for Climate Change (IPCC) regularly summarizes data and scientific evidence in special reports, addressing different topics an audiences [Pörtner et al., 2022]. Another example is the *Armed Conflict and Location & Event Data* (*ACLED*), an independent non-profit organization collecting data on violent conflict and protest in all countries and territories in the world.[1] The data are used by researchers, peace builders, and policians for information and decision making. Likewise, *Our World in Data*[2] compiles data sets on global demographics and health. It's website reports 'close to 100 million' unique visitors for 2023[3] and it is cited as data source across numerous news media.

What's interesting is that these respective data sets easily share many of the characteristics of big data such as the canonical Five V's: *velocity, value, variety, and veracity*. For example, ACLED lists 348,424 violent events for the past year alone (Aug 2023-July2024). The organization works with numerous volunteers around the globe as well as major organizations such as the UN to update the data daily and provide background analysis. Likewise, the online *IPCC interactive atlas*[4] contains high-resolution global geographic data data from 15 climate models with current and historic predictions, 11 current observational datasets, for 28 measures (such as temperature, snowfall, etc.), across 4 time periods, 4 possible future scenarios, and several seasons. Eventually, for each country the *Atlas of Economic Complexity* lists 5,000 goods across 10 categories, ranging from sometimes back to 1962 - 2021, updated on an annual base. Data in these contexts necessarily contains uncertainty and is of questionable quality (veracity) as it comes from different sources, at different time-points, and through different methods. In summary, these these initiatives are comparable to open data and government repositories (e.g., `https://data.gov`, `https://www.data.gov.uk`, `https://data.europa.eu`) in that they provide data. However, they differ in that these portals are *i*) created specifically around a wicked problem (a topic, rather than providing 'random' data about a country), with *ii*) the mission to making data about that topic accessible to a wider audiences. Most recently, we have seen a range of such platforms collecting data about the Covid-19

[1] https://acleddata.com/
[2] `https://ourworldindata.org`
[3] `https://ourworldindata.org/coverage`
[4] `https://interactive-atlas.ipcc.ch`

pandemic and providing the data to the public, policy makers, and analysts in the from of dashboards and data stories [Zhang et al., 2021].

'The Nine E's in Big-Data Visual Storytelling'

Given this situation—large heterogeneous and dynamic data sets about wicked problems with the mission to communication and accessability—it is helpful to think how such *Big Data Storytelling* is different from existing data-driven storytelling practices (e.g., [Riche et al., 2018, Cairo, 2012, Segel and Heer, 2010]) and what specific challenges it entails.

For that sake, let me discuss some of these challenges as the "Nine E's of Big-Data Visual Storytelling".

1/ Extensiveness—Compared to small-data storytelling, data sets are too large, they are too big to be explained in a single data article or dashboard. They require an ensemble of stories, linked, and embedded into a common context. The size of the data makes it hard to provide enough explanations in time, at least on a human scale. Explanation and understanding large data sets and complex topics will require an audience time and motivation.

2/ Explanation—Explanations need to explain more than key insights and findings. They need to explain methodologies, terminology, data structures and forms, computational models if used, analysis procedures and other decisions involved in the data preparation process. Explanations need to deliver a wider context and cater to the specific needs, expertise and expectations (see below) of their audience.

3/ Exploration—While also linked to size is the problem that stories alone cannot account for the dynamicness and vastness of the data alone as well as the different decisions and tasks people are required to perform. Hence, exploration is a key concept in big data storytelling, i.e., the interactive, open or goal-driven exploration and analysis of the data sets. This requires people being motivated to explore, able to ask questions and able to understand the respective user interfaces.

4/ Evolution—The dynamic nature of the problem and data requires updated data sets and stories. Updates can lead to inconsistencies among stories and data but also to challenges managing data updates in the visualizations and stories or tracking changes over time. For example, instead of updating their online visual atlas, the UN decided to create a new online atlas on their sustainable development goals, every so and so many years.

5/ Expertise+ 6/Expectation—Audiences will greatly vary in their expertise as well as their expectations. For example, governments need access to health data, climate modeling and conflict data for policy and law making. Businesses rely on such data to make investment decisions. Educators and students can base their learning on evidence from data and interacting with those data sets in their classes. Researchers perform further analysis and integrate data from different sources. Eventually, the general public can benefit from knowing about the respective data sets and decision making processes, or, use them to inform their own opinions, world views, and decisions. Like in traditional data-driven storytelling, these audiences are not necessarily domain scientist in the respective topic. However, these audiences are diverse in their goals and knowledge which requires interfaces and stories to be catered to more than one specific audience. Yet, most of the works published so far focuses on rather unspecified generic audiences such as 'general public'.

7/ Exceptions—Presenting large and heterogenous data implies numerous exceptions with respect to data completeness, their analysis process(es), and their applications. Some data sets are outdated, others are incomplete or uncertain. Methods for data collection, analysis, and reproduction might be unreliable and suffer from forking paths analyses [Gelman and Loken, 2013]. Any presentation and engagement needs mechanisms to be transparent about exceptions and provide the respective context and potentially warnings.

8/ Embargoes—Often, there are constraints on publishing and communicating data. For example, where this jeopardizes people's security or to prevent premature decisions. Some parts of the data might need to be treated confidentially and an entire openness and explorability needs to be carefully curated.

9/ Education—Data storytelling and communication in the contexts of these previous eight E's is more than explanation, presentation and communication. It is education in the widest sense, around concepts in data literacy, visualization literacy and the specific wicked problem. Big data storytelling can also play a role in general education as a means to foster these skills as part of training experts in specific wicked problems in data-driven and empirical methods.

The role of data visualization

These nine E's represent some loose reflection on the current challenges in data-driven storytelling. As evident from my professional background, this thesis is concerned with visual and interactive approaches to address the problem of how to support communication and presentation of big data in the context of wicked problems. **Data visualization**, i.e., the creation and use of external visual representations of structures and quantities in data for decision making and cognition [Ware, 2019, Munzner, 2014], is uniquely placed to deal with the data and context of wicked problems because of visualizations' ability to become an interface to the data and analysis. Visualizations can show complex patterns and relations in data in an accessible way. Abstract visual constellations, patterns, relationships become assembled into higher level meaningful structures and concepts [Boy et al., 2014a, Pinker, 1990] can create the mental models we need to (literally, mentally) visualize a problem and seek for understanding and solutions. Visualizations can provide overview, visualize complexity in data, deal with uncertainty, show different views on the same data, make information accessible to many people, and support the interactive exploration, the presentation of results, collaboration, and decision making in many ways.

While the superpowers of visualization have been known for a long time [Meirelles, 2013] and have routinely been used in traditional statistics [Tukey, 1977], map-making [Rendgen, 2018], and science [Lima, 2017, Lima, 2014, Rosenberg and Grafton, 2013], the past decades have also brought numerous innovations to visualization in the form of interactive interfaces, immersive display and interaction technology [Dwyer et al., 2018], the internet, better understanding of the psychology of visualization and deception, storyelling, and advanced algorithms for visualizations (e.g., network layouts, seriation, multi-dimensional projection, rendering). Eventually, creating and using visualizations is now increasingly simplified through the spread of computer, displays, and software applications dedicated to data visualization such as Tableau, RawGraphs, R-Studio or D3.

In this large context of wicked problems, big data, storytelling, and the expansion of data visualization capabilities, I see the following general trends that contextualize research in that area: *i*) more people are using visualization through better interfaces and libraries, *ii*) more people are creating visualizations for data exploration and communication, and *iii*) people are looking for better visualization solutions to apply visualizations to ever more complex problems, involve them over a longer period of time, and obtain greater results with them.

1.2 Use cases

I had the chance to witness many of the Nine E's first hand through interdisciplinary collaborations with researchers in history and archaeology, sociology, conflict research and social inequality, industry partners, and public sector organizations such as the Scottish Government, the BBC, clinicians, NASA's Visualization Studio, through numerous workshops and tutorials on visualization design and use, as well as through teaching university students and working professionals in dedicated visualization courses. I will briefly describe two of these scenarios here which have strongly influenced and motivated the specific research described in this thesis.

Case 1: Visualizing multivariate, geographic, and temporal networks

Networks (graphs) are a common data structure to model and explore many real-world phenomena including social interactions, trade, chemical relations, similarity relationships, or computational models and so many more. Many of these networks contain additional data associated with their respective nodes and links in the form of attributes: geographic positions, temporal change, qualitative and quantitative data. Exploring this range of data, understanding correlations between those attribute, exploring change over time, and analyzing parts of the network requires sophisticated (interactive) network visualizations. The range of visualization techniques for networks is consequently incredibly rich including node-link diagrams, adjacency matrices, timelines, maps, adjacency lists, arc diagrams and countless more designs mixing and matching different visualization techniques (see respective surveys of network visualization techniques [Beck et al., 2017, Von Landesberger et al., 2011]).

There is a range of problems associated with designing (and eventually applying and using) such visualizations. First, few potential users know of the existence of this rich space of visualization techniques and the insights they can gain from them. Second, many of these techniques may not be intuitive once explained to a user. Third, once explained how to read a given visualization technique, users need to develop strategies how to use and explore a visualization technique to solve their problem (Boy et al. [Boy et al., 2014a] call this the translation from domain problems into visualization problems and back.) Fourth, visualizations are interactive, so users need to understand how to interact with an interface (interaction affordances), and why to interact in the first place. Fifth, visualizations are complementary and they can be used in concert to explore or to learn new techniques. Sixth, the more specific a user task and data set becomes, the more a visualization design needs adaptation, potentially including constant updating as a task evolves; eventually leading to novel designs every so often as those users analyze their way through the data and gain more information about their problem. Eventually, the communication of the respective insights are often hardly explained using these highly specialized visualizations.

To help us explore and tackle these problems, I lead the design and implementation of an online platform for network visualization. The Vistorian (`https://vistorian.net`) is an open source platform to interactively explore multivariate, temporal, and geographic networks. The Vistorian currently includes 10 interactive network visualizations (Figure 1.1) build with our toolkit NetPanorama (Section 3.3): (force-directed) node-link diagram, circular node-link diagram, adjacency matrix, adjacency matrix with arc-diagram, time-arcs, arc diagram, adajcency list, incident list, scatterplot, and node-link map. Each visualization provides for interactive exploration through timesliders, details-on-demand, pan and zoom, as well

Figure 1.1: *Landing page of visualization application at `vistorian.net`. The Vistorian currently features 10 interactive visualization techniques.*

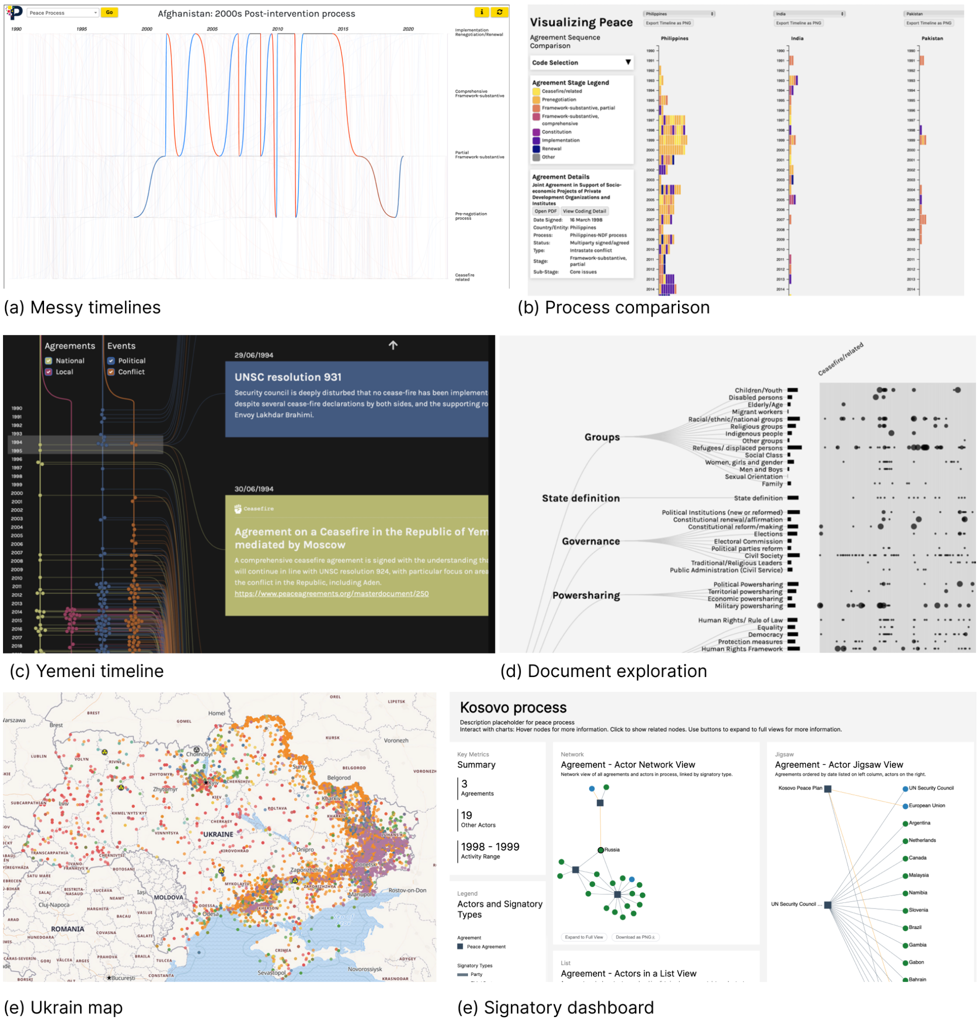
as multiple visual styling mechanisms for element sizes or transparencies. Visualizations can be used in parallel, synchronized through brushing and linking. Users upload and configure data tables through a wizard that helps them formatting and describing their data.

The Vistorian is targeted to support a wide range of audiences benefit from interactive network visualization for their work. It does not assume any previous knowledge and aims to support people's individual approaches to network visualization. There are two scenarios of using visualizations with the Vistorian. The first is the set of visualization techniques we provide while building the Vistorian: the visual mappings, visualization techniques, visual design, interactions, and the general user experience and context of the Vistorian tool. The second facet is about users curating their data, choosing a visualization, and adapting a visualization to their specific data and needs (e.g., through the visual design choices in the tool or as screenshots in a user's publication, talks, and notebooks). While the first part resembles a straightforward design process, supporting users in their design and exploration is non-trivial (see Chapter 4).

The Vistorian grew out of a personal collaboration with a historian and an afternoon of work in 2014 and has since been used in workshops and tutorials, keynotes, and supported and inspired research on data comics (Section 2.1), a network visualization toolkit (Section 3.3), automating storytelling (Section 3.2), understanding barriers to interactive network exploration (Section 4.1), building dashboards for user tracking [Wang et al., 2023], cheatsheets explaining visualizations (Section 4.3) and an approach to automatically explaining patterns in network visualizations (Section 4.4).

Case 2: Analysing and communicating peace and conflict

The second use case is in peace and conflict analysis and is the topic of a long-term collaboration with the PeaceRep program at the University of Edinburgh (`https://peacerep.org`) concerned with the systematic analysis of peace [Bell, 2024]. The



(a) Messy timelines
(b) Process comparison
(c) Yemeni timeline
(d) Document exploration
(e) Ukrain map
(e) Signatory dashboard

Figure 1.2: *Selection of visualizations and tools created as part of the PeaceRep collaboration. All tools are online on* `https://peacerep.org` *under the PeaceRep tools tab.*

research team asks *'What is good peace'* in order to inform ongoing and future peace negotiations, e.g., in Yemen, Myanmar, and South Sudan. With the team, based at the Edinburgh Law School, we build exploratory and explanatory visualizations to support the team, their collaborators, and stakeholders. Our visualizations show geographical, temporal, multidimensional, and relational data about peace agreements, their content, the evolution of peace processes. We have been working on interactive dashboards, storytelling, and interactive applications. A selection of those tools is shown in (Figure 1.2).

Designing visualizations for this project faces numerous challenges related to the seven E's, mentioned above. First, we need to create visualizations for a potentially wide range of people, including peace builders, journalists, politicians, students, educators, peace researchers as well as those directly impacted by these conflicts. Second, those visualizations need to carefully balance how they communicate and what to communicate; data needs to be scrutinized and verified; information needs to be framed in a specific tone, some information need to be withheld because they can cause political dissonance or jeopardize the peace building. Third, data sets can be extremely large such as the ACLED data based collecting data about any conflict on the world. Making such large data sets accessible to people for monitoring (Section 2.2) and use beyond downloadable data dumps, i.e., through visualization, requires large repositories for visualization or visualization atlases (Section 2.3), but also training in the use, design, and creation of visualizations (Section 4.2)

1.3 Outline and Contributions

Given these experiences, the central question of my research has been *How to communicate complex information with visualization?* and *How to support people in creating such forms and the respective visualizations?*. Visualization requires and supports human intelligence: creativity, design, iteration, readjustment, understanding the context of the wicked problems. It requires ideas and imagination to create and design appropriate visualizations for complex data, to build and use interactive interfaces, and to create compelling presentations of the results. Perhaps hardest to understand is the fact that visualization is an inherently iterative process—both their design as well as their use, a fact essentially blurring the boundaries between design and use: designing is using and using is designing [Hinrichs et al., 2019]. In designing a visualization, a person thinks about the problem, tries to understand the questions their are interested in, searches for structures in the data and anticipates what patterns and information a visualization should show. They explore solutions and assess the amount and quality of insight gained from each; they play with the data and gradually build an understanding of those data while working towards a valid visualization design. Likewise, using a visualization is designing, as people judge the value of a visualization, ask new questions and consequently seek new visualizations. Eventually, the explanation of findings and insights requires the same processes of designing clear explanations and visualizations, structure information and interactions, and anticipate the use of those visualizations and explanations through others. Clear and engaging presentations and explanations are key to make complex systems, relationships, and decisions understandable and to obtain buy-in from many people.

In all this, visualizations are horizons—as soon as one sees them, one realizes the limit of what one sees and knows; yet one understands where to go next. In other words understanding data through visualization is *understanding (data), one visualization at a time*. While separating the known from the unknown, with the right equipment and approaches, horizons become movable frontiers, discovering one step at a time. At the same time, they demand guidance for the novice and training in the required techniques. Inspired by these scenarios and my experience in working with all the people I have worked with, I focus this document on three dimensions of my research.

1/ Forms: How to communicate data about complex issues?

Chapter 2 talks about *forms* of visualization for data-driven storytelling, i.e., the types and genres [Segel and Heer, 2010] embedding visualiztions for data-driven storytelling. These forms are higher-level structures and platforms for the storytelling and can include many hundreds of individual visualizations to account for extensiveness of large datasets. The chapter presents two novel formats **data comics** (Section 2.1) and **visualization atlases** (Section 2.3). Both formats are complementary approaches to presenting large and complex data sets in the context of explanation, exploration and some of the other Nine E's. The chapter then provides the first analysis of design patterns for **dashboards** (Section 2.2) to help streamline dashboard design. Related contributions to formats not explicitly described in that chapter include an analysis of **data articles** [Hao et al., 2024] and the first description of **DataGIFs** as a novel medium for data-driven storytelling [Shu et al., 2020].

2/ Tools: What software tools are required to design visualizations?

Designing visualizations for exploration, presentation, and storytelling—dashboards, comics, atlases, network visualizations, etc.—requires designing, engineering, and iteration. However, existing tools and toolkits for visualization design are either focusing on interactive exploration, the design of traditional individual visualizations, or require purely functional programming. In Chapter 3, I present authoring tools to create visualizations and data-driven storytelling as well as explore large data sets by novice audiences. The chapter introduces an authoring tools for **interactive data comics** (Section 3.1). Then, the chapter introduces **NetworkNarratives** (Section 3.2) a system and interface to create data tours for the semi-automatic guidance of novices for large datasets supporting explanation and exploration, as well as education and exceptions. Eventually, it introduces a toolkit, **NetPanorama** (Section 3.3) for network visualization inspired by declarative programming and formal visualization grammars such Vega-like [Satyanarayan et al., 2016].

3/ Skills: How to train people in using visualizations?

In the third and last Chapter 4, I talk about approaches to train people in using and creating visualizations. I refer to training in the widest sense, in the sense of education and understanding as required in the scenarios I described above: domain experts and working professionals aiming to solve real-world problems with their visualizations. While I have run workshops with students and working-professionals on visualization design and visualization tools, most analysts work without a visualization expert supervising them. This results in cases where people struggle on their own to understand good visualization design, to find and master visualization and tools, and to create their own design knowledge from scratch.

To that end, Chapter 4 first summarizes **barriers** we observed in our field-research about network exploration visualization with the Vistorian (Section 4.1) and then summarizes general **challenges in education** originating from discussions at a Dagstuhl seminar (Section 4.2). Last, I introduce two novel techniques to support the training of visualization which focus on informal educational settings, i.e., no specific teacher, unstructured time, self-driven learning—**visualization cheatsheets** (section 4.3 and **interactive pattern explanation** (Section 4.4).

I conclude this thesis with a reflection on the lessons learned and open questions in the field of data visualization design in Chapter 5.

Chapter 2

Forms for Communication and Presentation

The effective communication of data through visualization requires means dedicated to exactly that task. The task of presenting for communication, storytelling, or explanation can be described as inherently different from an analytical and exploratory scenario. In analytical and exploratory scenarios domain experts scrutinize, analyze, and interpret large amounts of data over days or months in dedicated lab settings with possibly large screens, virtual reality, and supercomputers. In explanatory scenarios, we usually find authors presenting and set of highly curated findings and messages to a more or less specific audience, in a possibly very wide range of contexts: life presentations, mobile phones, websites, news, museums, etc.

While analytical and exploratory scenarios can be described as *data-centric*, explanatory scenarios can be described as *human-centric*. In such a human-centric scenario, the goal is not to understand and scrutinize data, but to talk to people, to teach them something new, to catch their attention, to engage them, and to possibly entertain them. In most of the cases, there is a specific goal of inciting some behavior, such as being aware of a topic, understanding data about a topic, or making decisions. Explanatory visualizations have to be effective in their communication, i.e., provide context about a given topic, explain data and visualizations, clarify potential biases, and adjust to the audience's level of visualization literacy [Boy et al., 2014b] and data literacy.

To structure approaches to data-driven storytelling, Segel and Heer's *narrative visualization* [Segel and Heer, 2010] provide a taxonomy of genres: data videos, annotated charts, magazine articles, comics, partitioned poster, flow chart, and slide show as common genres for data-driven storytelling.

Outline

In this chapter, I describe additional forms and ideas for data-driven storytelling, communication, and presentation, each offering complementary affordances and design possibilities. One of these formats—data comics—I have studied in more depth, while for others: dashboards (Section 2.2) and visualization atlases (Section 2.3) I offer analytical descriptions in the form of design pattern and specific genres.

2.1 Data Comics

Data comics are inspired by the way traditional comics tell stories and combine text with visual context. Comics can be broadly described as *sequential art* [Eisner, 2008] and *juxtaposed images and words* [McCloud, 1993]: they represent time through space by panels and their ordering. 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]. Comics can use transitions [McCloud, 1993], simultaneously providing detail and overview (*co-presence* and *configuration*) [Groensteen, 2013], thereby preserving visual permanence [Yang, 2008] and allowing readers to read the story at their own pace. Besides their linear structure, comics have a rich history of integrating unexpected logics into their structure, such as hierarchical linguistic grammars [Cohn, 2013], repetition [Shores, 2016], or musical elements that carry temporality, emotion and a joyous disregard for staying in their expected place [Brown, 2013].

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]. Comics can blend explanation (e.g. schemata, illustrations, data visualization) with narration, characters and dialogue. From a structural point of view, a notion central to comics is the *panel*, which encapsulates a specific *message* (or information) represented as an integrated combination of text and picture [McCloud, 1993]. The (mostly) linear order of panels creates a sequence of view-points which together can build a deeper understanding in the readers' mind [Sousanis, 2015]. Empirical evidence from using comics as classroom material [Hosler and Boomer, 2011a, Spiegel et al., 2013a, Short et al., 2013a] as well as for health-related communication [Delp and Jones, 1996, Tekle-Haimanot et al., 2016, Diamond et al., 2016, McNicol, 2017] suggests that comics may be more engaging and easier to comprehend, especially for non-expert readers.

Comics have consequently been used for explaining a range of phenomena, blending schematic drawings and illustrations with the narration and characters of traditional comics, for example to facilitate classroom education [Hosler and Boomer, 2011b, Spiegel et al., 2013b, Short et al., 2013b], conveying scientific phenomena [Tatalovic, 2009, Farinella and Ros, 2014], explaining concepts related to health [Williams, 2012], and to storyboard interaction design [Dykes et al., 2016].

Despite being named as one of seven genres by Segel and Heer [Segel and Heer, 2010] (based on two instances of comics, none of which actually contining any data), comics are not yet a common genre for data-driven storytelling. The other official mention of data comics was by Zhao and Elmqvist in their 2015 paper *"The Stories We Tell About Data: Surveying Data-Driven Storytelling Using Visualization"* [Zhao et al., 2015], describing the idea of using comics for data-driven storytelling and presenting a simple editor that allows creating panels, visualizations, and text. Byond that, few examples of data comics existed when we started working on this topic and there had not been any structured analysis nor guidance for their creation. The Nature provides a high quality example on the challenges of climate change [Monastersky and Sousanis, 2015] in 2015 combining data visualization, narration, linear-and-non linear flow, into a compelling, visually striking and even emotional piece. Cisneros created an interactive comic using Tableau Graphics [Cisneros, 2011]. Additional examples include a scientific report on the change of water temperatures [Bernaerts, 2006] and graph comics [Bach et al., 2016a].

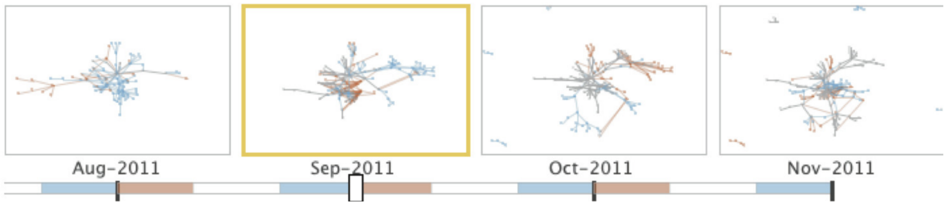


Figure 2.1: *Small multiples in an analytic exploratory scenario, showing*

Outline

With so few examples existing, there were no guidelines, methodologies, tools, or empirical evidence about the effectiveness of comics and to inform their creation. In our research, we explored the meaning of comics as a new genre [Bach et al., 2017a], explored and created comics and design patterns for data comics for networks (Section 2.1.1), statistical analysis in the context of controlled user studies [Wang et al., 2020], created general design patterns for comics (Section 2.1.2), studied comics empirically in the context of infographics and illustrated texts (Section 2.1.3) as well as studied comics for explaining visualization literacy and build our own tools (more on that in Chapters 3 and 4). Sections based on published or submitted papers reference those papers and co-authors at the beginning of each respective section.

2.1.1 Data comics for Dynamic Networks

Our systematic investigation of data patterns started with data comics to facilitate showing changes in dynamic networks. This research was inspired by my previous work on visualizing dynamic networks through small multiples [Bach et al., 2015a, Bach et al., 2013, Bach et al., 2014] in which individual time slides of a network's state are shown side by side (Figure 2.1). Small multiples conceptually seem one step away from comics, yet comics would need to focus on explaining changes, rather than showing everything. To understand how comics would be created for such dynamic networks, we created graph comics (one example is shown in Figure 2.2 with more examples online `https://aviz.fr/~bbach/graphcomics`) by hand for various data sets and application scenarios we were familiar with according to a structured process (Figure 2.3). Our consequent analysis of this process and the resulting comics resulted in eight types of design patterns (called Factors in the original paper, Figure 2.4).

Published as: ***Bach, B.***, *Kerracher, N., Hall, K.W., Carpendale, S., Kennedy, J. and Henry Riche, N., 2016, May. Telling stories about dynamic networks with graph comics. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 3670-3682).*

Design patterns for graph comics

Those design patterns—representing specific solutions for the design of graph comics—were grouped into the following categories and illustrated through examples (Figure 2.4). More detail and examples can be found in the original paper.

- **Visual Representation of Graph Elements**: To show characters and objects, most traditional comics use depictions of the real world. The elements we consider in graph comics are more abstract and, for graph comics, limited to three classes: nodes, links, and groups (or clusters). Our patterns include visual symbols for showing nodes, links, clusters, and visual attributes in networks.

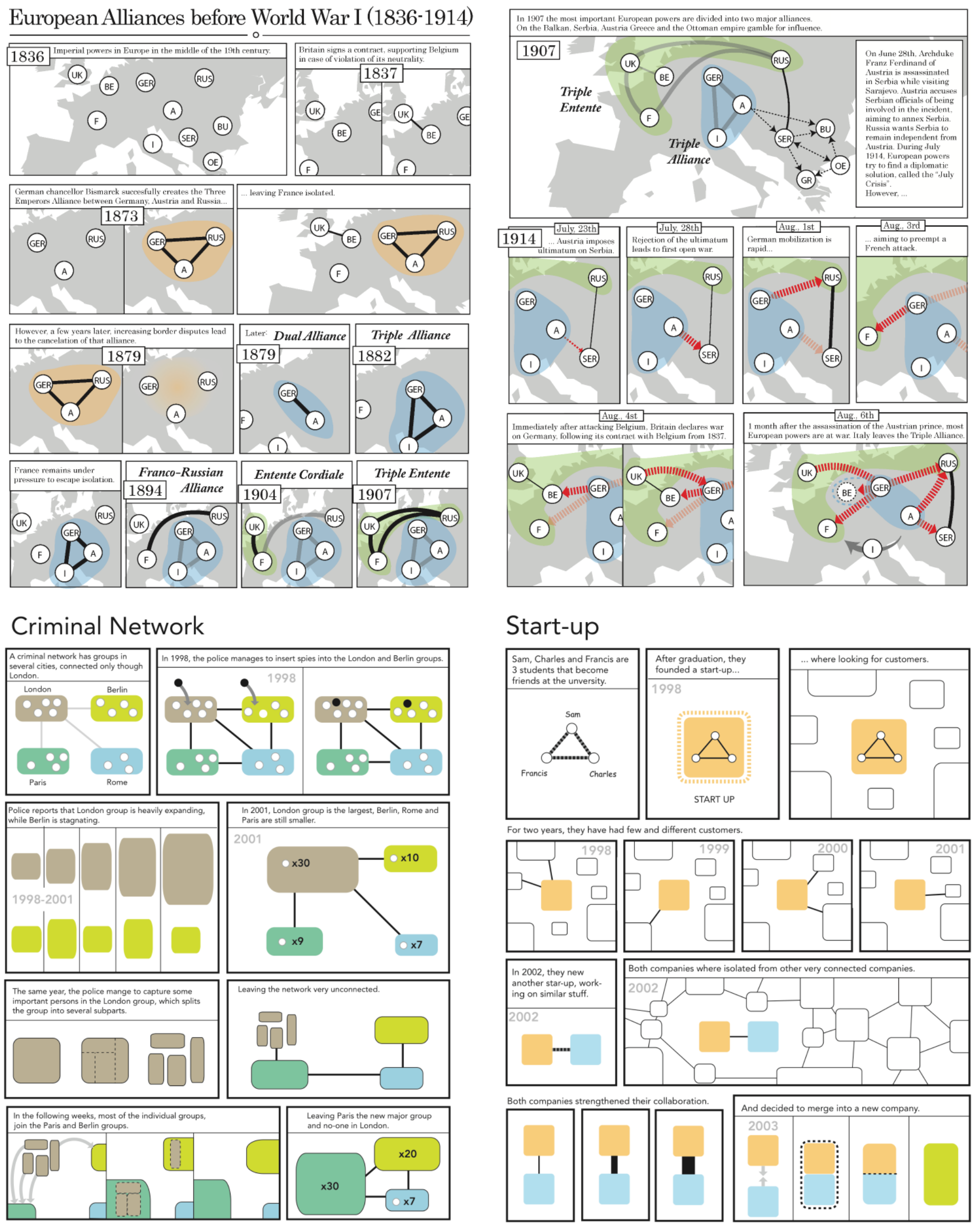


Figure 2.2: *Graph comic examples. Top: the evolution of alliances in Europe before the 1st world war. Bottom-left: changes in a criminal network. Bottom-right: the story of a start-up connecting with other companies.*

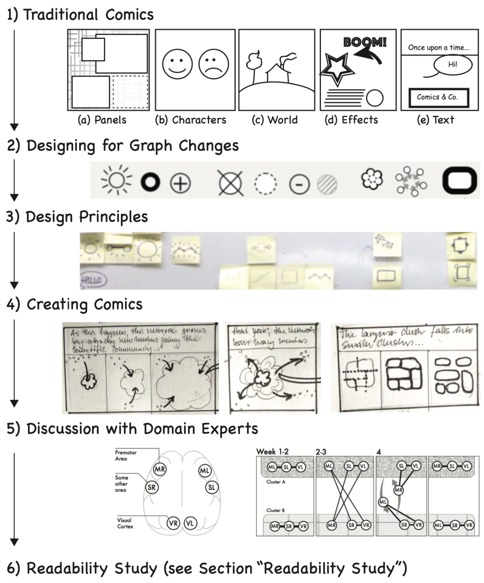


Figure 2.3: *Multi-step graph comics creation process.*

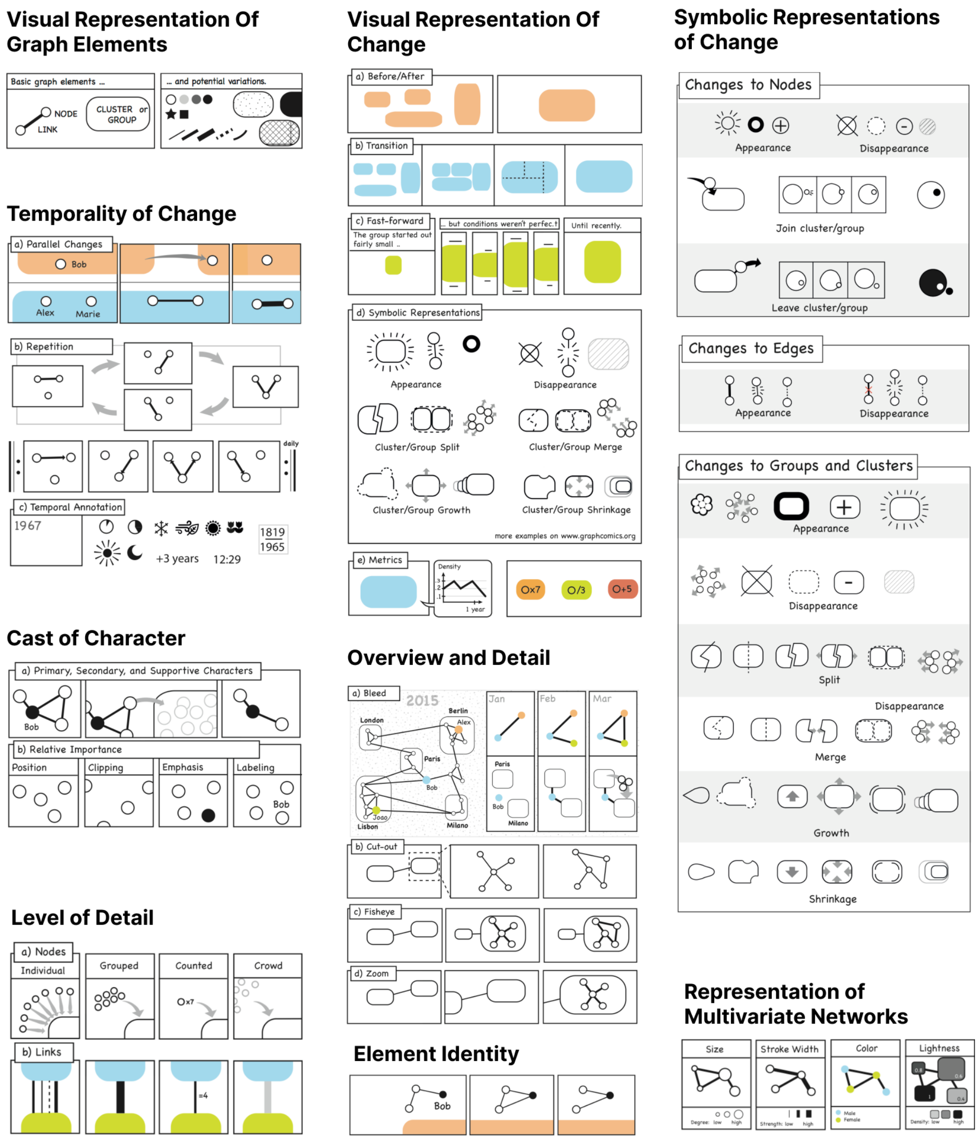


Figure 2.4: *Examples of graph comics patterns used to convey information in graph comics.*

- **Visual Representation of Changes**: In traditional comics, spatial layout and sequencing of panels defines the storyline and its temporality. For graph comics, we propose to convey temporal change through transitions, before-and-after settings, metrics, or symbolic representations.
- **Temporality of Changes**: Readers of traditional comics may reasonably guess the duration of events by the number of panels used to represent an action and by analogy to the real world (e.g., one panel depicts nighttime, the next depicts daytime). Graph comics require additional techniques and annotations to illustrate changes that occur in parallel or repeat, and to explicitly encode temporal duration: changes can happen in parallel, changes can repeat, be paced in a specific way, or indicated through specific graphical symbols.
- **Element Identity:** In traditional comics, identification of recurring elements occurs through the identifiable features of a character's physical appearance, and through the character's position within a scene. In graph comics, we propose four alternatives for identifying recurring objects, e.g., textual labeling, specific colors, identical places or shapes.
- **Cast of Characters:** The main character(s) in comics is usually one (or a few) elements, on which the story focuses. Supporting characters are related to the main character, e.g., by connection, or are otherwise important for the understanding of the story and are countable. Extra characters are characters that represent other elements in the graph whose identity and precise number is not of importance. We propose three strategies to distinguish between these types of characters: highlight the main character, while showing other characters (e.g., nodes) more salient, or place them at the margin of the panel, or crop or otherwise hide them, or simply not labeling them.
- **Level of Detail:** Related to the previous consideration is the level of detail in describing changes. Figure 10 illustrates an example of addition of nodes to a group (a) and the connections between two groups (b) at four levels of details.
- **Overview and Detail:** Often the focus of a story switches from a larger collection of elements, e.g., an entire network or a set of groups, to fewer elements, e.g., nodes in a group. Graph comics offer several ways to show transitions between overview and detail through, e.g., illustrating a zoom, using cut-outs or a fisheye representation.
- **Representation of Multivariate Networks** Additional information about graph elements, such as node type (e.g., male, female), edge type (e.g., married, friend), edge weight or group density can be shown through visual variables such as colors, line thickness etc.

User study

As part of this first investigation of data data comics, we ran a study with ten participants reading our comics. Our hypotheses were that (H1) graph comics can be understood without training, i.e. readers can correctly interpret the story, and, (H2) that ambiguities can be resolved via discussion. We prepared three initial comics including visuals and annotations. After each session, we iterated on the designs that were ambiguous to participants, and steadily decreased the amount of annotations to assess how interpretable the visuals were without text. None of our participants was knowledgable in graph analysis or visualization. To ensure participants would vocalize their interpretations, we ran sessions in pairs. We held five 1-hour sessions under the same conditions. Each participant received a color copy of the same comic and was encouraged to first read and annotate them before discussing among themselves to reach a consensus. Participants then reported their interpretations and the experimenter asked for clarification and findings participants did not report explicitly if necessary Video and audio was recorded at all times.

In a nutshell, our results show

- All **participants generally correctly interpreted all three comics**, independently of the amount of annotations present, confirming H1.
- While **annotations undoubtedly helped participants** resolve ambiguities and interpret comics faster, these were not essential for understanding comics, confirming H2.
- **Change is understood in context.** The specific story context (criminals, family, start-up) has a strong impact on the interpretation of graph changes and helps resolve possible ambiguities in the graphical representations.
- **Symbolic representations require annotations**. While symbolic representations and conventions naturally require learning, we hypothesized that encodings restricted to depicting graph changes were discoverable without any annotations or training. However, better to provide textual annotations the first time a symbol or graphical entity is introduced.
- **Visual encodings borrowed from node-link diagrams are appropriate** and participants understood before/after states, transitions and metrics to encode graph changes. The number of panels was perceived as proportional to the duration of a change.
- Relative **positions, size, and color** are important to maintain element identity across panels, though color receives priority in making decisions:

Summary

Through our investigation, it became clear to us that comics are a flexible and expressive medium, enabling communication with a variety of visuals and text. We refined a set of key design patterns and used these as guidelines to generate graph comics more systematically. While there are still a large number of possibilities to convey the same dynamic network story, we offer these design considerations to provide guidance for generating consistent graph comics. Our reading study with 10 novice participants provided initial evidence that graph comics can be understood by a wider audience, with minimal textual annotations and without any training.

Following this work, we were curious how a design process and design patterns would generalize and expand to other types of data beyond dynamic networks.

2.1.2 General design patterns for Data Comics

To create a higher-level and potentially large collection of data comics design patterns, we composed and refined a set of design patterns by following three complementary methods. First, the authors of this paper—an interdisciplinary team involving a comic artist, an art graduate, as well as researchers in visualization—created over ten data comics for different data and audience (Section 3). By crafting and iterating over these comics, we reflected on an initial set of design patterns that we found useful in the creation process and that reoccurred across the comics we

Published as: ***Bach, B.***, *Wang, Z., Farinella, M., Murray-Rust, D. and Henry Riche, N., 2018, April. Design patterns for data comics. In Proceedings of the 2018 chi conference on human factors in computing systems (pp. 1-12).*

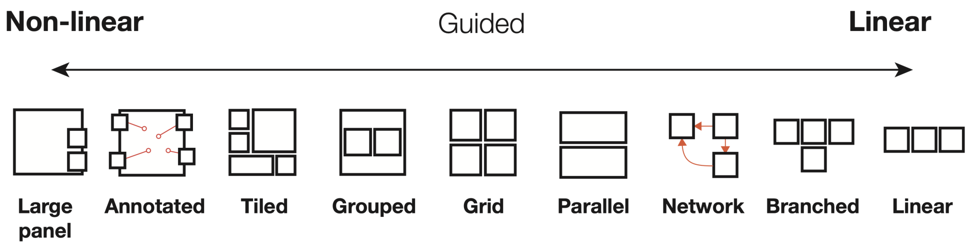


Figure 2.5: *Panel layouts found in infographics, loosely ordered by their degree of linearity. Red elements are flow-marks, indicating reading direction.*

generated. Second, we identified a design space to systematically describe our patterns leading to the description of new patterns (Section 4). The design space is informed by i) coding spatial layouts common in the space-oriented genres of infographics and traditional comics, combined with ii) linear narrative progression between panels (frames) in the time-oriented genres, building from Hullman et al. [Hullman et al., 2013]. Third, to assess the value of our design patterns for teaching and creation of data comics, we ran a workshop with 23 participants creating data comics for multiple datasets (Section 5). We printed a set of design patterns cards akin to IDEO [ide, 2016] and provided them to the workshop participants. Participants leveraged these cards to create a set of comics from data we provided them. Subjective feedback from the workshop suggests that the cards were useful in the ideation and execution process. All material in this paper including examples of existing data comics, comics that inspired our patterns, our own data comics, design cards, as well as the entire material of the workshop, can be found online: `http://datacomics.net`.

A design space for data comics design patterns

As data comics combine spatial presentation (panel organization on a page) with narration (text), we define a data-comic design-pattern as *a set of panels with specific layout and content relation*. Consequntly, our design pattern design space combines one dimension for spatial panel-layouts with one dimension for content relation between panels. A design pattern can be described by a combination of both dimensions (Figure 2.6):

- **Dimension 1: Content Relation** describes how panels in a pattern are related by their content, i.e., how they create a narrative and what they want to say, e.g., a temporal change, the introduction of a visual encoding, a set of facts about the data, the explanation of a visualization technique.
- **Dimension 2: Panel Layout** describes the specific spatial structure of the layout and how panels are arranged in a temporal and logical order (Figure 2.5)

A design pattern fills one cell in the design space in Figure 2.6. Patterns for the empty cells might be possible but we did not find any particular purpose for them. In the following, we describe types of patterns according to the classification along the Content Relation dimension.

- **Narrative patterns** create connections between the data visualizations and the reader, the narrator, and the context of the data such as introding the data context, problems and questions, address the reader and involve them into the story, or highlights different aspects of the data across several patterns, keeping the visualization the same to avoid overloading a single large picture.

LAYOUT → / CONTENT RELATION ↓	Large panel	Annotated	Grouped	Tiled	Parallel	Grid	Network	Branching	Linear
Narrative		State panels ★		Multiple-explanations ★		State panels ★	Flashback ★		Exposé ★ Multiple explanations Question & answer ★
Temporal		Time-overlay	Time-nesting	Moments	Before/After ★	Time-grid	Time-States	Alternative tracks	Time-Sequence ★ Overview+detail ★
Faceting				Multiple facets ★	Contrast ★ Alternatives ★	Multiple facets ★		Alternatives ★	Overview+detail ★ Gradual Reveal
Visual Encoding	Legend ★								Build-up ★ Legend ★ Annotated transition ★
Granular	Overview+detail ★ The larger picture ★	Cut-out Lens							Zoom Cut-out
Spatial		Space-annotations ★		Tiled-polyptych ★	Parallel-polyptych ★	Grid-polyptych ★	Space-walkthrough ★		Pan

Figure 2.6: *Design space for data-comic design-patterns: panel layout (horizontally) and content relation (vertically). Darker cells indicate more patterns. Some patterns occupy several cells as they can be expressed with different layouts. The full list of patterns with names and examples in found online. (see supplementary material). Red stars mark patterns we have found or created examples for. No stars mark patterns generated systematically from the design space.*

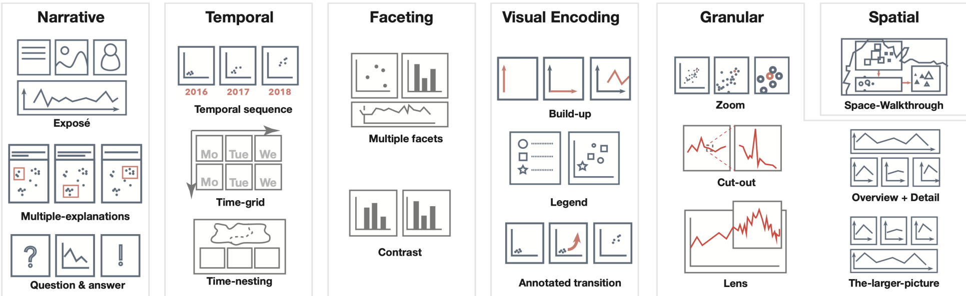


Figure 2.7: *Illustrations for some example design patterns in data comics, organized according their content relation. Stars mark patterns that we identified through making comics (Section "Data Comics Use Cases").*

- **Temporal patterns** communicate a temporal change in the data through temporal grids (like calendars), sequences, hierarhical nesting to indicate hierarchical time, varying the panel width to indicate shorter or longer events into a panes summarizing longer periods (e.g. important events in each century).
- **Faceting patterns** deliver complementary views on different parts of the data (similar to comparison in [33]). In a multiple facets pattern, panels are loosely assembled into a tiled layout or a grid layout. Panels are narratively connected by explanation and the delivering of facts. Con- tent of two panels can be placed in contrast by isolating two panels and place them into a parallel layout. Alternatives are shown as sequences, branching out of a linear sequence.
- **Visual encoding patterns** describe explanations that help the reader to understand the presented visualization. A build-up can explain the structure of a visualization through a sequence of panels, by gradually introducing the parts; e.g. introducing axes, scales, as well as visual marks and visual mappings. Similarly, a legend can be shown in a single panel prior to showing the visualization in order to make sure the reader is at least aware of the visual encoding.
- **Granular patterns** relate panels with different levels of detail. The most prominent example is a zoom, which can be expressed by a sequence of panels that guide the readers. Alternative patterns are a cut-out or a lens. A little more high-level are overview+detail and the-larger- picture. Both involve a large panel, in combination with a sequence of smaller ones. In the first case, the large panel shows the overview for a complex visualization. The following panels then explain details which all relate to the issue shown in the first panel. The-larger-picture is the inverse in that a larger detailed panel finalizes a series of smaller panels and can close a story with a general overview, potentially inviting the reader to further

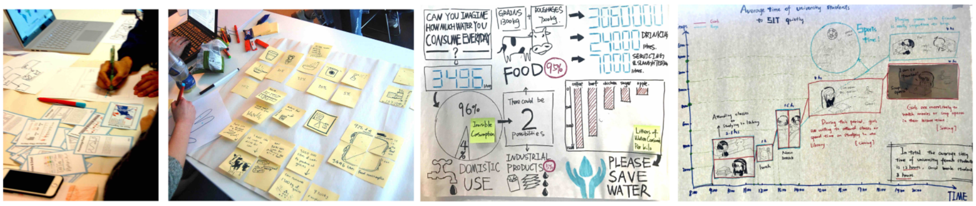


Figure 2.8: *Data comic workshop using design patterns printed on cards, guided activities such as storyboarding and low-fidelity protyping.*

exploration.

- **Spatial patterns** create a narration through the same visualization picture. A space-walkthrough consists of a single large picture in the background, while over- laying panels in a sequence at the important parts of the larger image. For example, a map showing the route of a travel while each station of the journey features its own panel. Panels may be related through the flow-mark of the journey path.

Design Workshop

We ran a workshop (Figure 2.8) to investigate: *i*) the perceived usefulness of data comics; *ii*) whether our patterns help novices to create data comics; *iii*) how difficult it was for novices to adapt design patterns to their own stories; and *iv*) whether we could find novel design patterns. The workshop asked participants to create one or more data comic storyboards—with an emphasis on sketching rather than final polished versions.

Participants were given physical design-pattern cards, each showing a pattern name, a brief description of the pattern, an abstract illustration, as well as an example data comic on the backside. Workshop participants were 23 master students in design and computer science. No specific skills were required. For our 3.5 hours workshop, participants worked in pairs to support rapid working and iteration. Participants decided themselves on a data set, either by browsing the visualizations in our Github-wiki (global migration, CO2 emissions, global health) or searching visualizations on their own initiative.

During the workshop, we obtained 11 comics. Our findings include:

- Asked for **the potential of data comics** to communicate information participants gave an average rating of 5.6 on a 7-step Likert scale.
- **Usability and usefulness of design patterns** were generally rated very positive with an average rating of 7.8, (mode=9) on a 10-point Likert scale. The design cards have been proven very useful, quickly helping workshop participants to engage with a new medium, with only little or no knowledge at all in visualization and comic making. The design patterns allowed our participants to draft both a story and a comic layout in a very limited amount of time.
- coding patterns in the resulting comics suggests that **participants incorporated the design patterns** from the cards into their comics
- While at first sight, some creations looked like infographics, most contained **a clear structure and a driving narrative**, making the results different from most presentation-oriented infographics

Our workshop showed a general interest in and the usefulness of both data comics, storytelling, and our design patterns. Consequently, we got approached by various university scholars to run extended versions of our workshop as part of their course

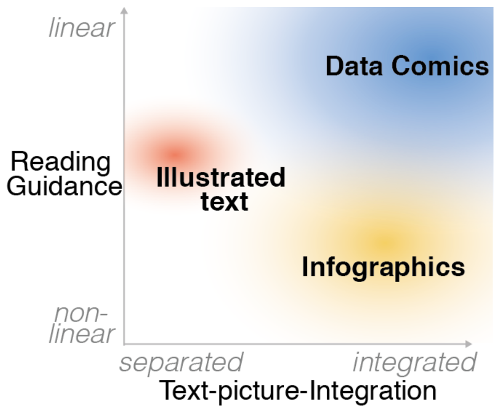


Figure 2.9: *Design space showing the differences between illustrated text, data comics, and infographics on two dimensions: Text-Picture-Integration (loose, to close), and Reading Guidance (low to high).*

curriculum (graphic design as well as computer science). The results of one such workshop are detailed in another paper of ours [Wang et al., 2019a].

Summary

This systematic space for design patterns is a tool to help thinking about data comics and it may not capture all future design patterns. Additional aspects of data comics might lead to complementary and types of design patterns, e.g., visual-stylistic patterns, patterns about the usage of text, patterns for specific visualizations or data types, or higher-level patterns that structure entire stories.

2.1.3 Comparing data comics to infographics and illustrated text: a controlled and in in-the-wild study

Published as: *Wang, Z., Wang, S., Farinella, M., Murray-Rust, D., Henry Riche, N. and Bach, B., 2019, May. Comparing effectiveness and engagement of data comics and infographics. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-12).*

Inspired by our data comic examples and qualitative studies (Sections 2.1.1 and 2.1.2), we wanted to better understand how data comics would be read and understood. To that end, we compare data comics to its most closely comparable other genres for narrative visualization: infographics and illustrated reports (text+image). 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 (Figure 2.9). Both formats 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. Such texts provide high reading guidance but low integration of picture and textual content. In contrast, data comics provide both high text-picture integration and high reader guidance.

Generally, information is remembered better when it is supported with pictures [Levie and Lentz, 1982], even more so when presented through both channels at the same time [Paivio, 1990, Clark and Paivio, 1991, Baddeley, 1997, Sultanum et al., 2018]. Data visualizations have been found less memorable than natural scenes [Borkin et al., 2013],

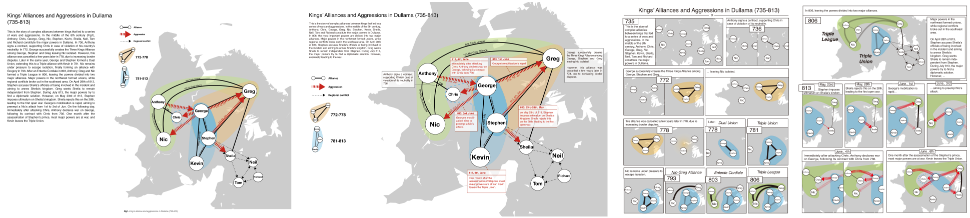


Figure 2.10: *Sample from the study showing the same story visualized in as Illustrated text (left), infographic (center), and data comic (right).*

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. 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, 2013, Cohn, 2014] used in traditional comics through eye-tracking [Foulsham et al., 2016]. Other studies have found that in cases where text is unnecessary, 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].

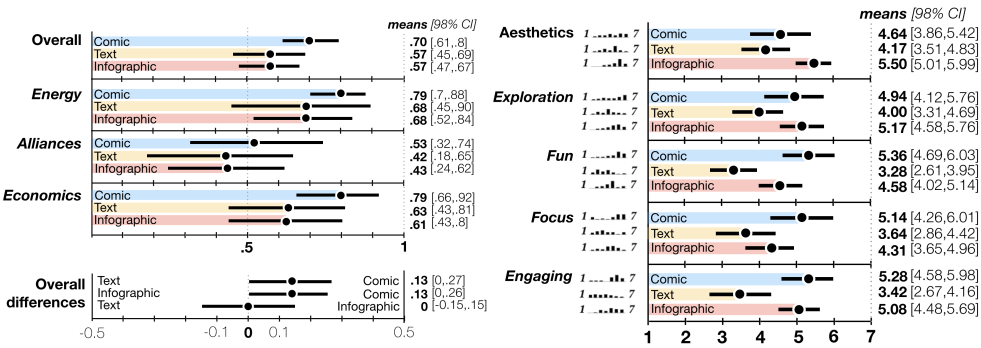
However, while most of the evidence from previous studies supports a strong motivation for the use of comics, the effects were observed in traditional or scientific comics and may or may not generalize to *data* comics. In particular, graphics in data comics focus more on abstract data visualization [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 point to the importance of consistency in creating sequences of visualizations [Qu and Hullman, 2018], a factor that very likely influences the acceptability and usability of data comics.

We compared these techniques in a controlled lab study, where 36 participants read the same 3 stories, each presented in a different format. 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. Study materials for the Lab study, in the wild study, questions, some interview quotations and examples of reading sequence record, can be found online: `http://datacomics.net`.

Controlled Lab Study

In this study, participants read three data stories explained (full details can be found in the original paper [Wang et al., 2019b]). We selected different types of common visualizations including multivariate data, dynamic networks, and geological distribution. All comics were the same length and counterbalanced for visual and content complexity, with visualizations about social networks (Figure 2.10), multivariate energy data, as and global interest rates and tax burdens with data taken from Gapminder [Hans Rosling and Rönnlund,].

We recruited 38 participants with a background in Design, Art, Computer Science, Engineering, Linguistics, Philosophy and Psychology to read each of our story (each story in a different format). After reading each of the stories in a given order, we asked participants questions about the understanding of the stories, e.g., about



(a) (b)

Figure 2.11: *Left: measures efficiency for data comics, infographics, and illustrated text (measured as correct answers). Right: subjective ratings of the same formats.*

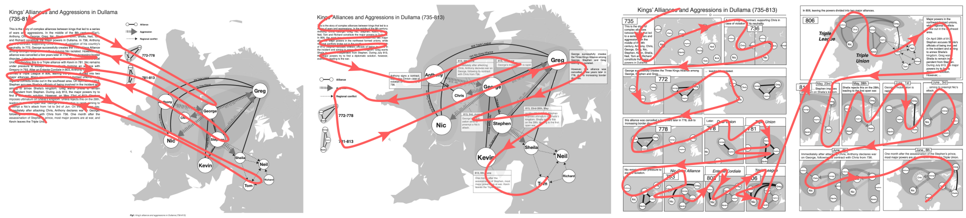


Figure 2.12: *Reading paths as explained by participants in the study (no technical eye-tracking was involved, but paths had been described in interviews.*

distributions in the data, temporal change, individual facts, outliners, and specific visual encodings. We also asked participants to explain how they read the comics, infographics, or texts, i.e., in which order they went through the visual material. After one day, we brought participants back and asked them to re-tell us all the stories and then generally discussed all the stories and stimuli with them to get more qualitative feedback.

The main results from this controlled study include the following findings, with some of the quantitative data shown in Figures 2.11a to 2.12.

- Across stories we found **comics more accurate** than infographic and illustrated-Text.
- on **subjective rankings, we found Comics to be highest rated** (averages) on three measures: fun, engaging, and allowing people to stay focused. Comics scored slightly **less for aesthetics and exploration** compared to Infographic.
- participants **preferred to use comics in future**, over any of the other formats
- **for temporal data, infographics performed worse** because comics are better in showing temporal sequences.
- **Comics were appreciated for their clear reading** order and their ability to break down the complexity into pieces .
- **clear reading order in comics** was found use support memory, esp. for temporal content
- participants liked that **comics could group higher-level messages** into rows and potentially pages.
- comics can **overview information** easier through their layout and multiple panels, akin to small multiples in analytic scenarios.
- Comic **visual repetitions** can be distracting if too many
- where comics were not so good was in cases where details had been overlooked by participants. This happened when **too many visual elements** populate the

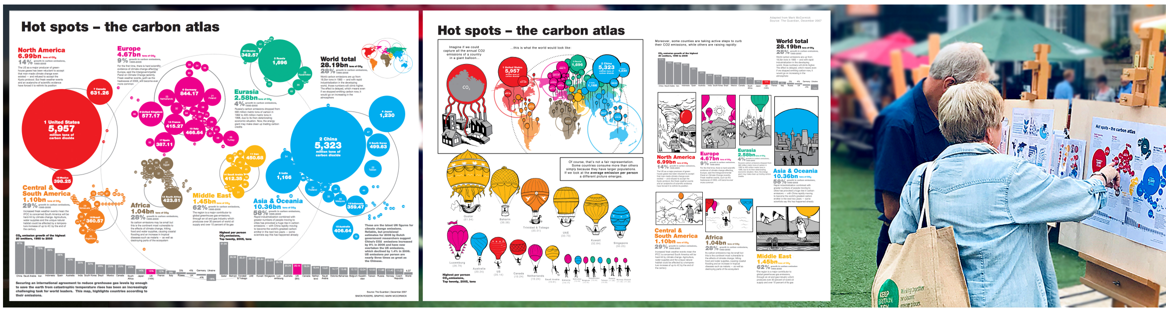


Figure 2.13: *Example of infographic (left), data comics (center), and visitors of both in the in-the-wild study.*

comics and panels.

- comics might miss overview picture, i.e., to orient the viewer in the story and data. that's easier with infographics.
- **infographics were better for spatial data**, also give good overview, but lack a clear reading order and explanation of temporal data.
- **illustrated texts had been described as clean and familiar**, however jumping between text and picture causes disruption as readers have to create their own connections, complaining about the high density of the text

In-the-Wild Study

To overcome the somewhat controlled and made-up nature of the controlled study, we were further interested in how people at an international art festival would engage and interact with actual, professionally created data comics and infographics. We removed the text condition here since evidence from our earlier study did not show major benefits compared to the other two formats. Visitors of the art festival varied in age, interest, and cultural origin, enabling us to study a more diverse set of participants and more casual encounters of data comics on the street. Hiding in some distance, we mounted a data comic and an infographic—both created by a professional comic artist and about the topics of global CO2 footprints and water consumption—each showing a different story and took notes of people's behavior as well as conducted semi-structured interviews with some of the viewers (Figure 2.13).

During 4 hours, a total of 43 visitor 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. Our main findings from observations and interviews include:

- We found **no difference in time** spent on either format or story.
- we found **slightly more *engagement* with Infographic** as measured by discussion and pointing. We attribute this the more exploratory and less guided nature of infographics.
- more groups found **comics more understanding** than the infographic (based on interviews).

2.1.4 Data Comics Design Guidelines

Based on these results, we defined the following design guideliens for data comics:

- **Balancing repetition and highlighting** to avoid too many details and focus attention. Possible solutions include:
  - Explicitly highlighting changes between panels

  - Using cut-out patterns to emphasize small changes.
  - Combining messages into one panel with annotations, like infographics.
  - Large-picture patterns to serve as a mental map before detailed explanations.
- **Balancing sequence and overview:** Data comics have been criticized for lacking overviews, especially when showing details of a bigger visualization. Solutions could be:
  - Carefully pacing overview pictures, i.e., include large pictures that show more context.
  - Making sure zoomed-in content is understood within the larger context.
  - Using larger pictures with annotations for detailed visualizations.
- **Using the layout to structure information:** Comic layouts can structure information using panels, rows, pages, and even panels inside panels. Panel size, number, and layout can be used to group messages, control reading pace, and show importance. A clear page layout with overview panels can support information lookup during reading.
- **Reducing visual complexity:** Panels can be overloaded and cluttered, especially when small. Possible solutions include:
  - Creating larger panels for complex content (fewer panels per page).
  - Consistent visual elements to reduce overall clutter, as long as changes are highlighted.

Conclusion

This first study on the effectiveness of data comics yielded plenty of results across its two studies. We could conclude that data comics led to more correct answers on average, 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: 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. These results suggest that increased text-picture integration and more reading guidance can lead to better understanding.

2.1.5 Further work on Data Comics

Based on those results, we conducted more data comics workshops and described a detailed methodology for teaching data comics and data-driven storytelling to university students [Wang et al., 2019a]. We believe teaching data comics in visualization classes and beyond is an easy and cheap means to allow students to think beyond the pure visual encoding in visualizations; to think about the audience, the message, the purpose, the importance of the communicating the data as well as why to chose a specific visual representation/metaphor. Potentially, this could help put the practice of data visualization into a larger (social) context and may help participants develop a critical perspective towards the construction of data presentations and surface the questions of why particular structures are employed in different contexts and how to successfully communicate.

Published as: *Wang, Z., Dingwall, H. and Bach, B., 2019, May. Teaching data visualization and storytelling with data comic workshops. In Extended abstracts of the 2019 CHI conference on human factors in computing systems (pp. 1-9).*

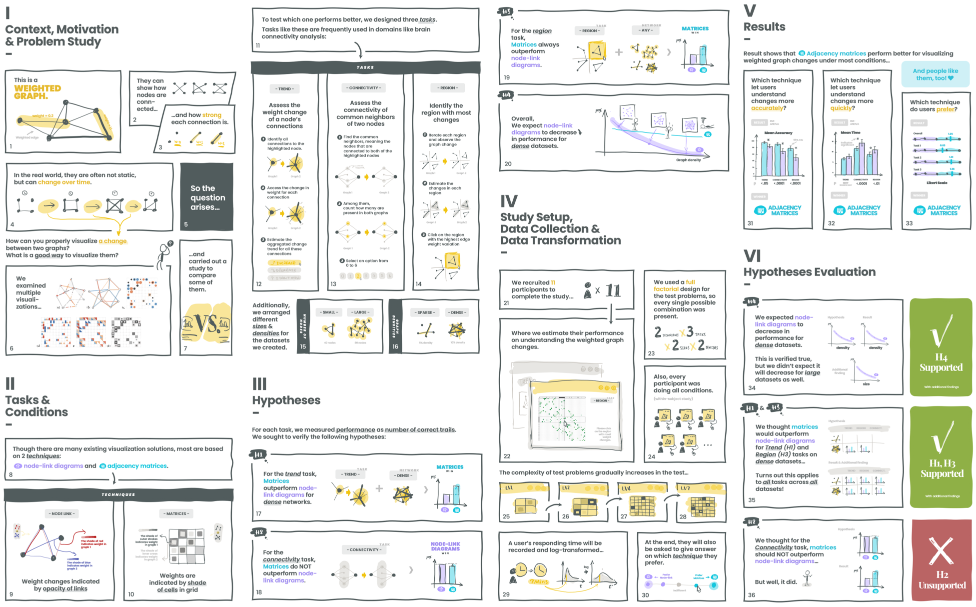


Figure 2.14: *Example of a data comic illustrating a study that compared two visualization techniques for weighted networks. Created by one of the authors through an iterative process, based on information provided in the original paper* [*Alper et al., 2013*].

We further investigated data comics for visualizing statistical data analysis in the context of controlled user studies in Human-Computer Interaction [Wang et al., 2020]. Similar to graph comics (Section 2.1.1), we identified design patterns and solutions based on our own practice in creating comics for various studies reported in the the HCI literature. The goal of the project was to understand how we can visualize and explain concepts related to data analysis and statistical concepts. That paper describes patterns for reporting on the context and motivation, tested conditions, hypotheses, tasks and dependent Variables, stimuli and materials, study setup, data transformations and checks, result presentation, and hypothesis evaluation (Figure 2.14). The full set of comics created in this project can be found online `https://statscomics.github.io`.

Our latest work on data comics focuses on data comics for educational settings and to explain concepts in visualization literacy [Boucher et al., 2023]. We also created two creation tools for data comics DataToon [Kim et al., 2019] and ComicScript [Wang et al., 2021], the latter of which is described in Chapter 3.

Summarizing Data Comics

As part of our investigation of data comics, we created a lot of data comics, defined our own comic creation processes and described patterns for graph comics, statistical comics, and general data comics, build workshops, and studied comics through co-design sessions, workshops, qualitative studies and quantitative controlled user studies. We argue that data comics offer a unique and powerful approach to data visualization, distinct from traditional infographics and other communication methods and that emerge as a powerful and adaptable tool for data storytelling, offering a unique blend of flexibility, structure, and narrative potential.

Here's a breakdown of the key points:

- **Structured storytelling:** Unlike infographics, data comics leverage panels

(explicit or implied) to break down information into manageable chunks and guide the reader's attention. Each panel serves as a self-contained world, allowing for frequent "clean breaks" and a dynamic presentation logic. This flexibility enables data comics to create unexpected connections, break their own rules, and present information in innovative ways.
- **Rich design space:** While data comics draw from established visual patterns, they offer considerable freedom in how data is presented and structured. This openness allows creators to leverage a wide range of principles to enhance understanding. The authors propose formalizing common comic layouts to identify storytelling strategies and explore new possibilities within this design space.
- **Beyond infographics:** Data comics offer a bridge between narrative and data visualization. The design patterns presented here not only aid in creating data comics, but can also influence the development of infographics and presentations by incorporating storytelling elements from comics. Blurring boundaries: Data comics exist at the intersection of various media, including infographics, illustrated texts, posters, dashboards, and even animations. This makes them a versatile tool for data-driven storytelling, allowing translation of information across different media. Notably, the low-tech nature of creating paper comics fosters a readily accessible space for generating ideas.
- **Pioneering a new medium:** This work represents the first structured exploration of data comics for data visualization. The proposed design patterns focus on the relationship between content and presentation, aiming to inspire communication strategies across various media that utilize time, space, and interactivity. This exploration paves the way for innovative storytelling formats, including interactive web comics, large-scale walk-through comics, and interactive presentations.
- As **challenges in creating and using data comics**—observed from our studies as well as during our own practice in creating many comics—we see their perceived complexity and association with art and illustration; their possible association with fiction and certain visual cultures; the possibility to overload a comic design with too many details; including complex visualizations; making a comic graphically compelling, visually consistent, and overaly engaging and informative.
- **Comics or elements of comics can blend** into other genres such as infographics, texts, slideshows, etc, introducing ideas such as panels, and sequence, layouts, allowing a designer to switch between the genres as they find it useful.

2.2 Dashboard Design Patterns and Genres

Best Paper Honorable Mention

Published as: *Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y. and Chen, M., 2022. Dashboard design patterns. IEEE transactions on visualization and computer graphics, 29(1), pp.342-352.*

In most of the cases *narrative visualization* and *data-driven storytelling* imply an author, creating a story, selecting the data, providing the data, and creating the visualization piece [Lee et al., 2015a] making sure information and data are clearly explained and presented. Clear explanation and presentation is also important in another form of visualization—dashboards—and which are used for *both* exploration and overview as well as presentation and communication of key insights so called *key-performance indicators* (KPIs).

Dashboards offer a curated lens through which people can view large and complex data sets *at a glance* [Few, 2006, Kitchin et al., 2015]. They combine visual representations and other graphical embellishments to provide layers of abstraction and simplification for numerous related data points, so that viewers get an overview of the most important or relevant information, in a time-efficient way. Their ability to provide insight at a glance has led to dashboards being widely used across many application domains, such as business [Few, 2006, Noonpakdee et al., 2018], nursing and hospitals [Mlaver et al., 2017, Khairat et al., 2018, Wilbanks and Langford, 2014, Bernard et al., 2018, Elshehaly et al., 2020, Wilbanks and Langford, 2014], public health [Lechner and Fruhling, 2014], learning analytics [Charleer et al., 2016], urban analytics [Lee et al., 2015b], personal analytics, energy [Goodwin et al., 2021] and more, summarized elsewhere [Sarikaya et al., 2019, Yigitbasioglu and Velcu, 2012, Few, 2006, Rasmussen et al., 2009]. These examples, designed mainly for domain experts, have since been complemented by dashboards for public health or political elections, designed for a more general audience and disseminated through news media [Zhang et al., 2021] or dedicated dashboard and tracker websites [Dong et al., 2020, Allison et al., 2021, Serrano et al., 2020].

There are many informative *high-level* guidelines on dashboard design, including advice on visual perception, reducing information load, the use of interaction, and visualization literacy [Few, 2006, Yigitbasioglu and Velcu, 2012, Elshehaly et al., 2020, Rasmussen et al., 2009]. Despite this, we know little about *effective* and *applicable* dashboard design, and about how to support rapid dashboard design. Like comic design, dashboard design is admittedly not straightforward: designers have access to numerous data streams, which they can process, abstract, or simplify as they see fit; they have a wide range of visual representations at their disposal; and they can structure and present these visualizations in numerous ways, to take advantage of the large screens on which they are viewed (vs. individual plots that make more economic use of space). These choices can be overwhelming, so there is a timely need for guidance about dashboard design—especially as dashboards are increasingly being designed for a wider non-expert audience by designers without a background in visualization or interface design.

Outline and Methods

Inspired by our analytical work on data comics, this section presents design patterns for *effective* and *applicable* dashboard design to support rapid dashboard design. As in our work with comics, we analyzed the visual design of 144 dashboards. By coding 144 dashboards found in the internet and in a previous collection of dashboards [Sarikaya et al., 2019], we formalized 42 design patterns (Fig. 2.15) and 8 genres for dashboards. Like our patterns for data comics, they capture key solutions for dashboard design. Likewise, we ran a two-week dashboard design workshop with 23 participants to understand the usefulness and applicability of these patterns for dashboard design. A detailed description of all our design patterns, design guidance and the workshop are online `https://dashboarddesignpatterns.github.io`.

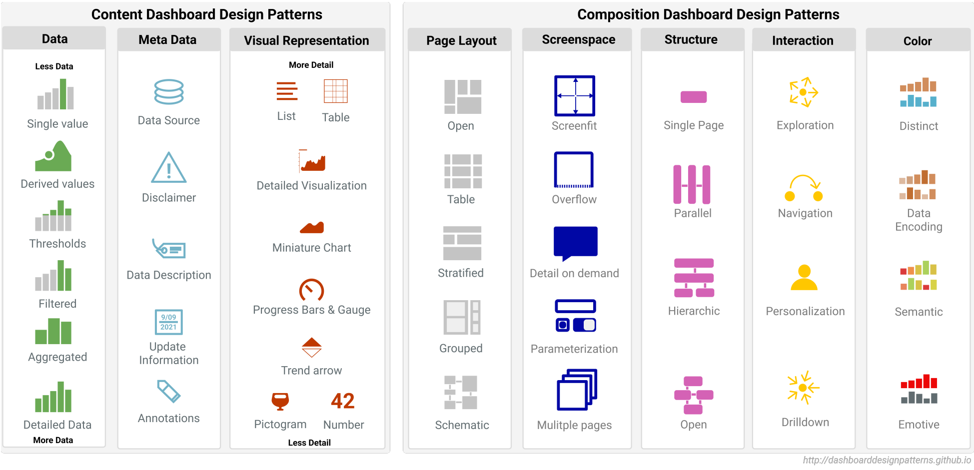


Figure 2.15: *Design patterns for dashboards identified from 144 dashboards.*

2.2.1 Design Patterns & Genres

We group the patterns into two high-level groups: content patterns describing the dashboard content and composition patterns that describe how the content is compositioned into a dashboard (Figure 2.15).

Content Design Patterns

The Content of a dashboard is made up of individual dashboard elements, the crucial 'ingredients' relating to the data and its presentation. We identified three groups of design patterns relating to content while disregarding any visual components used purely for decoration or embellishment, e.g., illustrative pictures, dividers, borders. The types of content patterns we found can be grouped as follows:

- **data information patterns** identify the types of information presented and the extent of abstraction use: a single value directly from the data (e.g., a current value), a derived value (e.g., mean, sum of something), thresholds (e.g., 3% over the suggested value), a filtered subset of a data (e.g., last 1 week), aggregated data (monthly averages), or the entire detailed data.
- **meta information patterns** capture additional information used to provide context and explanation: data source, disclaimer (e.g., 7 days rolling averages), data description, update (e.g., last update this morning 4am), annotations (e.g., peak reached on May 20th).
- **Visual representations patterns** show data and information is encoded visually in the dashboard, ranging from more detailed to less detailed and more abstract: lists and tables, detailed visualizations, miniature charts (usually missing labels and descriptions), progress bars and gauges, trend arrows, pictograms and numbers.

Composition Design Patterns

Dashboards show multiple information elements and their structure and layout on a page are meaningful design decisions. We identified five aspects of composition

- **Page layout patterns** describe how *widgets* are laid out and, often, implicitly grouped in a dashboard: open, table layout with an inherent structure, stratified with a clear top-down hierarchy, grouped by topic or any other relationship, and schematic, e.g., a spatial layout plan for a plant.

- **Screenspace patterns** describe solutions used to fit content onto a single screen: either making everything fit the screen, but if this is not possible then, use an overflow and zoom to navigate, show details on demand, e.g, through tooltips, allow for paramterization, or spread content across multiple pages.
- **Structural patterns** capture relations between multiple pages of a dashboard, if there are multiple pages: the structure can be parallel with pages showing similar content and structure (e.g., one page per topic, data value, spatial unit, etc.), hierarchic (e.g., summary information, detailed information), or entirely open.
- **Interaction patterns** describes common interaction approaches found within dashboards for *exploration* (what is in my dashboard), navigation (how to get around), personalization, and drilldown (what are the details).
- **Color** can be used for different purposes in dashboards and comes with cultural implications, we examined the use of color at dashboard-level—i.e., across multiple widgets and views. Color can be used to differentiate data sets and attributes (e.g., category A vs. category B), encode data values (e.g,. high vs. low), have semantic meanings (e.g, extremes, good vs. bad), or emotive to create a certain mood related to the topic at hand (e.g., Covid, casualties).

2.2.2 Genres

Genres group dashboard with shared characteristics and common design patterns. These can be related to other genres of information visualization, such as Multiple Coordinated View systems [Roberts, 2007] or infographics. Our review and pattern collection help us discuss design tradeoffs and discuss possible design frameworks. Percentages in brackets relate to the frequency among our 144 dashboards.

Static Dashboard (21%) capture the traditional notion of a dashboard as a static (*no-interaction*), single page, screenfit display of information. Static dashboards often feature concise information and representation such as single-values and derived values, miniature charts, arrows, and numbers 42.

Magazine Dashboards (2%) were created by news agencies and similar media outlets. These dashboards are found as integral part of journalistic articles and resemble visualizations of the *magazine* genre [Segel and Heer, 2010]. The text provided alongside visualizations goes beyond basic meta information to provide additional commentary and storytelling about the data.

Analytic dashboards (73%) is what Few would call a *Faceted Analytic Display* [Few and Edge, 2007]. We see parallels to the concept of Coordinated Multiple Views (CMV) [Roberts, 2007], while there are clear differences between both concepts: by definition, the focus in CMVs is on coordination and individual views react to interaction in another view to give complementary views for the task.

Infographic Dashboards (6%) are similar to infographics that include decorative graphical elements and other non-data ink shown alongside data representations. They are meant to provide more background and context and narration about the data, i.e., embedding the dashboard into a narrative context. Similar to magazine dashboards, they use non-data media to annotate and embellish data.

Repository Dashboards (17%) list a multitude of charts across multiple pages, with *overflow* structure that impedes proper analytics, i.e., comparing views. Common examples include repositories from governmental and academic institutions, e.g., for open data and data on Covid statistics. Often data can be downloaded.

Embedded Mini Dashboards (4%) were found to be embedded into other applications such as news websites. These concise dashboards only occupied a small area on screen and usually come with concise visualizations (42). Mini dashboards requires interactive features for navigation and to parameterize the content quickly.

2.2.3 Workshop

We ran an online dashboard design workshop to help us scrutinize and refine the design pattern collection as well as our discussion on design trade-offs. In the workshop, the participants applied the patterns to their own dashboard projects. The workshop ran for 2 weeks and was open to everyone with a dashboard design task at hand. It aimed to guide people through the main stages and decisions of a design process while using the design patterns for ideation and discussion. Rather than implementing a dashboard in a tool such as Tableau or Power BI, the course was intended to finish with interactive mockups in Figma. We had 23 participants in total from diverse backgrounds, including data science, medicine, psychology, economics, bioinformatics, design, etc., including a mix of industry and academic partners. Each participant worked on their own dashboard within a context they defined themselves and which was relevant to their work.

Reflections on the Design Process

We obtained many quotes from our questionnaire thatnicely capture participants' reflections and which do not need much commenting:

- *Start with simple designs that highlight just the most important data, then add to this as you go; this will help to avoid just putting everything you can on the dashboard, keeping it focused. (This might not apply to repository style dashboards).*
- *I learned not to try to ask too many questions for one graph. One graph can give several different answers, but the question should just be one question.*
- *Simpler is often better when it comes to charts: bar charts, line charts, and tables are often clearer and easier to read than their more fancy looking counterparts.*
- *If you try to simplify [your design] too much, you risk imposing your own story.*
- *Use filters, menus, buttons, and parameters to reduce the number of static visuals shown at any one time [to] keep things readable.*
- *In medicine we often hand out 15 leaflets and tell patients [...] to come back with a decision [...] but I have not given them information about how to* use *[the information] and how to* decide.

Pattern and Workshop Utility

Participants valued the individual feedback onto their designs as well as discussing other participants' design. They also valued the framework, guidance, and community the workshop provided.

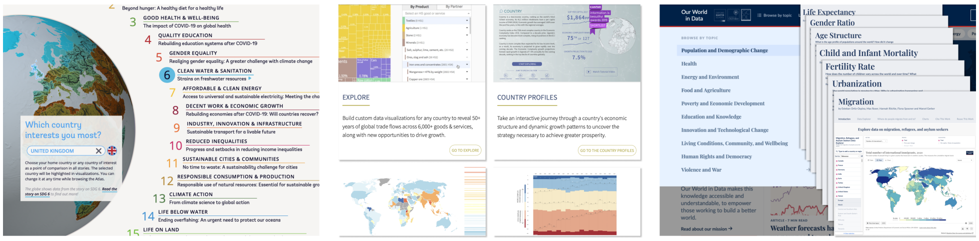
- *[Design patterns and their terminology] limits down [the design complexity] from endless design options to 2-3 good candidates [...] I think it will be really useful for co-creation or kick-off meetings [...] so everyone is very clear on what the key elements are and what we're trying to achieve right from the start etc*
- *[I used the patterns] first as guidance for designing [then] I did any changes with the design principles in mind. In the end, I used [the patterns] as a 'check-list' to review my design decisions.*
- *If you walk through each of those patterns/categories, you basically end up with a written plan for your dashboard which is clear to everyone e.g. 'I'm planning to build an analytic*

dashboard. Because of XYZ factors, we're going to use a paginated design and allow user interaction etc.'

- *I think it's very useful to now be able to ask things like 'is this a static dashboard? is this an analytic dashboard?'. Or on structure, things like 'is a hierarchical structure the best? [...] it makes it a bit easier to envision and make choices.*
- *in my experience as a data analyst, building dashboards is something I've had to pick up and learn myself with not much outside training or workshops [...], that's a somewhat common experience for analysts.*

2.2.4 Summary

- **Filling a design gap:** While dashboards share similarities with infographics, data comics, and other visualizations, they often lack specific design guidance. This work introduces the first set of practical design patterns tailored to dashboards. These patterns address key aspects of dashboard design, including data abstraction, screen organization, element grouping, relationship visualization, and interactive exploration.
- **Benefits of design patterns:** These patterns offer several advantages. They establish a common vocabulary for discussing dashboards, highlight their unique characteristics, and position them within the broader visualization landscape. By fostering connections to infographics and other forms, the patterns encourage cross-pollination of design ideas.
- **Defining dashboards by function:** We argue against a rigid definition of "dashboard" and propose viewing it as an idea to communicate information and within a visualization user interface with the need for overview, control, and conciseness for decision making. Those ideas can be integrated into variour other forms of visualizations as shown by our genres. These genres represent common combinations of patterns, offering designers a practical framework for creating dashboards.
- **Building a shared resource:** The paper contributes a substantial dashboard corpus, a collection of real-world examples. This corpus, similar to established resources for physical visualizations or data storytelling, provides a valuable foundation for future research, design practice, and discussions on dashboards.
- **The design process:** In the future, we need more structured design processes to guide designers in making effective trade-offs. We outline a potential design process (in the original paper, omitted here for clarity) that integrates the proposed design patterns and focuses on mid-level design decisions. This iterative process assumes prior requirement analysis that defines users, tasks, and datasets.
- **Tool considerations:** Likewise, user feedback from the workshop indicates that existing dashboard design tools facilitate easy iteration and prototyping. However, no single tool emerged as clearly superior, and many lacked support for specific design patterns or optimal use of screen space. This suggests a need for more advanced tools that can cater to the full range of dashboard design considerations.
- **Future work** on dashboards can include means for personalization, a better inclusion of dashboards and narration for, e.g., public storytelling and information, or dashboard as part of immersive mixed reality environ-



Example 1: ATLAS of Sustainable Development Goals 2023 **Example 2**: Atlas of Economic Complexity **Example 3**: Our World in Data

Figure 2.16: *Three visualization atlas examples.*

ments where access to the actual environment, e.g., machines, is required.

2.3 Visualization Atlases

Published as: *Wang, J., **Bach, B.**, Hinrichs, U., 2022. Visualization Atlases: Opening Complex Topics through Data, Visualization and Narration. 2024 IEEE transactions on visualization and computer graphics. UNDER SUBMISSION*

The third and last visualization format I want to describe in detail in this thesis are Visualization Atlases. Like data comics and dashboards, they present a specific way to communicating and exploring data through *many* data visualizations, assembled in a very specific way and come with their own set of superpowers and design options. Like comics and dashboards, atlases lack detailed description and like data comics have not been formally described yet.

Visualization atlases are an emerging concept in the internet, referring to polished web-based projects that use data visualizations to make complicated contemporary problems accessible. Projects like the *Atlas of Economic Complexity* [The Growth Lab at Harvard University,], *Our World in Data* [Roser, nd], or the *Atlas of Sustainable Development Goals* [Pirlea et al., 2023] (Figure 2.16) summarize and explain large and multidimensional datasets and their key insights about global challenges, such as climate change, sustainable development, economics, artificial intelligence, and cultural discoveries. Many of these projects are created by large organizations such as the World Bank and the UN, or research labs at prestigious institutions, with a mission to open data, methodological transparency, and education. These projects target a wide range of audiences, by making those data sets more accessible through the use of highly-curated interactive data visualizations and contextual explanations embedded in well-designed professional websites.

While the majority of these projects call themselves "atlas", "observatory", or "index", there exists no common definition or any shared understanding of the characteristics of these atlases. Yet, visualization atlases are a potentially powerful new format for data-driven and evidence-based dissemination of information of global—and often humanitarian—concern; what are the motivations behind designing an atlas and who are its intended audiences? There is no account of the mechanisms that current atlases use to address their mission and any essential dimensions to consider when designing a new atlas. There is also no understanding on how to inform atlas design for other topics and what we can learn from the processes and decisions involved in the creation of existing atlases.

For the sake of this project, we decided to term this new class of projects *visualization atlases*, given their perceived analogy to the traditional use of the term "atlas".[1]

[1]The term atlas, originating in Greek mythology, has been widely used since the 16th century as a collection of maps, initially in geography [Ortelius, 1570, Maps, 2016], and further extended to various domains such as history [Tan, 1982], language [Drexler and Wimpissinger, 1934], and science [Borner, 2010].

As an outcome of this work, we define a **visualization atlas** as

"a compendium of (web) pages aimed at explaining and supporting exploration of data about a dedicated topic through data, visualizations and narration"

Outline and methods

This definition is based on (a) structured survey of 33 visualization atlases, (b) resulting in definitions and design patterns, (c) interviews with eight designers and authors involved in the creation of some of these atlas, and (d) 5 high-level visualization atlas genres that describe usage scenarios while informing lower-level design decisions. All materials including atlas cases and detailed descriptions of design patterns can be found online: `https://vis-atlas.github.io`.

2.3.1 Example: Atlas of Sustainable Development Goals 2023

For illustration, this section details one atlas example—another detailed example is found in the paper.

The Atlas of Sustainable Development Goals 2023 [Pirlea et al., 2023] (SDG2023) shows the latest insights into global progress and challenges towards achieving the 17 SDGs [of Economic and Affairs, 2023] defined by the United Nations in 2015. The project was commissioned by the World Bank as sequel to their previous atlases in 2017 [World Bank, 2017], 2018 [World Bank, 2018], and 2020 [Pirlea et al., 2020]. Data are mainly from World Development Indicator compiled by the World Bank about global development and living quality. Its latest 2023 issue is the most sophisticated in terms of interactive visualizations and storytelling, created by a team of data experts, editors, and visualization designers.

SDG2023 consists of 17 web pages, each dedicated to an individual goal and starting with a brief description of the respective goal and its sub-targets. The hero section features a rotating globe, visualizing a selected metric about each country, e.g., the number of people in poverty (Goal1: No Poverty), followed by a highly curated and clearly written article with visualizations about detailed data insights for reaching this goal. Visualizations include a huge variety of well-designed linecharts, treemaps, small multiples, Mekko charts and others. Some visualizations are interactive, showing tooltips on demand. Many of the visualizations are explained through scollytelling, using piece-meal explanations and animations. Each article also features scientific references, cited throughout the text, as well as the ability to download the data, visualizations, and implementation code.

Each of the pages has its own color identity and is accessible through a condensed sidebar menu as well as through a landing that shows up when first visiting the atlas. The landing page features a central list of the titles of all 17 sustainable development goals and the same gently rotating globe as that on the individual pages. The landing page shows a small pop-up inviting people to search for a country of interest, which will be consistently highlighted across all visualizations.

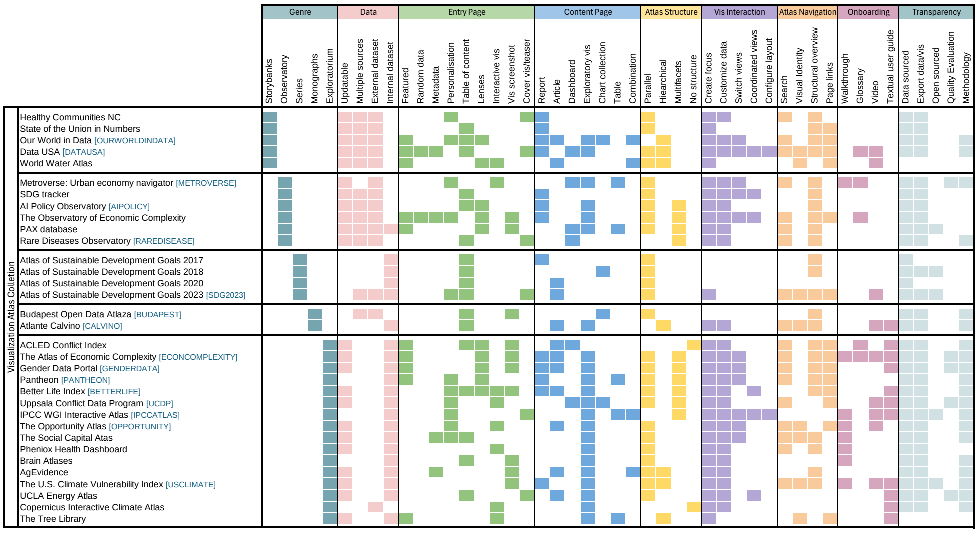


Figure 2.17: *Complete coding of 33 visualization atlases, including seven design dimensions, contextual dimension of data, and genres.*

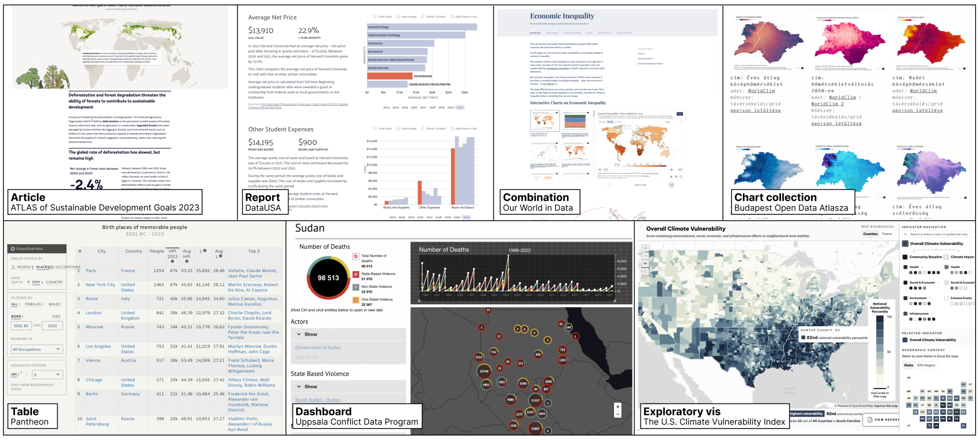


Figure 2.18: *Examples of content pages styles.*

2.3.2 Visualization Atlas Design Patterns

Figure 2.17 shows all atlases (rows) and their design patterns (columns). Colors differentiate design pattern categories. Patterns have been identified though rigorous double coding process and several iterations and discussions with all authors.

Visualization atlases are made out of *pages*, a notion that roughly corresponds to a webpage. Links on one page can lead to another. Across atlases, we identified two types of pages—content pages and entry pages—that serve different purposes. Across entry and content design pages, we observed seven types of design patterns, detailed in the original paper but only briefly outlined below.

- A visualization atlas **content page** can include visualizations, textual descriptions, titles, author, illustrations, links to further information, etc. As centerpieces in every atlas, content pages not only provide access and allow people to explore the data presented, but also explanations. Content pages come in different *styles*: report, dashboard, exploratory visualization, chart collection, table (Figure 2.18).
- **Entry pages** provide an overview to a wider topic (or subtopic) in the atlas, deliver respective explanations and content, and, most importantly, guide the navigation to specific content pages. All atlases in our collection contain at least one entry

page—the atlas landing page. Unlike content pages that have specific styles, entry pages use combinations of textual and visual interactive *components* such as table of contents, featured pages, meta data, screenshots, etc.

- **Atlas structure patterns—**The atlas structure describes how content pages are organized and linked. The page organization can combine different structure options; 12 atlases (37%) in our collection combine two structures: parallel, hierarchical, multiple.
- **Atlas navigation** across an atlas guides the user to individual pages, help orient which page they are currently seeing, and may suggest which parts of the atlas they have not yet seen through patterns such as search, visual identities, structural overview, and page links. This is supported through search boxes, menus, individual visual entire per page, directed page links.
- **Visualization Interaction Design Patterns** interaction within a content page to explore data and visualizations such as creating focus, switching views, exploring data with coordinated views, or customize the data shown.
- **Data Transparency Design Patterns** describe how an atlas' data, analysis, and visualizations are made transparent, e.g,. through disclosing data sources, offering data experts, making visualization and analysis code open, provide quality evaluations, and descriptions of the analysis methodologies.
- **Onboarding Design Patterns** describe help resources for visitors to understand an atlas' context, use, and visualizations: walkthroughs, glossaries, instruction videos, or textual user guides.

Designer Interviews

We interviewed eight visualization experts ("atlas creators") who were directly involved in the design of seven visualization atlases. Creators' professional backgrounds included 5 visualization designers, 1 full-stack visualization engineers, and 1 data journalist. Our main findings in a nutshell are:

- **What was your motivation & goals for creating a visualization atlases?**
  - Promote Selected Datasets for Exploration & for Taking Action.
  - Publish Insights & Stories.
  - Promote Topics in a Systematic, Data-Driven Way.
- **What are the intended audiences of your visualization atlas?**
  - Internal & External Experts
  - Policy makers.
  - The General Public.
- **What do you think are key characteristics of atlases?**
  - Defined topic and breadth of perspectives and Data Sources.
  - Visual curation
  - Debating comprehensiveness
  - Provide global and geospatial Perspectives
  - Offer up-to-date data and information.
  - Provide some degree of interactivity, but not too much

Atlas Genres

The design patterns described above provide low-level design solutions and common practices, while our interviews provide very detailed information about the rationales behind atlas design. Inof a practical middle ground, we now describe atlas *genres*, inspired by Segel and Heer's [Segel and Heer, 2010] genres and the genres for dashboards (Section 2.2). Likewise, those genres can serve as templates for specific purpose and scenarios while informing design decisions such as the choice of design patterns.

- **Storybanks** aim to promote a defined, yet broad topic in a systematic, data-driven way. They are driven by multiple data sources, often external, that are updated regularly, use a large collection of *articles* or *reports* followed by an extendable structure of *hierarchy/multiple facets* plus *parallel* pages, and provide multiple entry points for visitors to pick-up recent highlights. However, storybanks have limited onboarding features, which, combined with their breadth and amount of content, can make navigation more difficult.
- **Observatories** share similar design features as storybanks but the topics are more narrowly defined and systematically tracked based on key performance indicators, e.g., indicators being sourced externally, processed, and presented together by-country and topic to track the latest updates in AI policy.
- **Series** are published in more or less regular intervals. The visualization atlas series focuses on a high-level theme and each issue will not be updated systematically. Consequently, the overall style, featured pages can be unique to each atlas issue and evolve over time.
- **Monographs** are visualization atlases focusing on highly specific topics with the purpose of promoting these for exploration and taking action. They are typically finished pieces without regular updates, are usually more deep and feature detailed stories and insights.
- **Genre: Exploratoria** promote the active and open-ended exploration of particular topics by making extensive use of interactivity. They typically feature at least one *exploratory vis* page as the core exploration tool, sometimes, but not always with supporting pages of either *articles* or *reports* to explain different perspectives on the topic in *parallel*. There are also often multiple entry points and navigation strategies which are communicated through a range of onboarding strategies.

Defining visualization atlases

Based on all the patterns, genres, and interviews, we are now able to more precisely define visualization atlases using three key characteristics.

- **Topics**—Visualization atlases provide data-driven views on complex topics, targeting a wide range of audiences. They focus on specific topics, rather than being an open repository for data and information. Visualization atlases report and analyze typically multiple data sources around the given topic. They provide context about the topics, mixed with data, and analyses.
- **Curation**—Visualization atlases are highly curated in terms of content, data and presentation. Content curation concerns the choice and framing of the topic, the choice of data, and the framing of stories. For example, in some atlases, certain data and related stories are deliberately excluded to maintain scope, to keep the content accessible to large audiences, but also to reflect on the nature, responsibilities and agendas of the organizations involved.
- **Visualization**—Visualization atlases rely data visualization for explanation, exploration, and illustration. Data visualizations serve as the main medium to engage with the topic and data presented in a visualization atlas. Visualizations can serve a variety of purposes in a visualization atlas. They can highlight specific

messages, invite exploration, attract curiosity, or simply function as illustrations on entry pages to communicate the overall mission and topic of an atlas.

Summary

- **An emerging genre:** We believe visualization atlases are an emerging new form for communication, data-driven storytelling exploration. They extend our current reportoire of genres and formats for these purposes, offering a structured middle-ground that combines exploration and explanation for large, complex, and urgent global topics understood through data—providing access to data, promoting knowledge, educating people, supporting further analyses, make data open. We believe they can become a crucial tool for a data-literate society and a culture of data-driven (inclusive) decision making, balancing access to data and critical engagement. In that, atlases might be particularly suited to address wicked problems.
- **Comparison to other genres:** While atlases share many characteristics with other forms of visualizations—a combination of text, image and other media; they are accessible through the web; they are curated and often involve data-driven storytelling; they can provide interactive visualizations, they can vary to the extend they are reader-driven or author-driven [Segel and Heer, 2010]. However, they are also different from these other forms or genres of visualization, analytic interfaces, dashboards [Bach et al., 2022b] or narrative visualizations [Segel and Heer, 2010] such as articles [Hao et al., 2024], comics [Bach et al., 2018a], or videos [Amini et al., 2015]: visualization atlases are high-level structures that incorporate and combine these forms/genres in different ways, depending on the topic, goal and audience of the atlas.
- **Atlas design:** Design patterns and genres can inform the design of future atlas projects. They can inform design card decks [Roy and Warren, 2019, Bach et al., 2018a], design activities [Huron et al., 2020]. Future research could look into specific toolkits and web-design platforms for making atlas design, development, and maintenance easier. We hope this paper inspires future research to imagine new genres of visualization atlases that may also expand the design dimensions and patterns presented here.
- **Future work:** We did not encounter visualization atlases incorporating visualizations such as data comics [Bach et al., 2018a, Wang et al., 2021], data-gifs [Shu et al., 2020], or data videos [Amini et al., 2015, Shen et al., 2023]. These and other forms/genres of data representation (e.g., sonification or data physicalization [Jansen et al., 2015]), as well as related approaches (e.g., interactive notebooks, explainable explorables [Victor, 2011], immersive technology [Isenberg et al., 2018], or streaming [Zhao and Elmqvist, 2022]) offer a vast range of design opportunities for visualization atlases.
- **Critique:** As all curated data and visualization work, atlases can introduce biases from the author side and imply power relationships. The amount of content and interaction required could be overwhelming to users. Eventually, care should be taken to provide disclose analysis methods and make the atlas curation transparent. We identify those issues as important areas for future work.

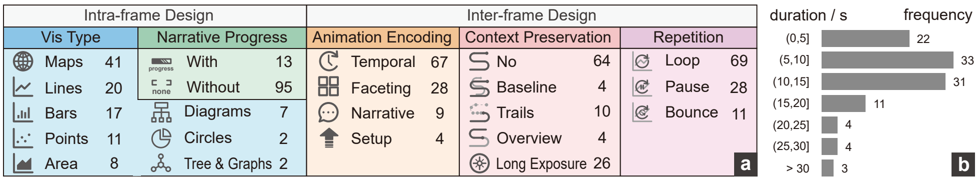


Figure 2.19: *An overview of the design space (a) and the duration distribution (b) of 108 data-GIFs in our survey.*

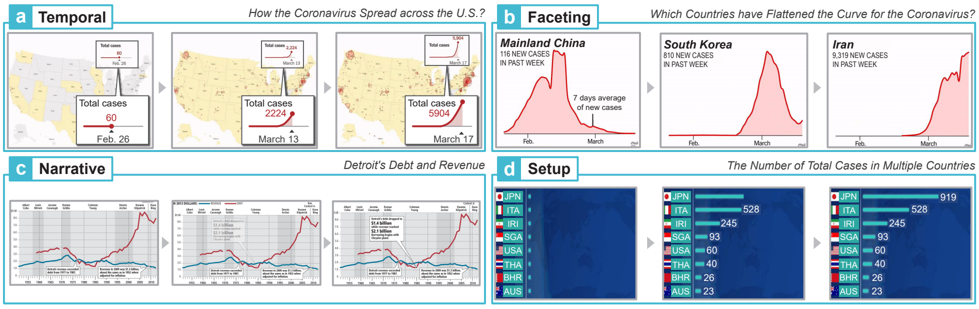


Figure 2.20: *Four data-GIFs with tailored keyframes demonstrating four animation types of data-GIFs. Text is enlarged to make the figure clear. (a)* ***Temporal****: It shows the evolvement of the coronavirus in U.S. over time. Shadowed callouts are added by paper authors to make the line chart clear. (b)* ***Faceting****: It presents the curves for the coronavirus of several countries one by one. (c)* ***Narrative****: It narrates a story by building up the visualization scene. (d)* ***Setup****: It animates the creation of a bar chart. Original GIFs are attached in the supplementary materials.*

2.4 Other formats: DataGIFs and Data Articles

To complete the tour of visualization formats, I helped investigate DataGIFs and Data articles, both presenting yet more complementary designs and options to data comics, atlases, and dashboards.

DataGifs

Published as: *Shu, X., Wu, A., Tang, J.,* ***Bach, B.****, Wu, Y. and Qu, H., 2020. What makes a data-GIF understandable?. IEEE Transactions on Visualization and Computer Graphics, 27(2), pp.1492-1502*

GIFs are enjoying increasing popularity on social media as a format for data-driven storytelling with visualization; simple visual messages are embedded in short animations that usually last less than 15 seconds and are played in automatic repetition. In this paper, we ask the question, *"What makes a data-GIF understandable?"* We have still little knowledge about the design factors and principles for "data-GIFs". To close this gap, we provide results from semi-structured interviews and an online study with a total of 118 participants investigating the impact of design decisions on the understandability of data-GIFs. The study and our consequent analysis are informed by a systematic review and structured design space of 108 data-GIFs that we found online (Figure 2.19). Our results show the impact of design dimensions from our design space such as animation encoding (Figure 2.20), context preservation, or repetition on viewers' understanding of the GIF's core message. The paper concludes with a list of suggestions for creating more effective Data-GIFs.

Data Articles

Best Paper Honorable Mention

Published as: *Hao, S., Wang, Z.,* ***Bach, B.*** *and Pschetz, L., 2024, May. Design Patterns for Data-Driven News Articles. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1-16).*

Data-driven articles (called magazine genre by Segel and Heer [Segel and Heer, 2010]) have gained attention from researchers and practitioners in the field of data journalism. We present a classification of data-driven news articles and related design patterns defined to describe their visual and textual components. Through the analysis of 162 data-driven news articles collected from news media, we identified five

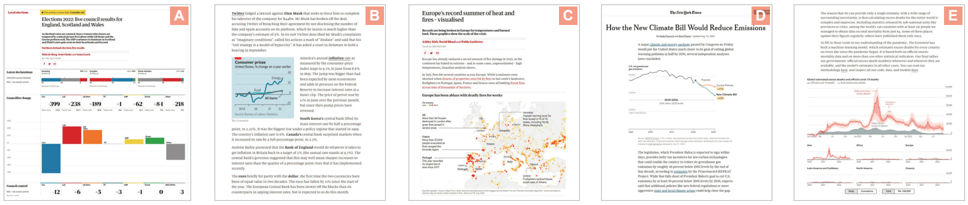


Figure 2.21: *Examples of different types of data-driven articles: (A) Quick update case Elections 2022(by the Guaridian); (B) Briefing case Business(by The Economist); (C) Chart Description case Europe's record summer of heat and fires (by the guaridian); (D) Investigation case How the New Climate Bill Would Reduce Emissions(by the New York Times); (E) In-Depth investigation case The pandemic's true death toll(by The Economist).*

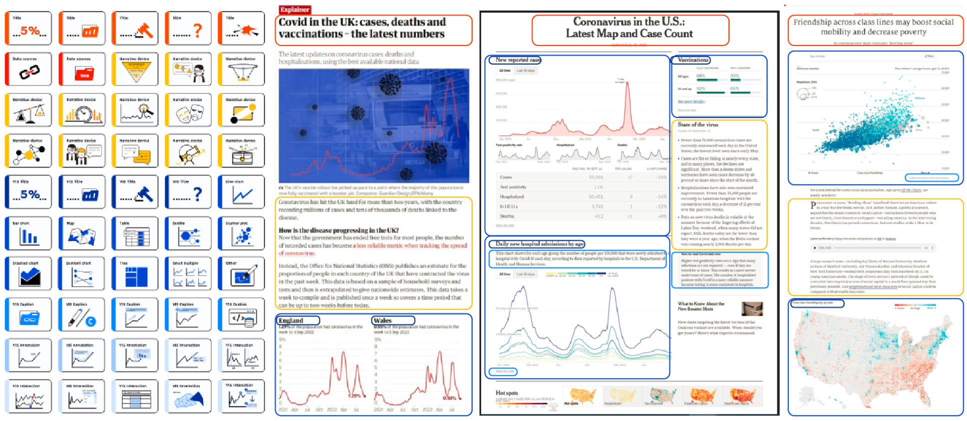


Figure 2.22: *Summary of our design patterns for data-driven article components and some cases of data-driven articles.*

types of articles based on the level of data involvement and narrative complexity: *Quick Update, Briefing, Chart Description, Investigation, and In-depth Investigation* (Figure 2.21). We then identified 72 design patterns (Figure 2.22) to understand and construct data-driven news articles. To evaluate this approach, we conducted workshops with 23 students from journalism, design, and sociology who were newly introduced to the subject. Our findings suggest that our approach can be used as an out-of-box framework for the formulation of plans and consideration of details in the workfow of data-driven news creation.

2.5 Chapter Summary

This chapter has described five possible forms to present and communicate data: data comics, visualization atlases, dashboards, dataGIFs and data articles. Data comics was an entirely new approach with a wide open design space and that we consequently studied most extensively through hands-on exploration, design patterns, and qualitative and quantitative user studies. Dashboards and data articles, while widely known were missing a hands-on framework to understand their design decisions and characteristic parts. Visualization atlases and DataGifs can also be defined as a new concept that we described for the first time systematically. All those concepts and ideas extend our understanding of the landscape of data-driven storytelling and the presentation of data.

Comparing and describing visualizations forms

Each form has specific affordances to solve a specific task, engage an audience, tell stories, and make data accessible. Comics might be preferred for wide and diverse audiences as they require little training and can make anything explicit. Comics can be a great choice when explaining temporal data by using individual panels to show progression over time. Comics can also potentially be drawn through a wide range of visual styles, making them more or less attractive to specific audiences. Dashboards are those concise representations of key-performance indicators. They are less about storytelling as about clear communication, visual conciseness, clear visual accessibility.

While data articles, data comics, and DataGIFs are part of Segel and Heer's seven genres [Segel and Heer, 2010], visualization atlases are higher-level constructs. Atlases are an umbrella concept that can can include any of the other visualization forms such as comics, dashboards, or any of the other genres. The focus of atlases are much wider and deeper at the same time; they aim for both explanation (presentation, communication, storytelling, reading) as well as exploration through navigating between pages, interactively explore visualizations, as well provide detailed context. In that case any of the discussed forms are complementary. Yet, often forms blend into each other and the boundaries where something stops to be a comic, but starts to be an infographic or a dashboard, or an interactive visualization atlas can be blurry. I think we should embrace this diversity and richness and use those forms as high-level ideas, genres, and get inspiration from the design patterns and how each form targets its intended purpose of informing, entertaining, explaining, exploring, data and information.

Ideally, we could compare those different forms through a more structured framework, providing a starting points for deciding when to use a specific form in a given context. Possible criteria for such a structured juxtaposition could be:

- how much **expression** a form allows (e.g., the visual style, the suitedness to explain temporal or spatial data, the type of visualizations it supported better, what types of insights and messages are best supported etc),
- which **narrative styles** it supports (e.g., linear, non-linear, reader-driven, author-driven)
- the **media and material** a form supports or dictates especially with respect to accessability (e.g., paper, web, large-walls, mobile phones),
- whether a form allows for interaction and updated data
- effectiveness in **specific scenarios** (e.g., learning, visiting a museum, discussing with a group, consuming news on a mobile phone)
- target **audience** (e.g., pupils, decision makers, museum visitors)
- how **complicated it is to create** a form.

I believe there is still a huge design space to explore for novel visualization forms and data-driven storytelling in different media such as immersive environments (Augmented and Virtual Reality), data physicalization, mobile phones, exhibitions, large wall, and possibly any other way humans encounter and interact with visualizations.

Design and Creation

This chapter has reported on design patterns for graph comics, data comics (in general), dashboards, and visualization atlases. Design patterns are incredibly effective conceptual tools to describe design solutions and to inspire creation. In fact, it is great to see that more and and more papers are published describe design patterns for visualization (e.g.,[Kim et al., 2021, Lee et al., 2023]). While no collection of design patterns will most likely ever be complete, I think design patterns are a very hands-on tool to connect application, theoretic research, and education.

While design patterns are conceptual tools, the creation of visualization forms such as data comics, visualization atlases, dashboards, or dataGIFs is *significantly* more complex and laborious and are less supported through existing visualization tools and libraries than 'traditional' single visualizations. The next chapter (Chapter 3) will talk about tools to create visualizations while the question about skills and methodologies for creation will be discussed in Chapter 4.

Chapter 3

Tools for data-driven storytelling and visualization design

Tools are essential to most human tasks, whether this is finding and preparing food, traveling over long distances or recording, treating and communicating information. Designing visualizations is also largely supported by an every growing body of software tools targeted at different audiences, technologies and tasks covering many steps from data analysis to visualization design to communication.

As just one example, Andy Kirk's website lists **367 tools** related to data visualization design on his website [Kirk,] (Dec 2020). The range of tools include commercial as well as open-source and research prototypes, coding libraries (e.g., D3.js [Bostock et al., 2011a]) and coding environments (e.g., Processing [Reas and Fry, 2006]), graph libraries (e.g., ECharts [Apache ECharts,]), interfaces for data exploration (e.g., Tableau [Tableau,]), tools for storytelling and annotation (e.g., Lyra [Satyanarayan and Heer, 2014b]), design plug-ins (e.g., the Figma Charts Infographic UI Toolkit [Figma,]), interactive visualization design (e.g., Data Illustrator [Data Illustrator,]), all-round workflows (e.g., Flourish [Flourish,]), tools for immersive environments (e.g., DXR [DXR,]) or data physicalizations, collaboration (e.g., Many Eyes [Viegas et al., 2007]), as well as support tools for helping with the generation of appropriate color schemes (e.g., ColorBrewer [Brewer and Harrower,]). Some tools created visualization from templates [Mauri et al., 2017] others are designed for more expressive and free-form visualization designs [Liu et al., 2018, Ren et al., 2018, Kim et al., 2016]. For data-driven storytelling, tools can create infographics [Zhang et al., 2020], personalized glyhs [Xia et al., 2018] or animated video clips [Amini et al., 2016]. Only few tools provide interactivity through, e.g., navigating a slideshow [Satyanarayan and Heer, 2014a], showing annotations [Ren et al., 2017], or scrollytelling [Flourish, , 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.

In my research, I have worked on many interactive visualization tools to facilitate visual exploration, e.g., for network analysis [Bach et al., 2013, Bach et al., 2015a, Bach et al., 2015c, Bach et al., 2014]. I have also been involved in building toolkits for immersive visualization [DXR, , Cordeil et al., 2019, Chen et al., 2020], data comics [Kim et al., 2019, Wang et al., 2021], tools for visualizing large data [Lekschas et al., 2017, Lekschas et al., 2019, Lekschas et al., 2020], tools for authoring responsive geographic visualizations [Schöttler et al., 2024] as well as an interface for annotating visualizations [Kauer et al., 2024].

In this chapter, I talk about tools to support data-driven storytelling, exploration, and visualization design and focus on three representative tools within the context of data comics and network visualization. The audiences of those tools vary slightly, ranging from novice analysts facing a large amount of data to storytellers with programming experience, as well as visualization developers. Each tool is representative for a possible scenarios and need in engaging with visualization: exploring large amounts of data, craft data-driven stories, and design and implement visualizations.

1. **Interactive data comics** (Section 3.1) are a direct extension to our work on data comics (Section 2.1), providing interaction techniques for data comics, implemented in a declarative JSON language to specify interactions onto existing data comics. Making data comics interactive aims to give more freedom to interaction, while being guided by a highly-curated story. Likewise, interactive comics can be used as a stand-alone format on a website, or its conceptual interaction techniques integrated into other forms of data-driven storytelling.
2. **NetworkNarratives is a tool to create automatic data tours** (Section 3.2)—i.e., guided walkthroughs—through a network a user uploads. The tool creates a slide show for the user to 'read' through information about a network as it was presented by a human analyst. However, at any point, a user can interact with the system, explore the networks on their own or ask for specific information about the network. The tool explores mechanisms to combine automatic recommender techniques and data-driven storytelling and interaction. In particular, this aims to guide novice analysts in the exploration of large data sets (e.g., atlases) as well as support the creation of data-driven stories in the first place, blurring the boundaries between human and machine creation of data-stories.
3. **NetPanorama is an expressive grammar for designing interactive network visualizations** (Section 3.3) used in exploratory and/or explanatory contexts such as network analytics applications, dashboards, comics, articles, or atlases. In our particular use case, we used it create a wide range of network visualization for The Vistorian (Section 1.2) as well as a set of dashboards to explain peace-agreement signatories in the context of the PeaceRep program. The grammar captures the entire pipeline of turning network data into an interactive network visualization—network creation, network transformations and metrics, layouts and seriation, as well as interactive visualization. NetPanorama is heavily inspired by other visualization grammars such as Vega [Satyanarayan et al., 2016] and fills a significant gap in the creation of network visualizations.

3.1 Making Data comics interactive with ComicScript

Published as: *Wang, Z., Romat, H., Chevalier, F., Riche, N.H., Murray-Rust, D. and Bach, B., 2021. Interactive data comics. IEEE Transactions on Visualization and Computer Graphics, 28(1), pp.944-954.*

DataToon [Kim et al., 2019] is a flexible pen-and-touch user interface to design graph comics (Section 2.1.1), using imported network data. That tool uses different brushes to create annotations, an infinite canvas to layout and experiment with panels, and some automatic tweaks to facilitate the creation of designs. This section shows how to make data comics interactive. Generally, for this research, we faced two challenges: *a*) *how to make comics interactive and what does this interaction support?* and *b*) *how to allow designers and authors add this interactivity to comics?*

So far, data comics encountered in the wild [Cagle, 2019, Bach et al., 2018b] and research are mostly static; the panel layout itself encodes the flow of information, guiding the reader through a predefined sequence of panels[McCloud and 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., 2017b, Romat et al., 2020, McKenna et al., 2017]. 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., 2018, **?**], or to curate their own version of the story given their own approach to explore the data [Dragicevic et al., 2019].

While a powerful static storytelling medium that works well on paper support, adding interactivity to data comics can enable non-linear storytelling, personalization, levels of details, explanations, and potentially enriched user experiences. 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 (Sect. 3). When attempting to design and create interactive comics building on these operations, we found no appropriate tool to support rapid prototyping for story authors (Sect. 4). 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.

Outline and methods

To mitigate this problem and allow story authors to more quickly explore interactivity, we propose a lightweight declarative scripting approach: ComicScript. 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., 2016] 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. We explored the design space of interactive operations and iterate on ComicScript in a two-week-long design process with six illustrators, designers, and postgraduate computer science students. A gallery of interactive data comics can be found online, alongside the editor, tutorials, and detailed documentation: `https://interactivedatacomics.github.io`.

3.1.1 Interactive Comics and Storytelling

To streamline our exploration and creation of interactive data comics, we define six goals for interaction in data comics.

- **Navigation:** Readers can choose how they explore the story, revisit information, and skip ahead.
- **Details on Demand:** Readers can access additional information about the data, visualizations, and context.
- **Change Perspective:** Readers can see the data from different angles and viewpoints.
- **Branching Narratives:** Readers can choose their path through the story and explore different scenarios.
- **Gradual Reveal:** Information is revealed slowly to build suspense and encourage engagement.
- **Input Data:** Readers can provide their own data to personalize the story and see how it affects the outcome.

To further inform our design for interactions, 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. The specific methods and findings are explained in detail in the original paper and include interaction for: navigation, details on demand, changing perspectives, creating and following branches, pause and reveal more content, and input data.

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

Inspired by these examples and their analysis, we created our own examples of interactive data comics. The story in this section presents a dynamic network of **historic alliances** between countries in Europe (3.1), based on a static data comic in Figure 2.2 from (Section 2.1.1). 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).

Example 2: Interaction for Data Input and Branching

The example in Figure 3.2 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

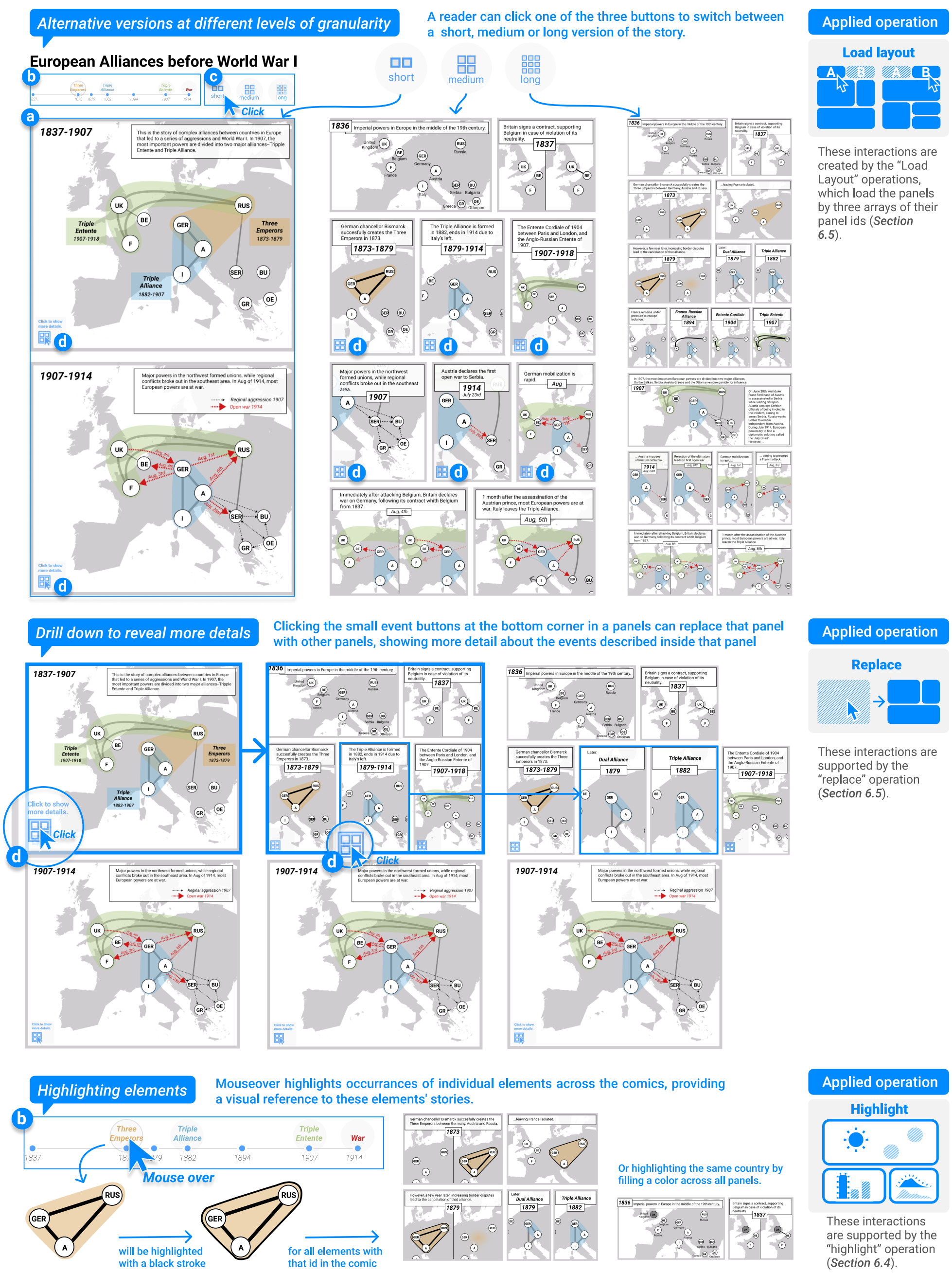


Figure 3.1: *The interactive comic of European Alliances before World War I presents a dynamic network with a set of countries in Europe, from forming alliances to regressions. With a goal of informing and educating, this interactive comic presents the story of different information granularity (*Navigate*), provides drill gown to acquire information on demand (*Details on Demand*), and highlights elements as visual navigation (*Navigate*)*

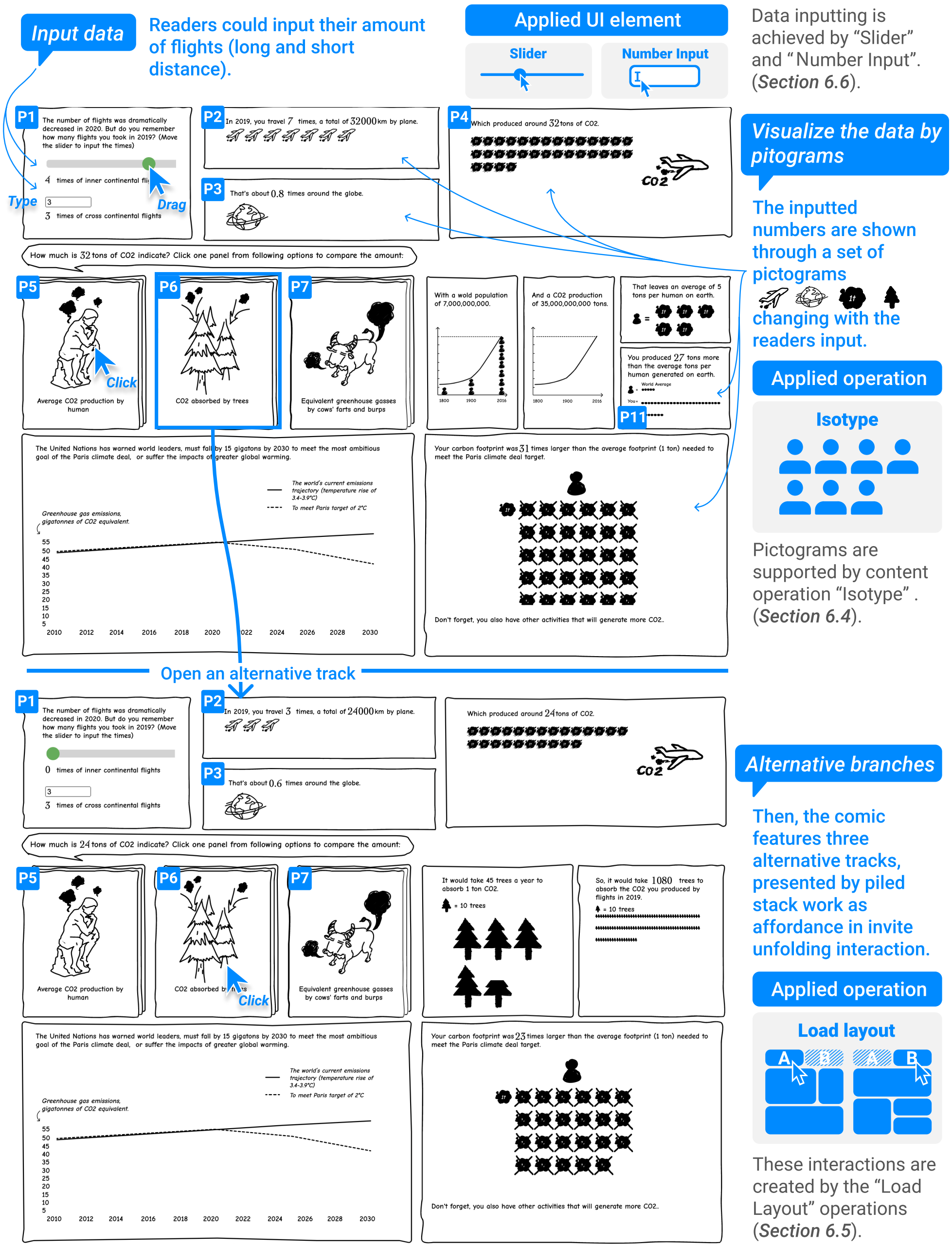


Figure 3.2: *The comic of CO2 emissions presents viewers with personalized contents by allowing data input, and provides different equivalent comparisons* (Branch).

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 CO_2 emitted to the average CO_2 produced by a human (P5), CO_2 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 CO_2 of the reader, based on their input.

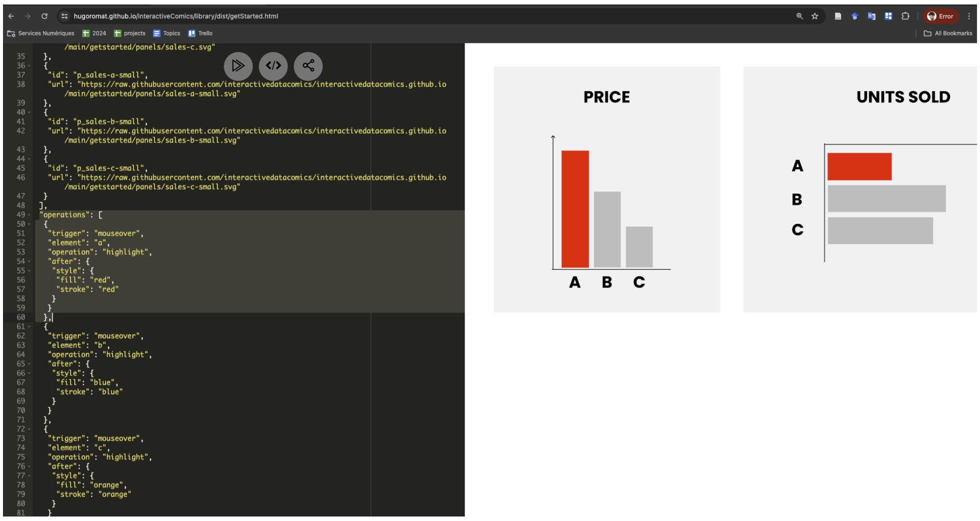


Figure 3.3: *Example of the web-editor with a ComicScript specification (left) and an interactive comic (right): hovering one of the elements inside the SVG panel hovers the 'same' element in another panel.*

3.1.2 Creating Interactive Comics with ComicScript

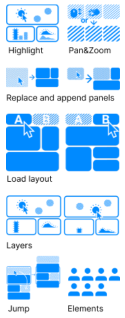
ComicScript is based on JSON and can be edited and interpreted by an online editor, and the resulting interactive comics can be exported and linked into any website. Instead of a fully functional editor like DataToon [Kim et al., 2019], we instead opted for a scripting approach to *formalize and implement* interactions first. User interfaces can be build in future work. To demonstrate the creation of interactive data comics, we implemented a web editor which imports panels, can write ComicScript and show and interact with interactive comics. The examples in the figures in this section are all created with that online editor.

ComicScript is used as follows: In the first phase, **comic preparation**, the designer plans the layout and content of each panel. This includes choosing the size and order for a clear reading flow. Designers can use their existing tools for creating visuals and panels, but the final format needs to be individual SVG or PNG files with public URLs for the web editor. In the **scripting phase**, the designer defines the list of operations that can be applied to the comic. Operations are attached to specific elements through their unique IDs and parameterized in the script.

In the scripting phase, ComicScript offers the following operations and constructs, some of which are illustrated in the figure to the side. Once panels are accessible through any URL on the web, they can be referenced in the ComicScript code. ComicScript then attached interactions to panel and and elements inside (in case elements are scalable vector graphics (SVGs), any element can be referenced by an ID.

ComicScript offers the following types of interactions, which are specified in the code as `operations`. Any operation has a trigger (e.g., a mouseover, a click) and then modifies visibility and order of comics. An example of a small interactive comic is shown in Figure 3.3, while the following summarizes the different types of operations supported in ComicScript. The full description can be found in the full paper.

- Creating panels and panel layouts
- Classes of elements onto which operations are applied, similar to CSS
- Data variables that can store and evaluate data in conditions, e.g., to trigger the visibility of a comic
- highlighting elements



- visual layers that can be shown and hidden
- adding, removing, and replacing panels
- loading a different pre-defined layout
- panning and zooming
- jumping the screen to a specific panel

3.1.3 Design Workshop and Qualitative Evaluation

We asked designers, illustrators and visualization practitioners to create interactive data comics with our tool on a topic and data of their choice. 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.

The six workshop participants were: a professional illustrator for 15 years , a postgraduate student in digital media with 6-year experience of creating illustration, a computer science postgraduate student with 15-year experience working as an illustrator, a postgraduate student in design and data science, a doctoral student in creative art and visualization, and a a team of data storytellers from the company Gramener [Gramener, 2019] with experience in creating data comics and interactive data comics by programming. Findings from the workshop include:

- Participants reported thinking of the interactions as they designed the comics and crafted the visuals, i.e., in an interactive manner.
- Participants spent the **majority of their time in graphics editing software** to create the story and design the comic panels. To create different layers or branches in ComicScript, they needed to make more panels and variations thereof.
- Participants **completed their comics with little to no help** from the facilitator.
- Participants commented that interactions made **data more engaging**
- A participant suggested that crafting interactive comics with ComicScript could both **foster creativity** and learn programming

Discussion and extensions

This is the first time interactive data comics have been described and supported through a creation tool. 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. 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.

ComicScript definitely involved a learning curve. Besides the technical use of the language and formatting visual content in images and SVGs, 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.

Participants suggested many possible 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 (Section 2.1.2) 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., 2016]. 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., 2015].

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. 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, we think 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.

3.2 NetworkNarratives: Data Tours for Visual Network Exploration and Analysis

Although powerful when used by an experienced analyst, feature-rich user interfaces present challenges for novice analysts who are required to learn possible interactions, understand the aim, perform interactions, and keep track of everything. Significant time can be taken up by repeating steps, applying them to different datasets, keeping track of one's exploration, employing layout and exploration strategies,

Published as: *Li, W., Schöttler, S., Scott-Brown, J., Wang, Y., Chen, S., Qu, H. and* ***Bach, B.****, 2023, April. Networknarratives: Data tours for visual network exploration and analysis. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1-15).*

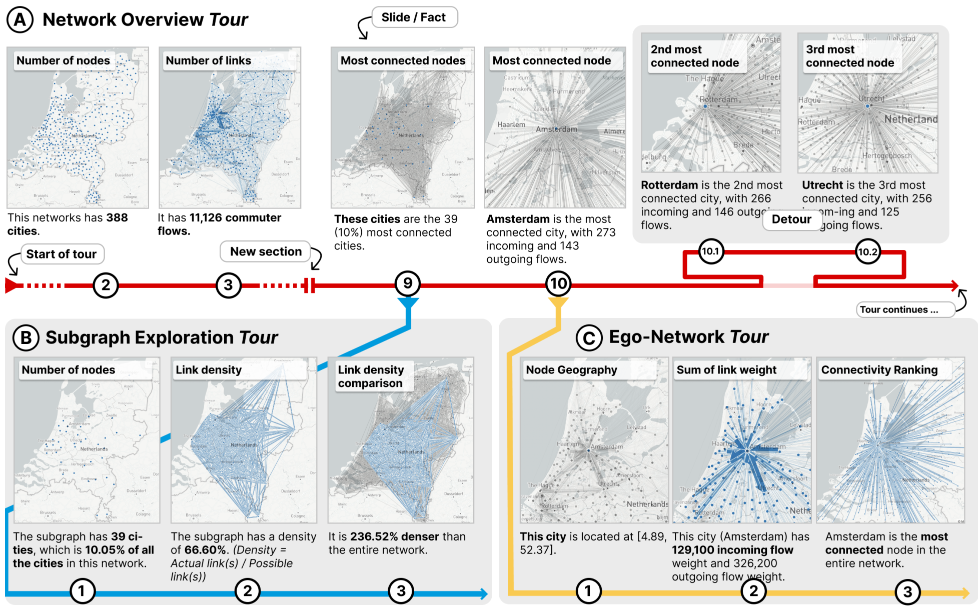


Figure 3.4: *Conceptual illustration of data tours in NetworkNarratives. An initial data tour of Network Overview (A, red line) showing facts (on slides) about the network. Each slide has a title and textual description. Numbers in circles indicate the number of facts in the tour. At times, e.g., when discussing a node or subgraph, a user can pivot to related tours (B, C) about that specific node or subgraph. Detours (D) include additional slides on demand for any given tour.*

and undoing interactions in case of mistakes. In such free-form exploration interfaces, analysts can become lost or overwhelmed [Yoghourdjian et al., 2021] or make analysis errors, such as succumbing to the *drill-down fallacy* [Lee et al., 2019]. For a novice analyst, a rich set of tool features can result in a steep learning curve that requires cognitive effort to understand each feature and its affordances and effects [Boy et al., 2015]. As highlighted in a recent study of ours [AlKadi et al., 2023] (Section 4.1), the open-ended nature of exploration can be overwhelming to novice analysts who may not know what information can be gleaned from a network and what questions can be answered with network visualization. Therefore, creating effective exploration strategies and learning about network exploration without appropriate training or experience are challenging.

In this work, we explored the idea of semi-automatic data tours to aid in network exploration. The idea of data tours goes back to Asimov's Grand Tour for multivariate data [Asimov, 1985] and has been recently described in a theoretic framework [Mehta et al., 2017]. We obtain further inspiration from ideas on guidance (e.g., as implemented in tools such as SocialAction [Perer and Shneiderman, 2009] and Small Multiples [van den Elzen and van Wijk, 2013]), recommender systems [Kim et al., 2017a], and data-driven storytelling (notably graph comics [Bach et al., 2016a] and interactive slideshows [Segel and Heer, 2010]).

In a data tour, the computer walks an analyst through an arbitrary network dataset, similar to viewing a slideshow presentation created by another analyst (Figure 3.4). The main difference is that the system mines the network for information and presents them. Each slide in a tour shows a specific piece of information about the network as explained in the caption. The work focuses itself on network visualization, given that it is directly informed on my personal experiences working on the Vistorian (Section 1.2) and training Vistorian users through workshops, online tutorials, and online manuals (Section 4.1). Our data tours aim to lower the barrier for novice analysts to learn interaction and exploration strategies and provide quick overviews of unknown datasets to expert analysts.

To demonstrate the concept of data tours, this work contributes

1. the concept of data-driven data tours for network analysis;
2. 10 extensible fully-implemented data tours, including 102 individual facts for multivariate, temporal, and geographic networks;
3. *NetworkNarratives*, an interactive user interface to experience data tours and that is publicly available and can be used as either a standalone application or an extension to the Vistorian (documentation and video demos are also available online: `https://networknarratives.github.io`);
4. two studies with 8 network analysis experts and 14 novice analysts respectively, which evaluate the usefulness and future potential of data tours and the NetworkNarratives system.

Our concept of data tours is defined with **six design goals** in mind and our 10 individual tours, currently implemented in NetworkNarratives, are designed in collaboration with five network analysis experts with backgrounds in social science (*Soc1*), history (*Hist1*, *Hist2*), epidemiology (*Health*), and archaeology (*Arch*). Qualitative feedback was obtained from three network analysis experts (*Soc2*, *Dev1*, *Dev2*) who explored their own data with NetworkNarratives. In addition, a comparative study with 14 novice analysts suggests that our tours save time during the exploration process and that a goal-driven approach can make the exploration more accessible: tours provide a simple set of entry points and allow analysts to choose from a well-defined set of tours, each representing a specific analysis goal.

3.2.1 Design goals

The design goals were based on extensive conversations with analysts, the literature, and our own experience in working with network analysts over many case studies (see paper for methodological details).

- **G1: Learn—Introduce exploration strategies, goals, and concepts to novice analysts**—In our interviews, we saw that exploration is rarely entirely open, and is typically influenced by factors including high-level research goals and prior knowledge about the data, as well as personal analysis protocols and methodologies. In an open-ended exploration, some novice analysts lack specific high-level goals when exploring a network [AlKadi et al., 2023]. They might also be unfamiliar with network concepts required to decode information from networks (such as clusters and communities, node degrees, the shortest paths, and link weight). Data tours can automate, demonstrate, and explain some of these concepts by exemplifying exploration for any given dataset.
- **G2: Reduce—Reduce cognitive and manual exploration effort**—All the analysts reported frequently switching between multiple tools to alternate between network visualizations (for topology and overview tasks) and metric calculation (for analysis tasks). Data tours can free an analyst from the majority of manual and cognitive labor required.
- **G3: Repeat—Repeat routine explorations**—Data tours can easily be repeated and reapplied to different networks. This feature can be useful when an analyst wants to repeat an analysis workflow with an updated dataset, apply the same exploration to subsets (e.g., connected components or clusters) of the complete dataset, or compare multiple versions of the same network (e.g., obtained by applying different filtering nodes and links).

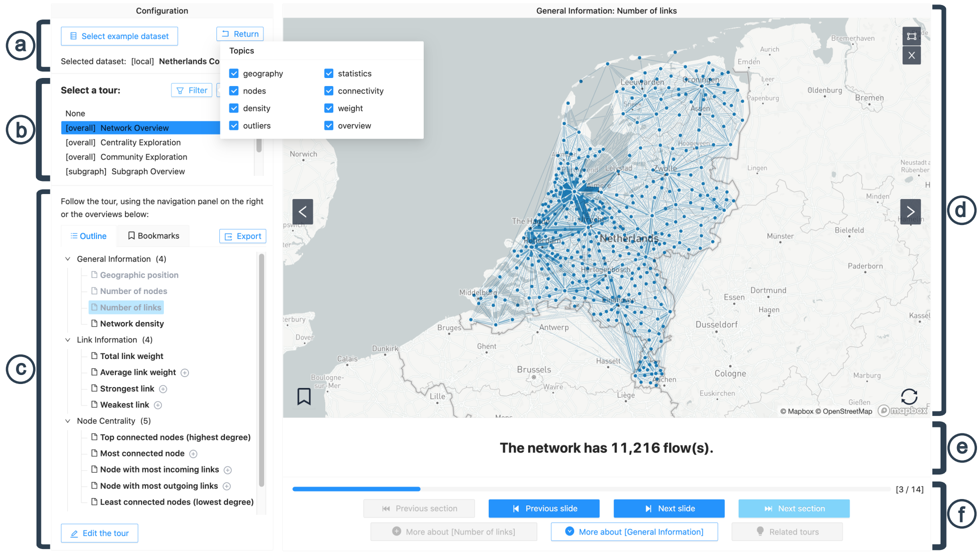


Figure 3.2: *User interface of NetworkNarratives. The left sidebar contains the data selection panel (a), data tour selection panel (b), and tour overview and starred slides panel (c). Each slide consists of an interactive visualization (d) accompanied by a textual description (e). The user can move between slides by clicking on the arrow buttons in the visualization panel (d), using the buttons in the navigation panel (f), or clicking on the slide names in the outline panel (c).*

- **G4: Balance—Balance prescription and agency**—Data tours need to be inspiring, not restricting. Similar to guidelines, they should give direction and provide detailed steps into that direction. (*"the more you produce networks, the more you have a chance to really get lost"* [*Hist1*]), but once analysts find something of interest, they may want to continue in this direction.
- **G5: Surprise—Support serendipitous discovery**—Despite being goal-oriented, an exploration process needs to remain open to serendipity and unexpected discoveries. Data tours should include a spectrum of information (time, link-weight, isolated nodes, etc), especially when networks are large and multivariate.
- **G6: Transparency—Keep information in tours and tour structure simple and transparent**—Our data tours follow regular templates scripted by a human author. We want to avoid complex recommender models whose decisions might not be transparent to the analyst. Moreover, human scriped tour templates can reflect common analysis routines, exploration goals, and story arcs.

3.2.2 User Interface

The user interface (Figure 3.2) of NetworkNarratives consists of six panels: (a) data upload and selection; (b) a list of available data tours; (c) a detailed outline of the content of the currently selected data tour; (d) the network visualization with (e) captions; and (f) a navigation panel. Below, we explain how a user applies NetworkNarratives to explore a dataset of commuter movements in the Netherlands [flowmaps.blue (based on data from Statistics Netherlands), 2020].

Importing data and defining terminology—After importing data, an analyst is prompted to specify a domain-specific terminology for nodes, links, link-weight, and subgraphs. For example, in the commuter network example, nodes might refer `cities`, links to `flows`, and link weight to the `number of commuters`. This feature was added in response to one of our analysts and aims to help relate explained facts easily to the domain [Boy et al., 2014b].

Choosing and scoping a data tour—A user can then choose any data tour from the selection panel (Figure 3.2(B)). Hovering the cursor over a datatour's name displays a tooltip containing a short description. This approach simplifies browsing and selecting tours of interest (LEARN, SURPRISE, REPEAT). Alternatively, the analyst can select nodes and subgraphs in the visualization view by clicking on nodes or using a lasso interaction and subsequently choosing a data tour exploring the selected subgraph or node (e.g., *Ego-Network Exploration*, *Subgraph Overview*, and *Subgraph Comparison*).

Starting a data tour—Once a tour is selected from the tour panel in Figure 3.2(A), its structure becomes visible in the outline panel (Figure 3.2(C)) showing a tour's sections and facts in a tree-view. The visualization view shows a popup with the title and description of the tour and prompts the user to start the tour. The example in Figure 3.4 starts with the *Network Overview* tour. Clicking the "*Click to start*" button in the popup loads the first *slide* in visualization view. Each slide refers to a single fact in the network. The first row in Figure 3.4 shows selected screenshots from the Network Overview tour. A slide includes *i*) a title representative of the fact shown on that respective slide (e.g., "Number of Nodes"), *ii*) an interactive visualization of the network (Figure 3.2(D)), and *iii*) a caption stating the corresponding fact (e.g., *"This network has 11,216 links."*). Any specific nodes or links mentioned by a fact are highlighted in the visualization. Whenever a fact mentions a term that might be unfamiliar to the novice analyst or that might require additional explanation (e.g., link density), NetworkNarratives displays a hyperlink for a popup window.

Navigating a data tour—An analyst can navigate to the next slide in the data tour by using the 'next slide button'. They can continue stepping through the slides by using the button, or choose to skip that section and press the button. Alternatively, the analyst can return or jump directly to any specific slide in a section by clicking on its title in the outline view. Within a visualization, an analyst can pan, zoom, and hover to highlight the connections of a specific node.

Detours—A detour inserts additional facts into a tour. To include a detour, an analyst can click the button in the navigation panel by appending additional relevant slides from NetworkNarratives' fact library that are recommended by the system. Likewise, detours can be included on a section level, i.e., to extend the current section. Clicking the button inserts slides with relevant facts about that section that are recommended by NetworkNarratives. An analyst can also filter slides about facts that are not of interest.

Pivoting to related data tours— Pivoting is possible for any slide discussing specific nodes or subgraphs. In these cases, NetworkNarratives displays a small popup that suggests to the user that they can pivot to a a related data tour (i.e., the yellow and blue tours in Figure 3.4). If the user decides to embark on a new tour, the new tour will be started (BALANCE). Likewise, NetworkNarratives suggests related tours at the end of each tour.

Personalizing and sharing data tours—Slides of interest can be bookmarked in any tour. In the editor panel, they can also add new slides from NetworkNarratives' library of 102 facts. Personalized tour templates can be exported in JSON format and shared with peers for reuse.

3.2.3 Fact retrieval and recommendation

In NetworkNarratives, tours are defined by a JSON specification, making the editing and creation of templates straightforward and flexible for developers. Each tour is made up of facts, described in the template and each represented as a slide in the tour. NetworkNarratives has a library of 102 fact templates about networks, including information about the network, subgraphs or individual nodes and links; about centrality measures, topological information, attribute information, comparisons, rankings, outliers, paths, geographic information, clusters, or connectivity trends are included. When starting a data tour with a new network, facts are calculated and inserted into the data tour. If a user requests a detour, i.e., additional slides for a slide or section, NetworkNarratives identifies the most relevant slides to display through TF-IDF score [Salton and McGill, 1984] as a measure of the importance of a fact to the current tour. To that end, each fact is described by a set of tags.

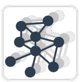
3.2.4 Data tour examples

NetworkNarratives currently implements 10 tours, whose goals and main facts are detailed as follows. The list is not meant to be exhaustive but show the richness of possible tours. The full list of all tours, including their facts, as well as some illustrations can be found online: `https://networknarratives.github.io/tours`.

Network Overview describes an entire network. It starts with four introductory slides, covering the geographic extent (skipped for non-geographic networks), number of nodes and links, and the density. The second section focuses on links, showing the total and average link weight and the strongest and weakest links in the network. The third section provides some details about node centralities and the overall community structure (number of clusters).

Subgraph Overview is similar to Network Overview, but focuses on a specific subgraph. It shows the subgraph's size and the percentage of the network's nodes. The tour also comprises important nodes such as the *Most connected node* in the subgraph, *Subgraph density*, and important links to the rest of the network.

Community Exploration explores and compares clusters in the network and shows their sizes, connections, and important nodes. For example, the *Most connected cluster* displays the cluster that has the most connections with the others. For community detection, we use the algorithm by Newman [Newman, 2004]. Advanced community detection algorithms can easily be included and used for comparison (e.g., k-means with different values for k) and shown on different slides (e.g., one slide for each value of k).



Centrality Exploration explores nodes based on different centrality measures (e.g., degree or betweenness). For example, we compute the *Average degree centrality*, the node with the *Highest betweenness centrality*, and the node with the *Highest Closeness centrality*. Possible extensions include comparisons of several centrality measures.

Subgraph Comparison compares two specified subsets of nodes and links (e.g., regions, subgraphs). Selecting this tour prompts the user to select two sets of nodes. The data tour first mentions the *Number of nodes* and the *Number of links* for each subgraph, then details important nodes such as the *Most connected node* in each subgraph, and finally reveals links between the two subgraphs (e.g., *Number of links*, *Total link weight*, and *Strongest link*).

Compare Two Nodes shows the links between the two nodes, compares their connectivities and total link weights, and finally shows the common neighbors. For example, general statistics such as *Connectivity ranking* and *Total link weight* of the two nodes are compared. Neighboring nodes that connect both of the selected nodes are shown in the last slide (*Common neighbors*) of the tour.

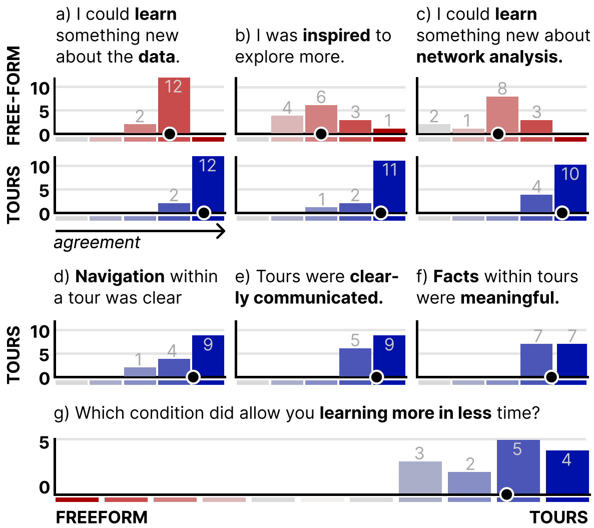


Figure 3.3: *Subjective ratings from the novice study. Darker shades imply stronger agreement with the question posed. Dark circles indicate average values.*

Ego-Network explores the network around a selected node and its neighbors. The data tour starts with the selected node and its position within the entire network. The tour then shows the node's direct neighborhood (nodes, links, strong connections), followed by their mutual connections, and finally its neighbors' neighbors.

Possible Paths explores a set of possible paths between two selected nodes. The data tour reports the path length, combined weights along each path, and the minimum link weight within each path. This data tour is motivated by *Arch*'s desire to explore historical travel costs between cities.

Follow a Path requires a selection of a set of connected nodes. The tour follows the path, explaining details about each node and its neighbors, and provides overall statistics of all the nodes in the path. This tour is motivated by *Arch*'s interest in nodes along geographic features, such as rivers, main roads, or political boundaries.

Temporal Exploration starts with an overview, showing the basic statistics about the network. It then demonstrates the connectivity evolution of the network based on temporal attributes. The data tour ends with the comparison of *Network density* over different time periods.

3.2.5 Evaluating Network Narratives

We performed two complementary user studies to understand the effectiveness of data tours in NetworkNarratives and the extent to which we achieved the design goals (LEARN, REDUCE, SURPRISE, REPEAT, BALANCE, TRANSPARENCY). One study involved the same network analysis experts as our initial interviews, plus three additional ones not involved in the initial design of the tours and feedback on the interface. The second study involved novices users without knowledge in networks or NetworkNarratives, using both a free-form exploration tool (FREEFORM), the Vistorian, Section 1.2) and NetworkNarratives (TOURS) for a comparative qualitative study. Methodological details for both studies are found in the paper. Here, I focus on the main findings from both studies *combined*. Figure 3.3 shows summary ratings from participant questionaires in the novice study while Table 3.1 summarizes advantages and disadvantages of both the tours and free-form condition from the

Cond.	Advantages	Disadvantages
Tours	Orients users with a narrative. Automatically provides facts. Can inspire deepened exploration. May lead to additional discoveries. Easy navigation. Saves time. Helps learn about analysis.	Can limit thinking and feel passive. Explanations need to be chosen carefully. Supports a rich set of views.
Free-form	Provides greater flexibility for exploration.	Harder to obtain deep insights or spot patterns with low prominence. Requires more time and effort to interact. Requires users to know where to look/have an exploration strategy.

Table 3.1: *Reported advantages and disadvantages for Tours and Free-form across both studies.*

novice study.

- When asked *"Which condition did allow for learning more in less time,"* the novices responded with an average rating of +3.71 towards Tours on an 11-point scale (Free-form=-5, Tours=+5, Figure 3.3(g)).
- Novices commented that producing insights during free-form exploration (Free-form) was difficult: *"I only get the most obvious insights;* [*it was*] *hard to further exploration and get more insights"* [N10].
- Tours and facts were perceived to be useful. The novices rated tours as well-communicated (Figure 3.3(e), avg=4.71 on a 1-5 Likert scale) and found facts within the tours as meaningful (Figure 3.3(f), avg=4.5). The analysts concurred, commenting that these facts are *"in fact, what I would look at myself"* [*Hist1*], and that they are *"good for hypothesis generation"* [*Hist1*], *"cover*[*ing*] *a lot of common steps"* [*Dev1*] and *"guides what you should look at"* [*Hist1*].
- Participants highlighted the simplicity of tours (*"I prefer because it directly guides me to see the next information"* [*Arch*]) (Reduce), its structure (*"I like* [*data tours*] *because the data facts are organized systematically"* [N13]), and that the provide a rich set of views and information about the network (*"provides multiple perspectives of the network."* [*Hist2*]).
- Novices found they could learn more about the data with Tours (ratings in Figure 3.3(a), Tours=4.86, Free-form=3.86), found Tours to teach them more about network analysis (Figure 3.3*c*, Tours=4.71, Free-form=2.86), and were more inspired to further explore the data (Figure 3.3(b), Tours=4.71, Free-form=3.07). *"*[*Data tours*] *teach me how to analyze the network. Tours are like stories with different steps. I don't need to remember the key concepts. The network visualization explains well and clear*[*ly*]*"* [*Hist2*]. Similarly, *Arch* suggested potential for educating students about network analysis, its concepts and methods and thought that the tours would be *"helpful to share with the new colleagues"* [*Arch*] for an introduction to network analysis.

Discussion and conclusions

In summary, our studies provide strong evidence for the benefits of data tours and show that our design successfully supports our initial goals (G1-G6). The main findings from our studies and collaboration with network analysis experts can be summarized as follows:

1. **Tours are an extensible concept.** Our current tours showed what our analysts were interested in. Yet, the sets of facts that can be shown in data tours are potentially very large, and the tour's power grows with the number of facts that they include. New facts can easily be added to our framework, which currently contains 102 individual facts.
2. **Human-scripted tours can reflect exploration strategies** and help teach students and novices, while automatic fact recommendation allows to follow individual interests.
3. **Data tours are complementary to free-form tools.** The novices acknowledged that they found different insights under each condition (Tours or Free-form). They commented that data tours provide sufficient insights, but using them could be a passive activity. Thus, both conditions have their unique values to the users.
4. **Data tours are a means to accelerate analysis and exploration** and reduce manual labor (Reduce). Especially when dealing with numerous networks, the analysts are required to have consistency in exploration and analysis, as well as to explore these networks quickly (Repeat).
5. **Quick overview can prevent analysts from getting lost** in too many options and help them keep track of their previous exploration history, e.g., using a specialized tool [Heer et al., 2008]. For example, the analyst can overview a network through different tours, star interesting slides, and follow up on these slides in a second iteration.
6. **Sequential tours support novice analysts** who are getting started with network visualization and learning about analysis methods and concepts (Learn). They can be used by novices without previous knowledge about networks or specific network concepts or goals, who would otherwise struggle with a very open and free-form approach to network exploration [AlKadi et al., 2023]. Tours could help onboard and familiarize novice analysts with specific analysis routines, and even help engage and communicate networks and analysis to a broad audience.
7. **Tours provide a serendipitous element** to exploration (Surprise). Comments from the user studies suggested that tours could also provide information quickly while keeping the door open for exploration.

This work is the first to explore data tours in the context of networks. Rather than implementing a full-fledged recommender engine, we opted to create *goal-oriented* tours inspired by real-world analysis practices. We see data tours as complementary to recommender approaches, as well as free-form exploration. At the same time, attempts to strike a balance between statically defined tours, data-driven recommender systems, and open-ended free-form exploration by including semi-automated techniques to retrieve related facts on user demand and allow for basic forms of personalization and exploration. We believe that there is a promising future for future tools that can automatically identify and visually communicate insights by drawing on storytelling techniques.

3.3 NetPanorama: A Declarative Grammar for Network Construction, Transformation, and Interactive Visualization

Published as: *Scott-Brown, J., Pister, A., Bach, B. NetPanorama: A Declarative Grammar for Network Construction, Transformation, and Interactive Visualization.*

The last two sections have introduced tools for data-driven storytelling and some form of interactive guidance and exploration. For most of these tools and any of the other formats—visualization atlases, dashboards, articles, GIFs, videos, etc—we need visualization and which can require specific visualization design tools. One particular problem is the design of clear visualizations for *network visualizations*, one possible domain of complex data mentioned in the context of the Vistorian (Section 1.2), GraphComics (Section 2.1.1) and Network Narratives (Section 3.2).

Visualization design requires early prototyping and user testing, quick design iterations, mixing and matching ideas from existing designs, testing different data sets [Walny et al., 2019], and ensuring visual consistency across multiple visualizations. When stakeholders are unfamiliar with proposed visualization designs, it may be necessary to demonstrate them by applying them to their own data. However, software implementations of specific techniques may be unavailable or unsuitable for integration into new projects, whilst general-purpose frameworks for network visualization (e.g., [Franz et al., 2015, Sciences-Po médialab and OuestWare, 2013, Tyner et al., 2017]) lack many of the concepts and functionality for network visualization design beyond the node-link diagram. Many network visualization techniques also rely on computing specific metrics or applying transformations to the network topology (e.g., clustering, aggregation, projections, faceting, filtering), and implementing visualizations is often time-consuming and requires specialist knowledge; the effort further increases for applications that involve multiple views or interactivity, and those requiring visualizations to be robust to different types and sizes of networks. In short, network visualization consists of many steps that need design, implementation, and often iteration.

- **C1:** Describe network semantics, including multivariate, geographical, and temporal networks, links with direction, and with node and link attributes that can be numeric (such as edge weights), categorical (such as link types), temporal (such as times corresponding to each link), or geographic (such as node positions).
- **C2:** Specify a range of visual representations, such as adjacency matrics, arc diagrams, pivot graphs, and those described in the wider literature (such as the reviews [Lee et al., 2006, Ahn et al., 2013, Beck et al., 2017, Schöttler et al., 2021, McGee et al., 2019, Nobre et al., 2019a, Von Landesberger et al., 2011])
- **C3:** Construct, wrangle and analyze networks, i.e., define a network from one or more data tables, to calculate network metrics such as node degrees, or to modify a network by projecting, filtering, or aggregating.
- **C4:** Integrate with existing (and future) libraries and toolkits to increase reuse.
- **C5:** Allow interactive exploration and control over styling.

With NetPanorama, we introduce a unified domain-specific language (DSL) in the form of a declarative grammar [McNutt, 2023] that can specify a wide range of network visualization techniques and their required transformations. NetPanorama provides a set of novel primitives and routines (*constructs*) describing networks and their semantics across multiple sources and tables, calculating network metrics and applying transformations to a network's topology, specifying diverse layouts and node seriations [Fekete, 2015], creating rich visual glyphs for nodes and links, and defining strategies for label visibility and interaction specific to network data. This

```
spec := height?, width?, pan?, zoom?, x?, y?, data+, network+,
grouping*, ordering*, table*, layout*, scale*, vis+, parameter*,map?

    data := name, (values | url | localStorage), format,
            dataTransform*
 network := name, (simpleNetwork | complexNetwork),
            networkTransform*
grouping := name, dataRef, common
ordering := name, dataRef, allowTies?, orderingCriterion
   table := dataRef, rowOrder, colOrder, group?, rowLabels?,
            colLabels?, dragToReorder?, symmetric?
  layout := name, dataRef, (attributeLayout | topologicalLayout)
            layoutTransform*
   scale := name, domain, range, scheme?, border?, layout?,
            drawAxes, overlapStrategy?
     vis := dataRef, mark, vis*, layout*, table*
```

Figure 3.4: *Top-level constructs in NetPanorama (red); apart from `vis`, instances of these have a name and may point to the name of another object (dataRef) such as a network or a data array. Instances of `data`, `network`, and `layout` can be modified by transforms. As well as being defined at the top-level of a specification, instances of a `vis`, `layout`, and `table` can be nested within the definition of a `vis` instance. Gray text indicates alternative forms of a construct.*

approach is similar to *graph level operations* [Stolper et al., 2014] in that it allows fine-grained design decisions, however NetPanorama is more expressive and adopts a declarative programming paradigm.

The challenge in creating a grammar for network visualization is to review, define, and formalize common mechanisms in visualization design, and to integrate these into a coherent framework. Our respective choices are informed by five design goals which in turn are derived from analyzing existing libraries and processes for network visualization. Inspired by other visualization grammars [Satyanarayan et al., 2016, McNutt, 2023] including our own ComicScipt (Section 3.1), NetPanorama specifications are written as JSON and interpreted by our reference implementation. Our reference implementation is partly based on Vega and other network visualization libraries and renders in a browser. However, unlike Vega, NetPanorama provides the missing constructs for network data within a common framework while exposing some functionality from those other libraries (layouts [Dwyer, 2013, Auber et al., 2017], matrix seriation [Fekete, 2015], and visual marks [Satyanarayan et al., 2016]).

3.3.1 NetPanorama concepts

NetPanorama specifies a network visualization as a combination of *constructs* of different types (`data, network, grouping, ordering, table, layout, scale, vis`; see Figure 3.4). Apart from `vis`, instances of these have a name that can be referred to elsewhere in the specification; they take a data file or an in-memory object as input and either modify this in-memory object, or create a new in-memory object. This allows structures along the visualization process, e.g., data tables, layouts, orderings, and different network models, to co-exist in parallel, to be displayed at the same time, or to be extended by different constructs. Instances of `data`, `network`, and `layout` can be modified by transform constructs, which are *local*: they do not produce named output, and act to modify a specific object. For simplicity, when referring to syntax in the paper, we use `red` for constructs and attributes, and `teal` for attribute values. To express quantity and option, we use: `*` zero or more, `+` at least one, `?` optional, and `|` alternatives. Note that some of the constructs in Figure 3.4 are actually arrays of arbitrarily many constructs (`*`,`+`). The full doc-

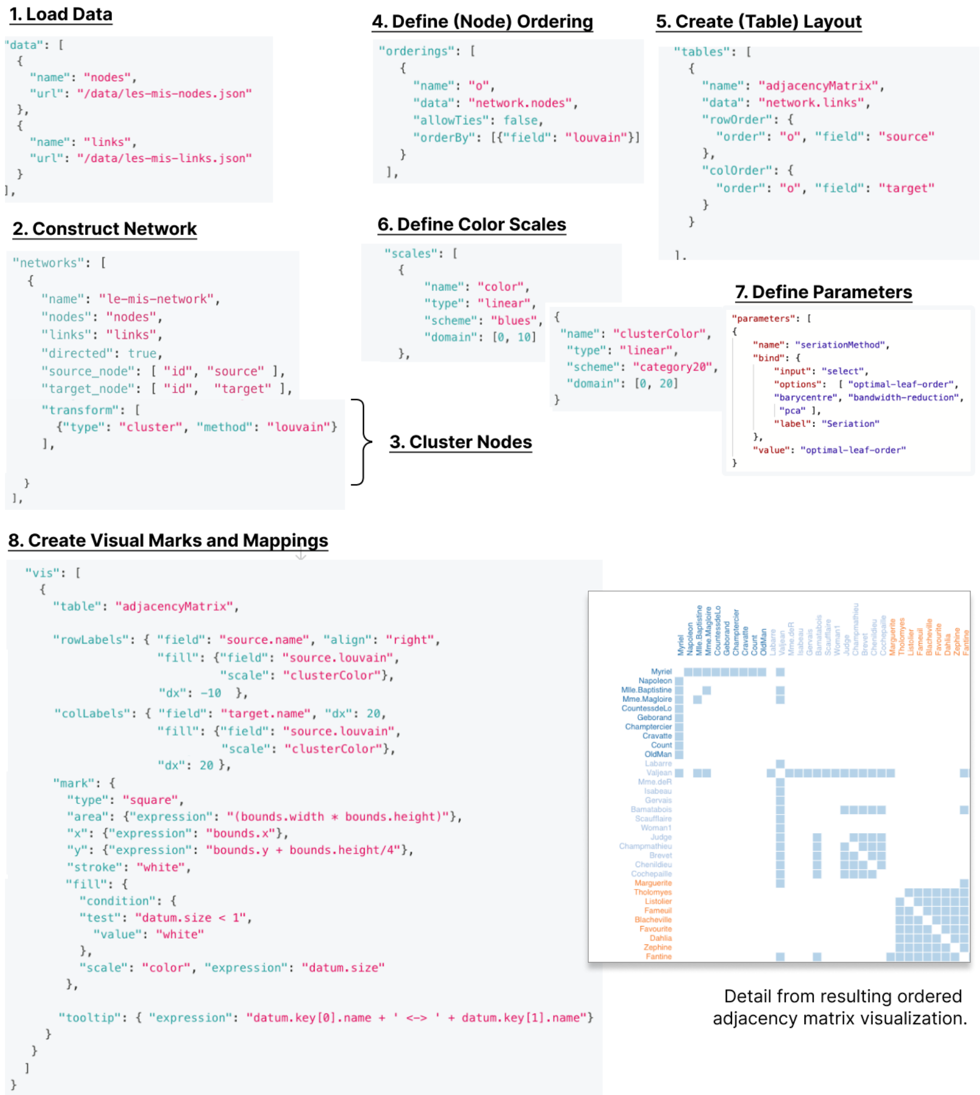


Figure 3.5: *Example of an interactive matrix visualization in* NetPanorama.

umentation of the grammar and all options can be found online along examples: `https://netpanorama.netlify.app`. An example of a full specification for an adjacency matrix with interactive node ordering selection is shown in Figure 3.5.

Figure 3.6: *Layout overview. Whether a dimension of a layout is discrete or continuous depends on what the* `scale` *or* `order` *that is used: an* `order` *will always lead to a discrete version, a* `scale` *can lead to either.*

The individual constructs in NetPanorama are listed below and explained in detail in the full paper. Figure 3.8 shows a simple network visualization example in the online editor.

- **Data Import and transformation** allows to import from many different data sources, including multiple tables and common graph data formats.
- **Constructing and defining networks** defines which columns and attributes in the imported data.
- **Calculating network metrics** such as node degrees and shortest paths.
- **Topological network transformations** calculate clusters, projections, and can filter nodes and links based on different criteria.

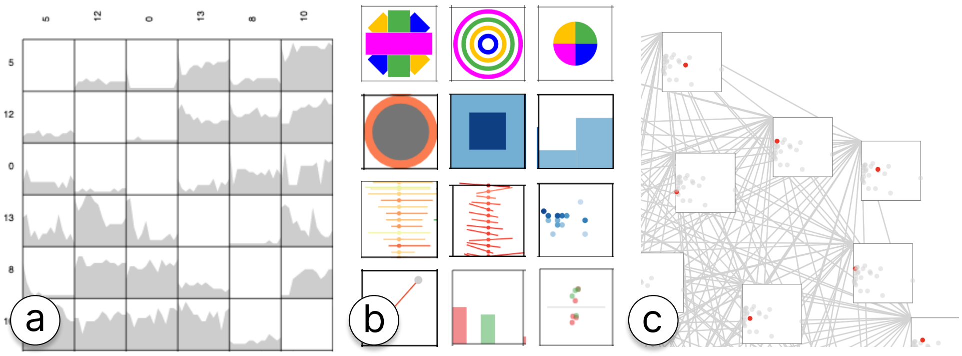


Figure 3.7: *Glyphs and shapes: (a) an adjacency matrix containing area charts showing link data, (b) 12 different glyph designs that could be used in an adjacency matrix, (c) scatterplot glyphs showing multivariate node data on a node-link diagram.*

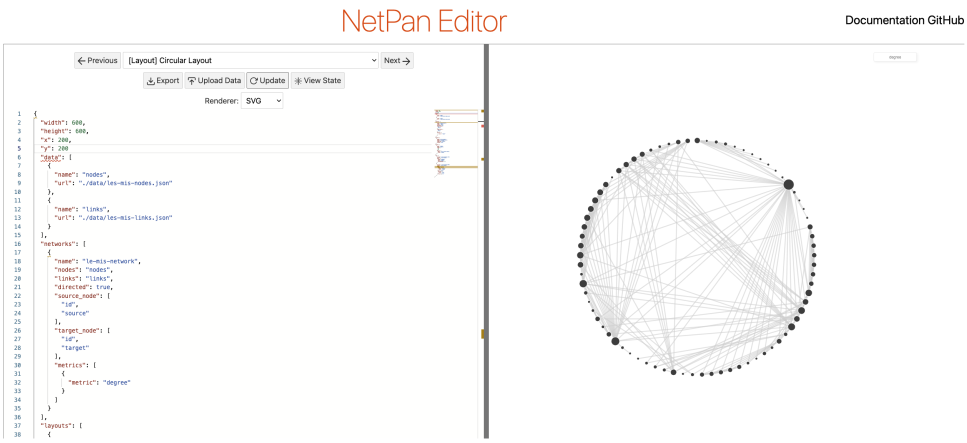


Figure 3.8: *NetPanorama online editor with specification code and preview.*

- **Calculate seriations and orderings** to optimize patterns in adjacency matrices or minimize edge-crossings in arc diagrams.
- **Layouts** (Figure 3.6) can define traditional graph layouts as well as an array of more geographic layouts, e.g., to allow designs such as Semantic Substrates [Shneiderman and Aris, 2006] or graph scatter-plots [Bezerianos et al., 2010].
- **Visual encodings, including glyphs** (Figure 3.7) are used to encode information in the visual marks for nodes and links. NetPanorama specifies a specific shape for links (linkpath) taking coordinates automatically from the respective nodes and links a link is connecting.
- **Node-label placement and visibility** allow to customize and automize the visibility of node labels for better overall visibility.
- **Interaction**, eventually, allows to show and hide nodes bases on specific criteria and include a set of UI elements such as timesliders (to navigate through time) and options to select and tweak the visualization design.

3.3.2 Example Designs

Specific Visualization Designs

Using NetPanorama, we are able to create many common network visualization techniques, as listed below. Some of them are shown in Figure 3.9 with more on our website. We are not aware of any other toolkit providing this many non-standard designs.

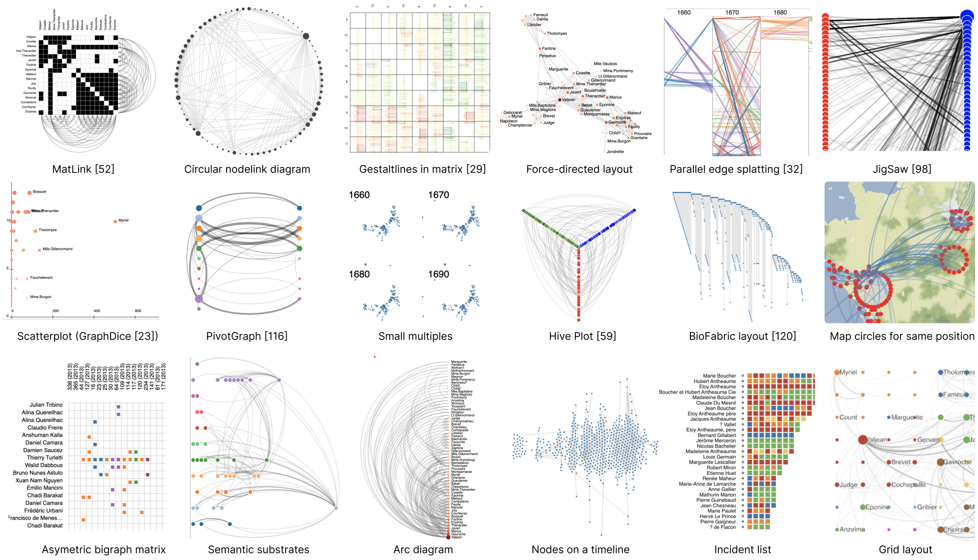


Figure 3.9: *Network visualization designs created with* NetPanorama*.*

- **Dynamic networks:** time-glyphs in matrices [Brandes and Nick, 2011], small multiples for nodelink diagrams [Bach et al., 2014], a biofabric layout visualization [William JR, 2012], animated nodelink diagrams with timesliders [Baur, 2008, Bach et al., 2015b], parallel edge splatting [Burch et al., 2011], time-arcs [Burch and Diehl, 2008], linkwave [Riche et al., 2014].
- **Multivariate networks:** Pivot Graphs [Wattenberg, 2006], Jigsaw [Stasko et al., 2008], GraphDice scatterplots [Bezerianos et al., 2010], Semantic substrates [Shneiderman and Aris, 2006], hive plots [Krzywinski et al., 2011], matrix cell glyphs for multivariate links [Vogogias et al., 2020], link weights [Chang et al., 2017], and link encodings in nodelink diagrams [Holten et al., 2011], wiggly lines from Abyss-explorer [Nielsen et al., 2009], bimatrices
- **Modifications of nodelink-diagrams**: arc-diagrams [Wattenberg, 2002], circular and concentric node-link diagrams, and layered graphs
- **Dense networks**: adjacency lists, adjacency matrices [Henry and Fekete, 2006] with diverse reordering methods [Behrisch et al., 2016a], node grouping [Bach et al., 2011], and glyph encodings for comparisons [Alper et al., 2013, Vogogias et al., 2020] and time [Yi et al., 2010, Stein et al., 2010], also matrix small multiples [Perer and Sun, 2012].
- **Hybrid techniques:** NodeTrix visualization [Henry et al., 2007], MatLink [Henry and Fekete, 2007], circular diagrams on maps, node-link diagrams inside cluster nodes.

3.3.3 Specific Usecases

We used NetPanorama to specify network visualizations for a set of projects that required bespoke designs in specific applications and contexts. Working on these projects whilst developing NetPanorama helped to inform many of the features in NetPanorama and to improve its usability. It motivated the inclusion of specific functionality including strategies for handling overlapping node labels and fixing the position of labels during panning and zooming, interactive parameters for styling, as well as the general range of visualization designs NetPanorama should be capable of expressing.

For The Vistorian (Section 1.2) we build 10 interactive visualizations (July 2024). Each visualization is highly interactive with tooltips, highlighting neighbors, pan, zoom, label occlusion strategies, a time-slider (for temporal data) and diverse visual styling parameters. All visualizations have parameterizable visual attributes and a timeline to filter nodes and links by time. Some visualizations, such as matrices, arc-diagrams, and circular diagram can chose among a set of orderings and seriations. By default, label visibility is determined by our occlusion mechanism while labels in matrices, arc-diagrams, timelines and adjacency lists are indifferent to pan and zoom. All visualizations are synchronized through brushing and linking and we made sure visual encodings (link weight, link direction, node size, color schemes) are consistent across all visualizations, which took significant iterations to the designs. Internally, that platform generates the `data` and `network` constructs based on the user's inputs, and uses these to complete a template NetPanorama specification.

The three other applications we built include an analytical web-application for literature analysis featuring three visualizations of co-authorship networks; a storytelling website explaining social relationships among intellectuals; and an overview dashboard showing links between peace agreements and the associated countries in the context of the PeaceRep program (Section 1.2). In all scenarios, we prototyped and reused visualizations with NetPanorama and iterated on them with our project partners. Each application was then optimized for its respective tasks by adding or removing interactivity, creating specific network models, calculating specific metrics, and adapting the visual appearance to the given context and audience. Specifically, no context switching nor additional libraries were needed during the whole development pipeline from data processing to the visual encoding choices since every step was done inside NetPanorama using a consistent grammar. Moreover, each step of the development and design process was potentially reusable for other designs and features in the other projects and NetPanorama. For instance, once we designed an arc-diagram, it was very easy to create a new specification for a matrix visualization, using the same network, metrics, scales, and orderings. Customization was also easy using the low-level properties of the `vis` construct—such as the `fontsize`, `stroke`, and `fill` properties—and easily reusable across several design of the same dashboard when design consistency was needed.

Across all those use cases, NetPanorama

- has **sped up our design process** by shifting our energy from implementation to discussing design requirements and complementary views for each application.
- helped us **show unfamiliar designs to our collaborators** ready with their data, instead of discussing abstract sketches and examples: we could show what information we can show about their data and discuss the actual usefulness of this information—and in consequence the designs.
- has **given our teams a language** to talk to our collaborators and hand over the designs for each visualization. As another result, we now have a wide range of ready-to-use NetPanorama-templates that can be adapted to different data formats, network schemas, visual design requirements, and interaction needs.

3.3.4 Implementation

We implemented a reference implemention of NetPanorama as a TypeScript library that accepts a specification in JSONC format (JSON with Comments), loads the required data, and renders the resulting visualization to an SVG or HTML Canvas element. This library is available as the `netpanorama` package on `npm`. We also implemented an interactive editor, which lets a user enter a specification and view the resulting visualization side-by-side. Accepting JSONC not only allows the inclusion of explanatory comments, but also allows users to comment-out sections of a specification as they experiment with alternatives. Internally, our implementation creates its own data structures for tables and networks.

NetPanorama uses several libraries that are part of the Vega project, for fetching and parsing data from a CSV/JSON file (`vega-loader`), generating a D3 [Bostock et al., 2011b] scale function from a JSON specification (`vega-encode`), parsing and evaluating expressions in a safe subset of JavaScript (`vega-expression`), and actually rendering a set of visual marks within an SVG or Canvas element (`vega-scenegraph`). However, our NetPanorama implementation is a separate system that re-uses some components of the Vega system, rather than being an extension of Vega. For topological layout computation NetPanorama delegates to D3.layout, WebCoLa/SetCoLa, Tulip [Auber et al., 2017], and Cytoscape.js. Most of these libraries run in the browser with the exception of Tulip, a C++ framework, for which we created a Python wrapper that provides an HTTP API: networks are sent by a POST request to our server, and the node positions are returned. Matrix seriation is performed using `reorder.js` [Fekete, 2015], the only suitable JavaScript library we are aware of.

Notably, the current system of reactivity and interaction—even if inspired by Vega—is specific to NetPanorama. NetPanorama models the relationships between the different items of a specification with a dependency graph. When an item of the graph is re-evaluated (for example, following an interaction or a change in a parameter bound to an input), the depending graph items are also recomputed. For instance, if the size of a visual mark construct depends on a `nodeSize` parameter, then changing this parameter value using an input will cause the re-evaluation of the mark specification and recompute the scenegraph values. However, marks not depending on this parameter would not be recomputed, enabling fast interactive visualizations. Some dependencies are also hard-coded; for example, if a text mark is specified with a label overlapping removal strategy, any zoom action will trigger a re-computation of the text marks as the zoom level affects which text marks are shown and which are not.

Summary

- NetPanorama is the **most expressive visualization grammar** for *fully* specifying network visualizations today. It covers the entire visualization pipeline from loading data, to cleaning and transforming data, to creating networks, perform calculations on the network (including clustering), create visualizations, and specify interactions.
- We imagine NetPanorama being used in web applications with network visualization components in analytical scenarios as well as data-driven storytelling where these visualizations are embedded into other comics and further adapted to the format of choice (e.g., [Bach et al., 2016a, Wang et al., 2021]).
- For example, with NetPanorama, **visualization designs can be shared online** and adapted for specific applications on demand while providing crucial transparency about network transformation (e.g., filtering) and visualization (e.g., layout) if published alongside the visualizations.
- formal ways of describing visualization designs with NetPanorama, can help with the **evaluation of the respective designs** and underlying al-

gorithms, e.g., changing one design parameter at a time and avoiding (re)implementation of many designs.

- NetPanorama could help with the **education** in network visualization by defining concepts and allowing easy prototyping and design space exploration, a topic discussed in Chapter 4.
- **Possible extensions** to NetPanorama include as 3-dimensional techniques (e.g., [Brandes et al., 2003, Bach et al., 2014]), set structures [Alsallakh et al., 2015], hypergraphs [Buono et al., 2021], Intermediate network structures such as ego-networks, graphlets, and subgraph patterns, motifs, animation and transition techniques [Bach et al., 2014], geographic network visualizations [Schöttler et al., 2021], Advanced interaction techniques for aggregation and exploration (e.g., [Riche et al., 2012, Bach et al., 2015a], semantic zoom, or portals [Hadlak et al., 2011, Ghani et al., 2011]
- We think NetPanorama could lay the **foundations for an ecosystem of tools**, including more concise languages and templates, and graphical interfaces that augment the textual specifications [Grammel et al., 2013].
- Given our many example specifications, **recommender systems** could propose visual representations for a particular dataset or task, e.g., by learning from a user's previous designs. We can also imagine higher-level specifications aimed at building entire network applications that, besides referencing visualization NetPanorama visualization designs, include concepts for annotation, multiple view layout, responsiveness, and accessibility.

3.4 Chapter summary

This chapter 3 described tools for the creation of visualization, including a tools in the context of storytelling, exploration, recommender systems and creating expressive network visualizations. Based on this, I summarize reflections and outline future avenues for research on visualization tools.

Differenet tool types Tools come in many forms and there will be more tools to come and to be researched in order to serve different needs for visualization use and design. Tools that adapt to different audiences, support different tasks, can handle different data types, run on different platforms and devices (e.g., virtual reality). Some of those tools will need to remain more specific to provide an adequate set of capabilities (tools for network analysis and visualization, tools for color palettes), while other tools want to support wider workflows (Tableau). We can also imagine visualization capabilities as part of generally non-visualization specific tools such as spread-sheet systems, accounting systems, learning applications, and interface design software (e.g., Figma charts[1]). Annotation and collaboration tools are another class of tools, such as our Viscussion, in line with related tools such as PixelClipper [Stein et al., 2010].

The need for interoperability This raises the question of how to keep tools sustainable and interoperable so that we keep a stock of more widely used tools that people

[1] urlhttps://www.figma.com/community/plugin/734590934750866002/chart

manage (see below) while offering new capabilities and modes of doing things. Especially, as researchers, it is often impossible to create full-fledged tools that people use and that can can in turn tell us how we can improve visualization tools in general. So, while aiming to explore novel concepts, we should see how we can align our efforts and translate our research efforts. For example, NetPanorama and Vistorian are both approaches aimed at being platforms, developed over many years, and aiming to bridge two gaps: the gap between research(ers) and use(ers) and the gap between individual features (data wrangling, visualization design, storytelling, learning).

Many questions remain on how we can we create a set of interoperable tools across tasks such as data collection, data organization and labeling, cleaning and formatting, analysis, visualization design, storytelling, deployment and engagement. Perhaps, what we need are standards such as as for the world-wide-web, focused on standard exchange formats for data, data errors, analysis routines (for tracking transparency), visualization types and visual mappings (e.g., visualization grammars), data types, and other APIs to allow for interaction and interoperability.

Tool design Those question call for analytical approach to tool-design that implies taking stock of the solutions we have and what we can learn. For example, on a low level we can look into common tool design patters used across tools. Common examples could include scales, visualization templates, grammars, interaction modalities (sketching), data binding, visual marks and channels. Design patterns could help compare tools [Satyanarayan et al., 2019] inform new tools, and perhaps even build tool-buildling toolkits, i.e., libraries that allow to build tools for specific workflows with the right set of features, data support, and the right level complexity for an audience.

Automation Other approaches may favor automation as already done in many tools (Calliope [Shi et al., 2021], InfoNice [Wang et al., 2018], and others summarized by Zhu et al. [Zhu et al., 2020]) as well as our own NetworkNarratives [Li et al., 2023]. Automation can be great for executing routines and help with creating drafts, the output of which then can be refined manually. Besides fully-automated recommender systems, users could perhaps create their own automation routines, similar to workflows in MacOS[2] where users have control over the features they need, every day. Other approaches could learn from a users repeated actions and suggest to do the work for them, suggesting a user's preferred color palletts, making common changes to data, suggest layouts and visualization techniques a user has used in the past. Transparency is important in such cases, both with respect to the actions taken (chosen visualization A, applied transform B), as well as why they have been taken.

Supporting Creativity and Desgin Thinking Because there are no silver bullets among the visualization tools, these tools need to allow for creativity and support users in exploring design options. *How can we enable creativity and break out of the cycle or re-using the same visualizations over and over?* By creativity, I do not necessarily mean artistic rule-breaking, but guided and informed creativity as in 'creating a wide range of options, organize a design space, discuss novel ways for problem solving', understanding visualization design as problem solving rather than artistic expression.

[2] `https://en.wikipedia.org/wiki/Automator_(macOS)`

In that sense, visualization tools support design-thinking and exploring would-be data: i.e., how can visualiation tools inspire data collection. Surprisingly to me, tools such as Charticulator, Lyra, and Data Illustrator were designed with authoring in mind, not the *process of designing* [Satyanarayan et al., 2019], i.e., exploring design spaces but for people who directly know what they want. In fact, their UIs require a set of very neatly defined steps and assignments, while sometimes leaving room for iteration. In that sense, these tools are are different from tools like our DataToon [Kim et al., 2019]), which is designed with some degree of data and design exploration in mind.

Understanding tool use and requirements To inform novel designs we should better understand current (creative) practices of users; we know too little about how people are using visualization tools in the wild. Apart from some recent work in understanding visualization novices [Burns et al., 2023], we have no idea who is using our tools, for which tasks, in which environments, to support which goals and how effective and efficient those approaches are. I think we need to understand those things to inform how we can build better tools and what tools can make the biggest change. At the same time, it is true that creative new tools can break with people's habits and teach them new things and new ways of doing things.

Improve tool evaluation Understanding tool use is strongly related to evaluating tool efficiency because ideally, we evaluate those tools in the wild and over a longer period of time. This is what we are trying with the Vistorian and which lead to identifying barriers to visualization exploration as well as interventions for supporting people in performing visualization exploration, and pattern explanation (Chapter 4). *What are those metrics and heuristics we apply to tools and how can we obtain the data necessary to assess them?*

Learning and Education Workflow, onboarding, and evaluation all point to the fact that we need to increase our efforts in training people in our tools and methods. This includes to provide mechanisms for training people in specific tools, user interfaces, and routines as well as the underlying principles and methods in visualization, commonly referred to as visualization literacy [Boy et al., 2014b]. Hence, the next chapter and third important strand of my research is dedicated to learning and education for visualization with a focus on understanding people's barriers and creating novel interventions for them learning about visualization tools and methods.

Chapter 4

Skills for visualization, data-driven storytelling, and visualization design

As visualization matures as a field and craft, with increasingly sophisticated artifacts (visualization types, interactive systems, tools, and forms such as data comics and visualization atlases), we need to take care of providing respective mechanism and tools to educate the current and future generations of users and designers. Yet, while a lot can be learned about education from other disciplines and general educational theories, education for visualization requires individual consideration for a variety of reasons. Visualization as a field integrates knowledge and methods from a wide range of disciplines such as statistics, design, art, geography, biology, psychology, cognitive science, computer science, and data science, to name just a few. It is a discipline that combines complex theory and applied craft, diverse domain knowledge and applied problem-solving skills, creativity, and criticality, design expertise and design thinking—all alongside a profound familiarity with data, technology, and human beings. Visualization also creates a growing set of very specific knowledge and methods for, e.g., visualization evaluation [Lam et al., 2012], visualization techniques and algorithms, visual variables and mappings [Carpendale, 2003], visualization design [Pandey et al., 2015] and design patterns (Section 2.2), interaction [Yi et al., 2007], case studies [Sedig and Parsons, 2016], technology [Ens et al., 2021], data-driven storytelling [Riche et al., 2018], task abstraction and theoretic models [Munzner, 2009], or programming libraries and tools for visualizations [Bostock et al., 2011b]. Eventually, while we see an increase in the number of textbooks, software tools, courses, workshops, and other resourcdaes [Liu et al., 2023] to aid with the reading, design, and creation of visualizations, visualization education needs to go beyond traditional classroom and student-oriented education and start cater to audiences as diverse as PhD students, children, researchers, working professionals, educators, decision-makers, and domain collaborators, taking into account their respective agendas, goals, and contexts. As educators and members of a global (scientific) community at the forefront of knowledge generation, it is our responsibility to engage in knowledge transfer, develop creative and practical approaches to education, and lead the research necessary to understand and improve education

My work on education in visualization can best be described through two high-level questions:

1. **What do people struggle with and what are the challenges in education?** In sections 4.2 and 4.1, I will talk about research on understanding the current practical problems and research directions in visualization education: once through the lens of the tool builder (Section 4.1) and once through the lens of visualization educators (Section 4.2).

2. **How can we support people learning about visualizations and visualization tools in informal settings, i.e., outside of formal classes or workshops?** In section 4.3 and 4.4, I show two projects about learning visualization techniques (cheat sheets) and its follow-up on interactively explaining patterns in visualizations to novice users.

All those projects are accompanied and partially inspired by the organization of international workshops [Huron et al., 2021, Huron et al., 2020, Rushmeier, 2006], a Dagstuhl seminar [Bach et al., 2023] and a special issue on visualization literacy [Bach et al., 2021], the creation and running of creative workshops on data comics, dashboards, and visualization tools.

4.1 Barriers to Network Exploration with Visualization

Published as: *Alkadi, M., Serrano, V., Scott-Brown, J., Plaisant, C., Fekete, J.D., Hinrichs, U. and* ***Bach, B.****, 2022. Understanding barriers to network exploration with visualization: A report from the trenches. IEEE Transactions on Visualization and Computer Graphics, 29(1), pp.907-917.*

Network analysis is becoming an established methodology across many disciplines such as biology, the social sciences, and the humanities and respective tools are popular. However, little is known about how analysts plan and engage in the visual exploration of network data—which exploration strategies they employ, and how they prepare their data, define questions, and decide on visual mappings. While the specific datasets are diverse and heterogeneous with respect to how they have been captured and what research questions they aim to address, they have in common that the analyst thought of their data as a *network*—i.e., as nodes and links. By applying methods from network analysis, such as calculating metrics or producing a visualization, the analyst aims to learn something about these nodes and their relationships. Hence, network analysis is a *tool* and a *lens* to model and interrogate data. It comes with assumptions, such as that a network is a meaningful representation for the data. And, it involves multiple steps in defining nodes and links, formatting data, applying visualization, and interpreting these visualizations correctly.

Over several years, we ran workshops, tutorials, talks, and interdisciplinary collaborations on network visualization and built our own network visualization tool, the Vistorian. During this time, we experienced first-hand some of the problems, challenges, and misconceptions that can hamper engagement with network analysis and visual network exploration. Many of the issues we encountered were very common, such as identifying the meaningful entities (nodes, links, attributes, time, geographical locations) from the initial dataset—e.g., a corpus of documents—that should represent nodes and links to create a network that meaningful to answering a specific research question. Other issues we observed were about using visualizations other than node-link diagrams to understand multivariate, dynamic, and geographic networks, or properly formatting a dataset for import into a tool.

While some of these issues are echoed in studies on expert data workers [Muller et al., 2019, Bigelow et al., 2020, Kasica et al., 2020, Bartram et al., 2021] or in the context of network visualization literacy [Börner et al., 2016, Sayama et al., 2016, Perer and Shneiderman, 2009], a comprehensive account of the problems novices face when engaging with network exploration was missing. Networks pose specific challenges due to their qualitative (i.e., structural) character, the combination with other data types, and the abundance of specialized visualization techniques for temporal [Beck et al., 2017], geographic [Schöttler et al., 2021], and multivariate [Nobre et al., 2019b] networks.

Here, we report on an in-depth study into the barriers some analysts face when preparing their data for visual exploration and using an interactive visual network exploration tool. For simplicity, we relied on a single tool, the Vistorian, that we consider a state-of-the-art network visualization tool comparable to Gephi [Bastian et al., 2009], NodeXL [Smith et al., 2010, Smith et al., 2009], or Palladio [Stanford University Humanities + Design lab,]. The Vistorian has been specifically developed to provide an easy entry point to network visualizations (node-link, adjacency matrix, timeline, map) and interactive exploration for data analysts without a technical background. In a first log study (Study 1) we tracked users of the Vistorian over several months to understand how they use visualizations and interactions in their day-to-day work. To complement these anonymous data with qualitative data about potential barriers, we designed a 6-week network exploration course and delivered this to 36 analysts whilst closely monitoring their progress, giving personal advice, and engaging in individual interviews (Study 2). The course was open to anybody and included analysts from computer science, the social sciences, and business. It introduced the basic concepts of networks, example visualizations for multivariate, temporal, and geographical networks, hands-on activities for sketching and formatting data, and hands-on tutorial with the Vistorian.

Below, I report only on the barriers and main findings, while more methodological details are found in the paper.

4.1.1 Users, Course, and Participants

The goal of the 12 weeks open course was to help participants—no special requirements were imposed, anyone with some network data to analyze and visualize could join the online course— *1*) define the goals of their data exploration, *2*) become familiar with a range of network visualization techniques, through theory and hands-on use, and *3*) use different types of interactive visualizations to explore their own data. The course comprised: Weeks 1 & 2: Network Data Preparation, Week 3: Data Shaping Techniques & Challenges, Week 4: Network Exploration Using Node-Link Diagrams, Week 5: Network Exploration Using Adjacency Matrices, and Week 6: Timeline, Maps, and Coordinated Views. Every participant worked with their own data.

In total, we had 36 participants from a wide variety of backgrounds, including history, social sciences, business & management, health & medicine, law & politics, natural sciences, and technology, education and event management registered for the course. In a pre-course questionnaire 18 (50%) participants stated that they had no experience with network exploration, 12 (33.3%) had explored one or two network visualizations prior to the course, and 6 (16.6%) participants stated experience with three or more network visualizations prior to the course.

4.1.2 Barriers

Across both studies the log-study and data from the course we observed 8 barriers hassling people along their analysis and exploration. The full paper contains a lot of detail about individual observations (grouped and numbers O1-O8), goals, more detailed mitigation strategies and information about individual participants' datasets and goals.

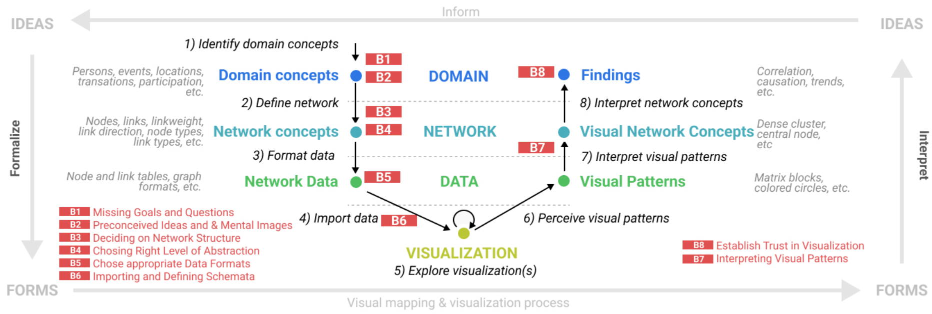


Figure 4.1: *Barriers (red) that may be encountered during visual network exploration process, while translating domain concepts (ideas) into network structures and visualizations (forms) and back into findings.*

1. **Missing goals & questions:** Many participants had no clear idea what they were looking for in the data. While missing goals for exploration are not a barrier per se and can inspire creative perspectives on the data, a lack of specific goals can lead to irrelevant findings or incorrect conclusions, e.g., via drill-down fallacies [Lee et al., 2019]). For example, some participants thought they needed cluster analysis or to look at central nodes. However, the higher-level goal of their research turned out to really be about the spatial distribution of nodes rather than about connectivity or some temporal evolution of people (nodes) joining a community.

2. **Pre-conceived ideas & mental Images:** Part of the reason for barrier 1 was that, some participants had a very clear idea of how their data and visualizations *should* look: e.g., which parts of the data would be nodes and links etc. However, such preconceived ideas and mental images hindered a creative and objective exploration process that allows for new perspectives on the data. For example, nodes and links in social networks were commonly *imagined* as people (nodes) and their social relationships (links). However, in some cases, modeling the types of relationships as nodes (e.g., family, friend, work-relationship) and the co-occurrence of these relationship as links would reveal totally different (new) perspective of the data that could correspond to different research goals.

3. **Deciding on a network structure:** Strongly related to to barriers 1 and 2, some participants generally struggled to decide on a network structure (nodes, links, link weight, node and link types) and how to think about their data in terms of networks. In some way this is the extreme opposite of barrier 2, while being independent from barrier 1 (missing research goals). Hence, whether goals are present or not, making the decision of defining nodes and links requires (informed) decisions that will impact an analysis. Failing to make an informed decision delayed subsequent processes and findings, especially for complex data sets.

4. **Choosing the right level of abstraction:** In choosing node and links, participants abstract from the initial data: they may remove some aspects, while highlighting others. Once defined, networks allow for further abstractions in terms of network aggregation (clustering), filtering (remove nodes with few connections), and other transformations based on the networks *topology* (e.g., projection). For example, the initial network visualizations of participants not always yielded insights, because an intuitive mapping of domain concepts to

nodes and links led to visually cluttered networks due to too many nodes or edges (or both).

5. **Choosing the appropriate data formats:** Choosing the appropriate (machine readable) data format and transforming data accordingly can become a barrier through, e.g., (1) not understanding the data format required, (2) confusing different formatting options, and (3) inconsistent formatting. Besides general benefits of tables for making data accessible [Bartram et al., 2021], node and link tables are relatively straightforward to understand, are editable in common applications such as Excel, and are successfully used in other network visualization tools [Smith et al., 2010]. However, some participants struggled to decide whether a node table or a link table serves best for their data.

6. **Importing Data and Defining Schemata:** A data schema creates an explicit mapping between the data and the visualization. For example, in an earlier version of the Vistorian, users uploaded a link table and then had to manually specify the role of each column. In addition, they could upload a node table and a location table, or click a button to retrieve geo-coordinates for locations mentioned in the tables. Observations from early workshops showed that planning and coordinating all these actions is a barrier for some analysts . Our import wizard addresses this issue, but can slow down the data import for advanced analysts.

7. **Interpreting visual patterns in visualization:** Only once barriers 1-6 had been mastered, our participants were actually able to work with visualizations—the initial main objective of our research. Interpreting visualizations includes reading low-level encodings of visual elements (e.g., through size or color) alongside their composition into more complex visual constructs. For example, blocks of cells in an adjacency matrix represent a densely connected cluster, while the same cluster may appear as a set of overlapping arcs in a time arc visualization. Likewise, interaction can lead to changes of visual patterns on-the-fly. To better understand visual patterns, participants reported on multiple coordinated views helping to interpret and verify a pattern visible in one visualization through another one. This challenge in particular motivated the research in Section 4.4.

8. **Establishing trust in a network visualization:** Trust in visualization is a generic topic [Mayr et al., 2019]. It can influence if analysts engage in a network visualization process in the first place, how findings are derived from the visualizations, and if a network visualization is deemed valid for evidence in (scientific) communication. One problem can be unfamiliar visualization (e.g., adjacency matrix) and their respective construction methods such as ordering (in the case of matrices), or layouts in node-link diagrams. Another problem is understanding the *provenance*, including the many decisions along the network creation: conceptualization of nodes and links, network transformations, and filtered elements.

4.1.3 Implications for Education

Many of the barriers observed we could mitigate through 1-1 discussions during the course; we discussed alternatives with participants, had them sketch their data and question, report to other participants etc. Below, we summarize our implications for overcoming the barriers we identified.

1. **Work with participants' data.** Allowing participants to work with their own data was hugely important for the success of our course. At the same time, clean demo data sets are important to explain and illustrate particular aspects of network visualization. Ideally, demo data should be relevant to the course audience and consistent across the course to allow for the comparison of techniques.
2. **Focus on understanding goals.** Early course activities should focus on helping participants to reflect on their own goals, which is also beneficial for course instructors to provide individual support. Pre-course questionnaires can help assess the extend to which missing goals are a problem.
3. **Provide activities to identify & mitigate barriers.** We found that sketching activities provide an easy entry point for analysts to engage in the data preparation stages of network visualization: namely sketching and clarification of once data, goals, and expectations from a visualization. Additional research is needed to explore how to frame sketching activities in this context and also think beyond sketching, and to develop further activities.
4. **Balance independent exploration & guidance with visualizations.** We tried to strike a balance between scaffolding participants' explorations and allowing them to come up with and explore their own strategies. Some participants will prefer to dive into their data and explore all sorts of visualizations, others need more guidance in understanding specific visualizations and the breath of possibilities, especially where interactive. We think there is a lot to do research on to understand how we can better communicate and teach *interactive* exploration with complex visualizations.
5. **Create teaching resources.** We need to start exploring additional resources that can facilitate teaching (network) visualization in ways that mitigate the barriers discussed in this paper. For example, success stories and galleries of successful network visualizations (beyond node-link diagrams) can help to show what is possible with network visualization and can motivate novices and expert analysts to consider network visualization as a method in the first place.

4.1.4 Implications for Tool Design

We can also deduce implications for the design of learning and exploration tools.

1. Support the analyst's perception of the data schema by offering a concept-demonstrating miniature (e.g., **concept map**) or diagram, an approach already taken by Origraph [Bigelow et al., 2019]. Changes and selections to this concept map can be reflected in the network.
2. Encourage **incremental network building** and iterative exploration; this is similar to the idea of enabling 'progressive evaluation' in the Cognitive Dimensions of Notations framework [Green, 1989]. Start with a simple network visualization with a limited number of visual encodings, then add encodings as users get familiar with the visual encodings.
3. Provide **coordinated multiple views** using different types of visualization. This not only allow the analysts to take advantage of their complementary strengths, but also allows them to learn how to interpret an unfamiliar visualization by observing how it changes as they interact with a more familiar one.
4. Support **creative approaches** to creating networks. For example, show possi-

ble networks derived from a given a "seed" network through aggregation, projection, and filtering. This could potentially be automated to create serendipity.

5. **Automatically recommend or highlight** potentially insightful views and data pattern.
6. **Explain visual patterns and encoding as part the tool**, e.g., using automatic annotations suggesting explanations for specific patterns or miniature galleries of visual patterns found inside a dataset.

Implications 5 and 6 lead to research on automatically generating stories and walkthroughs [Li et al., 2023] for network visualization and the automatic explanation of visual patterns in adjacency matrices [Shu et al., 2024]. While I am omitting the first paper here for brevity, I explain in Section 4.4.

4.2 Challenges from an Educator Perspective

Published as: ***Bach, B.***, *Keck, M., Rajabiyazdi, F., Losev, T., Meirelles, I., Dykes, J., Laramee, R.S., AlKadi, M., Stoiber, C., Huron, S. and Perin, C., 2023. Challenges and opportunities in data visualization education: A call to action. IEEE Transactions on visualization and computer graphics.*

This section looks at challenges for visualization education from an educator perspective and issues a call-to-action for more reserach and deliberation in this area. The work is the result of a 1-week Dagstuhl seminar we organized in the summer of 2022 [Bach et al., 2023] on *Visualization Empowerment: How to Teach and Learn Data Visualization*. I lead the collaboration on challenges, which then evolved over the following year through regular reading, reflection, discussion and knowledge exchange to share, situate and structure our **personal experiences with visualization education**. The group comprised 21 members and the process led us to describe 19 challenges grouped into seven themes: PEOPLE, GOALS&ASSESSMENT, MOTIVATION, METHODS, ENVIRONMENT, MATERIALS, and CHANGE (see sec:challenges). Our goal was to discuss these challenges in the context of visualization and capture a rich picture of the current state of visualization education as experienced by the 21 authors as educators and academics in data visualization. This bottom-up process aims at the first empirical reference point to inform discussion, research, and practical advice for visualization educators as well as inspire the community to work towards broader and better data visualization education. For each of the challenges, we also formulate *research questions* (Q1-Q43) that highlight the need for concerted activity: issues we need to better understand, solutions we need to invent, approaches we need to evaluate, theories we need to create, etc. Eventually, our call-to-action details five *opportunities*—`DIVERSITY+INCLUSION`, `COMMUNITIES`, `RESEARCH`, `AGILITY`, and `RESPONSIBILITY`. More info and supplementary material can be found on `http://viseducationchallenges.github.io`.

The 21 co-authors of this work include educators and researchers in academia (10 F, 11 M), teaching in the UK (7), Canada (5), Austria (4), Brazil (2), France (1), Belgium (1), and Denmark (1)—who all met at the Dagstuhl seminar mentioned above. Through a structured survey, we collected information about authors' visualization courses (i.e., topics, methods, tools, and materials) and target audiences. I lead the paper writing, its conceptualization and contextualization, it's approach, I organized meetings and discussions and took on the main writing and editing of the paper. Each challenge was worked out and drafted by a group of 3-5 of the co-authors, including myself.

In the following, I briefly lay out some of the main challenge themes discusses in the paper amended by personal reflections. I refer to the paper for a full description of all challenges and their associated research questions which are summarized in Table 4.1. The paper also contains the detailed methodology and limitations of our approach.

Figure 4.2: *Overview of contributions, contributors and methods.*

Tag	Challenge
	People
Ppl1	Impact of educators' diverse backgrounds on visualization teaching
Ppl2	Leveraging and catering to learners' diverse backgrounds, goals, and needs
Ppl3	Acknowledging and embracing diversity to allow for mutual learning
	Goals&Assessment
GA1	Identifying learning goals and designing objectives tailored to specific groups of learners
GA2	Assessing creative, project-based, and problem-oriented work in a fair and efficient manner
GA3	Assessing learners' work at scale and distance
	Motivation
Mtv1	Communicating the need *for* visualization education
Mtv2	Retaining motivation *during* learning visualization
	Methods
Mth1	Fostering core skills around visual representation and interaction
Mth2	Developing 'specific' and 'general' skills and competencies
Mth3	Adapting methods to learners and contexts
	Environment
Env1	Providing environments for hands-on, creative, and collaborative work
Env2	Using online, hybrid, informal & workplace environments
	Materials
Mat1	Finding, evaluating, and adapting materials
Mat2	Reusing and adapting materials
Mat3	Creating and updating materials
Mat4	Creating materials for informal, self-paced learning
	Change
Chg1	Understanding the role and effects of AI
Chg2	Overcoming inertia and adapting to change

Table 4.1: *Overview of our 19 challenges grouped into seven themes.*

Challenge theme: People (Ppl) We differentiate between *educators* (those planning, preparing, conducting, facilitating, and evaluating learning), and *learners* (those wanting to gain experience and skills whether obliged or voluntarily). As data visualization education spreads across disciplines, educators and learners increasingly come from a vast range of backgrounds and levels of education, industry and academia, from different segments of society, age groups, cultures, disciplines, and work environments, and have different objectives in learning about visualizations. In fact, many disciplines currently discuss challenges and strategies for incorporating data analysis and data visualization in their course offerings and curricula in, e.g., Art [Bertling et al., 2021], Computer Science [Bach et al., 2021], Design [Parsons, 2022, Moere, 2007], Digital Humanities [Ballentine, 2022], English Studies [Graham, 2017], Math [Setiawan and Sukoco, 2021], Political Sciences [Henshaw and Meinke, 2018], Statistics [Loy et al., 2019], Secondary Education [Kahn and Jiang, 2021, Lee and Wilkerson, 2018]). The challenges with people in visualization include *the identification of learning goals and designing objectives tailored to specific groups of learners* (GA1), the *assessment of creative, project-based, and problem-oriented work in a fair and efficient manner* (GA1), and assessing learners' work at scale and distance (GA3).

My personal experience with respect to this challenge came from a visualization course for working professionals which I lead in my time at Edinburgh for several consecutive years; I had (and still have!) very little ideas what visualization skills working professionals in Industry need. For researchers and aspiring visualization designers, I have a better understanding. Researchers want either a) specialized tools that do exactly the exploration and analysis they want in a very simple manner, or b) skills in making research findings available to larger audiences. Designers needs a fundamental understanding of visualization design rules and programming. However, in Industry, some employees simply want to know Tableau, others create a Dashboard for Excel data, again others some help with polishing graphics and infographics, and again other want complex analysis. There is a rich space for research here to better understand how visualization training can help the Industry and large real-world applications.

Challenge theme: Goals&Assessment (GA) Learning goals are important to inform course content and set expectations about what the learner will be able to know and do [H. Brown and D. Green, 2020]. They also inform teaching methods (Methods) and plan assessment against more fine-grained learning objectives [Melton, 2014]. Assessment, in turn, effects learning in a closely related cycle [Pereira et al., 2016]; it can happen *during* the process of learning (formative assessment) to help learners improve, or at the end for accreditation (summative assessment) [Biggs and Tang, 2011]. This section discusses three respective challenges in the context of visualization: identifying and communicating learning goals (GA1); the fair and reliable assessment of creative, project-based, and problem-oriented work (GA2); and fair and efficient assessment at scale (GA3). The challenges include the *identification of learning goals and designing objectives tailored to specific groups of learners* (GA1), the *assessment of creative, project-based, and problem-oriented work in a fair and efficient manner* (GA2), and *the assessment of learners' work at scale and distance* (GA3).

My experience with these challenges is grounded in teaching at the Masters level in a design and computing program. Creativity, problem solving, and reflection skills are time intensive to assess and often require a very detailed understanding of how the student works. This is impossible in large classes and with standardized tests. Ideally, we want individual assessments that highlight each student's individual capabilities. On the other side, I think we need more structure in our description of visualization knowledge. There is a lot of 'common sense' being built up through research and practice but research is limited to structure the rules and exceptions in visualization. Some of those challenges are reflected in our work on visualization guidelines [Diehl et al., 2018, Bach et al., 2022a]. To manage the marking effort (in a class of 80+ students), we grouped students and assessed the group project. This allows for more complex visualization projects at the cost of individual assessment and fairness. But, hopefully providing more experience to students with respect to larger visualization projects and collaboration. However, some students might prefer individual projects. How to assess professionals with their different backgrounds skills (and individual learning goals) remains an option question.

Challenge theme: METHODS (MTH) Common methods in visualization education include lectures, sketch-based exercises, brainstorming sessions, design critiques, co-design, project work, group discussions, game playing [Adar and Lee-Robbins, 2023], tutorials with visualization tools and programming, peer feedback, project work and presentations, and many others. This places the activities that learners *do* [Goodyear et al., 2021] at the center of much of our education methods. An educator then need to consider how to design [Goodyear, 2015] for this *doing*, that is, to *find (or invent!) methods for teaching core skills around visual representation and interactivity* (MTH1), *teaching soft-skills required for visualization* (MTH2), and *finding solutions for adapting existing methods and activities* (MTH3).

We are currently working on an online platform to collect, browse, and discuss education activities for visualizations. Again, this was a project originating from Dagstuhl discussions. In particular, I was working on an activity to make people familiar with the variety of visualization design tools (unpublished work), supported by our website `http://vistools.net`. Among those preliminary activities we crowed-sourced during the Dagstuhl seminar we find activities involving sketching, learning visualization tools, activities specifically for kids, etc. In the future, we hope to get the chance to collect more activities and to conduct a more holistic analysis.

Challenge theme: MATERIALS (MAT) As *material*, we consider any artifact (or content thereof) to support education, learning, and training. This can include lecture slides, recorded video lectures, text books, tutorial outlines, curricula,[1] learning goals, exams and quizzes, positive and negative visualization examples, websites and blogs, guidelines [Diehl et al., 2018], research papers, or descriptions of visualization activities [Huron et al., 2020]. While the number of visualization materials is steadily growing [Liu et al., 2023], few of them have been created for the context of education—i.e., with guidance and pedagogic principles in mind—and it is unclear which of these materials best support particular educational contexts. This seemingly paradoxical situation leads to a range of challenges, starting from *finding and evaluating materials* (MAT1), to *adapting existing materials to one's specific contexts* (MAT2), to *creating materials and sharing* them (MAT3), and *creating materials for informal, self-paced learning* (MAT4).

My personal experience here is based largely on a) the roll-out of the Vistorian platform and its consequent research (Section 4.1) as well as on the professionals

[1] `http://education.siggraph.org/resources/visualization/education`

course. The specific problem in both cases was that some people were not able to commit to the meeting / course times we had set and needed more material to play with on their own pace given that they were working on their own data. Tutorials and demo data, for example, required people to translate those instructions into their own data sets and research questions. This experience further inspired the two solutions I detail in Section 4.3 (visualization cheatsheets) and Section 4.4 (pattern explaination).

4.2.1 Opportunities for Research and Call to Action

The seven challenge themes—People, Goals&Assessment, Motivation, Methods, Environment, Materials, and Change—should be seen as lenses, rather than taxonomic groups: the themes reflect factors important to planning, understanding, and evaluating learning, while leaving room for overlap and prioritization. While many of these challenges occur beyond visualization—e.g., learner's diversity (Ppl2), assessing creative and problem-solving skills (GA2), or challenges in creating material (Mat3)—we deemed these challenges important in visualization and describe how they unfold in the specific context of data visualization. Other challenges—such as motivating the need for visualization (Mtv1) and teaching visualization core skills (Mth1) are more specific to our domain. In the remainder of this section, we use these challenges and research questions to discuss how we move forward from here to create better education in visualization.

- we need **reliable empirical evidence** about the effectiveness of our approaches, reported in ways that are open to scrutiny and enable close and conceptual replication [Brandt et al., 2014], in order to understand what works, how much and under which circumstances.
- We need to **provide guidelines** and possible solutions (education design patterns?) for planning and conducting educational activities starting from specific visualization activities, to courses and curricula, as well as help **formulate higher-level theories of visualization education**.
- We must also **embark on "theory borrowing"** from related educational fields such as computer science, design, or human-computer interaction. For example, to inform the design of education, we could start asking questions such as *Who*, *What*, *Why*, *When*, and *Where*. Such an approach might produce the kind of pragmatic mid-level models for data visualization education that are advocated by Tedre and Pajunen in their work on the role of theory in computing education research [Tedre and Pajunen, 2022].
- We need to plan and **conduct rigorous research** in visualization education. While some of the research questions mentioned in this paper could be solved with methods from human-computer interaction and user-centered research, others require **methods from the fields of psychology, pedagogy and education**. This includes **addressing ethical issues in research** around working with potentially vulnerable populations such as children, elderly, or disabled persons as well as testing different and replicable conditions in educational settings [Brandt et al., 2014]. Papers that report on empirical studies in visualization education provide excellent starting points (e.g., [Beasley et al., 2021]) but as is typical for human-centered research, it will take a long time and **many studies involving a diverse range of people and settings** to provide actionable evidence on the impact of specific teaching approaches.
- **Building interdisciplinary communities for exchange and dissemination** includes connecting to practitioners, educators, policy makers, and learners, and to

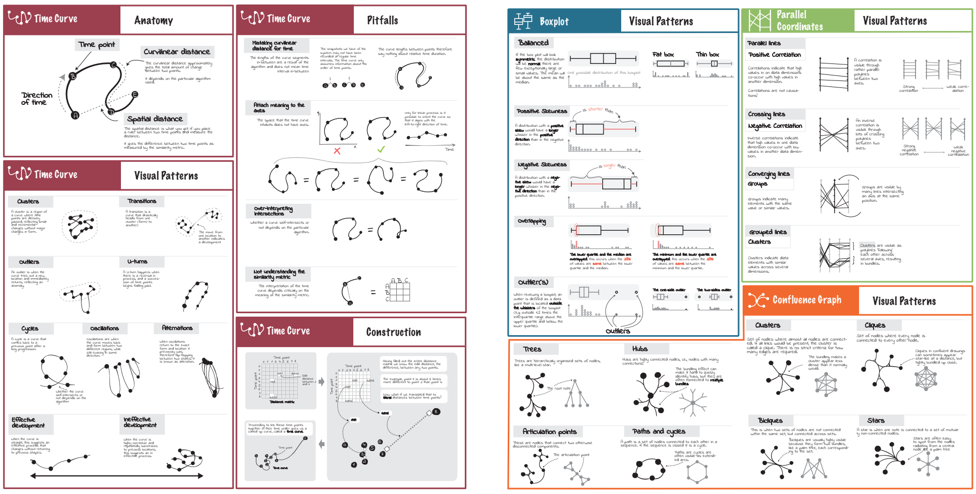


Figure 4.3: *Examples of customized cheat sheets combining several aspects of visualizations. Left: example of look-up summary sheets for the time curve visualization showing anatomy, visual patterns, construction, and pitfalls. Right: Cheat sheet showing visual patterns for several techniques: boxplots, parallel coordinates, and confluent graphs. More cheat sheets and information can be found at:* `https://visualizationcheatsheets.github.io`

engage with the academic education communities to cross-fertilize research and adopt accepted research method. For example, the Eurographics conference has a dedicated education paper track, the Design Research community has a call for *Futures in Design Education* by the *Education Special Interest Group*[2] and in 2023 both Information+[3] and the IEEE VIS *1st EduVis workshop* [Keck et al., 2023] explicitly invite submissions on education.

Eventually, as a community, we need to take responsibility about the future of our discipline and research. Because, through education, we will define the future of our discipline: who will use our research and tools? How much will they be used? And perhaps most of all: are we researching and designing the right things that are most useful to people? In particular this last question requires us to listen and observe to our potential users and learn from them, first and foremost.

In the next two sections, however, focus on interventions to help people better understand and use visualization techniques in the wild, while eventually helping to understand further barriers through those interventions.

4.3 Cheat Sheets for Data Visualization Techniques

Published as: *Wang, Z., Sundin, L., Murray-Rust, D. and* **Bach, B.***, 2020, April. Cheat sheets for data visualization techniques. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-13).*

Reading visualization techniques requires 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]. We found a lack of material that supports the understanding of specific visualization techniques, especially when a visualization technique is used outside of a teaching scenario: *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.

[2] https://shorturl.at/gjmR8
[3] https://informationplusconference.com/2023

Our *visualization cheat sheets* are inspired by infographics, data comics [Zhao et al., 2015, Bach et al., 2017a, Zhao et al., 2019], assembly instructions, and the 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.

We present cheat sheets for a selection of non-trivial visualization techniques representative to different data types (temporal, relational, hierarchical, statistical and 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 a 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.

To evaluate our designs and concepts, we ran a user study and co-design session with visualization teachers. We also refined our design guidelines and respective visual design-materials in our co-design workshops, eventually inviting everyone to create cheat sheets for teaching, personal learning, and for any new visualization technique that gets designed.

Design Principles

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

- **D1—Modularity:** create *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 to 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, 99]. 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].

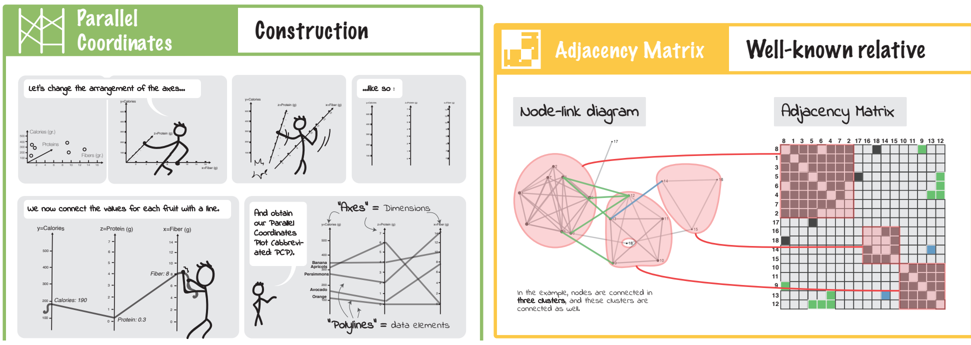


Figure 4.4: (*a*) *Part of Construction for PCP, showing its "creation" from three-dimensional scatterplot in a comic-strip.* (*b*) *Well-known relative for adjacency matrix, adapted from* [*Henry and Fekete, 2006*]. *Node-link diagram are familiar to most people, helping to understand the less familar marices: Blue and green represent the connectors between clusters in both charts.*

4.3.1 Cheat sheet types

In total, we we created 8 types of cheat sheets (Figures 4.4 and 4.5) for six visualization techniques. All sheets were created using vector drawing tools. The detailed 7-steps design and evaluation methodology is omitted here for clarity but is described in detail in the paper. The types of cheat sheets (for each visualization technique) were:

Anatomy explains the visual elements of a visualization technique, their composition, their specific terminology, and how they relate to the data. The visual components of a visualization can include individual visual marks (in the sense of Bertin [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.).

Construction explains the conceptual idea behind a visualization design and how it encodes data (Figure 4.4-left). It reflects the relation between the data and its visual encoding by demonstrating the process of constructing or providing a familiar metaphor 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, 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.

Well-known relatives highlights other approaches to visualizing the same data (Figure 4.4-right). Relating new knowledge to existing knowledge is fundamental for learning, and there are often complementary techniques for visualizing 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 4.4-right shows a node-link diagram as a relative of adjacency matrices, inspired by Riche et al. [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

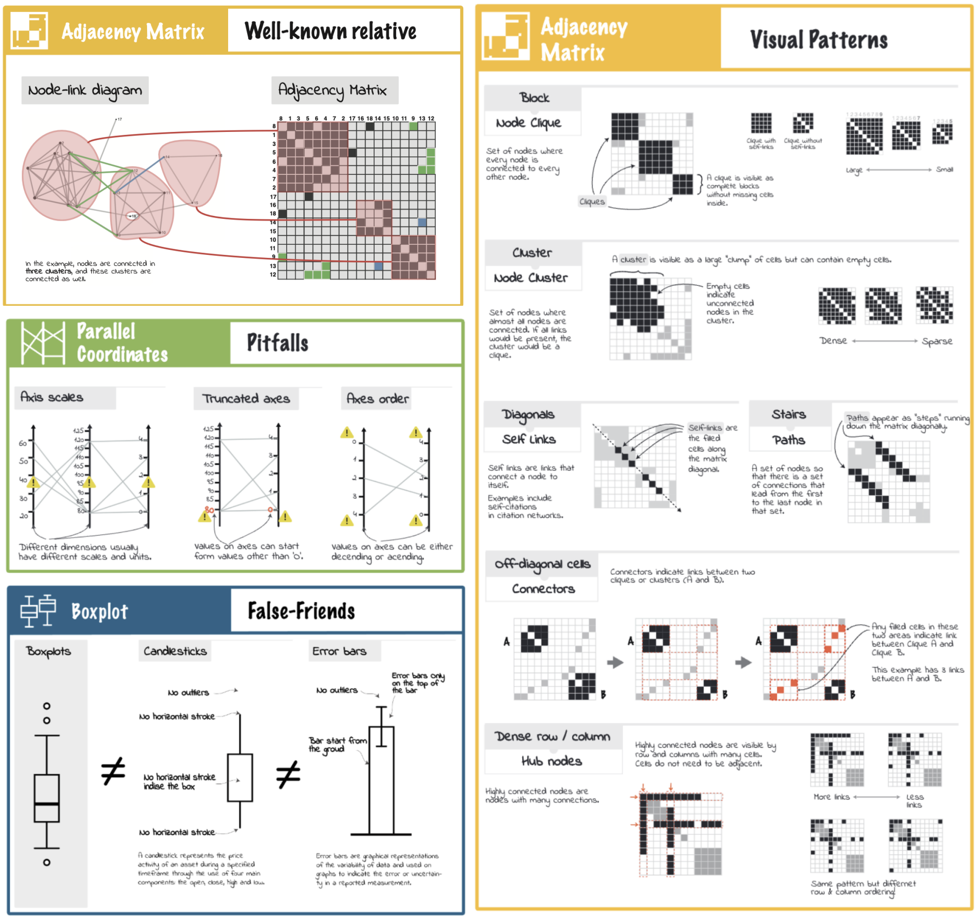


Figure 4.5: *More examples of cheatsheet types.*

meaningful patterns that may occur in a visualization (Figure 4.5-right). Spotting patterns can be key to correctly interpreting visualizations and making discoveries—for example, network cliques in adjacency matrices are visible as solid blocks relies on a good ordering [Behrisch et al., 2016a]). By providing a set of patterns and their explanations to users, cheat sheets can support users in developing understanding using a visualization.

Pitfalls show possible misinterpretations of a visualization (Figure 4.5-left-center). Distinct from general 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: *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 (Figut 4.5-left-bottom). For example, boxplots share visual similarities with candlestick charts and error bars in bar charts; parallel coordinates can look similar to line charts, time curves can look similar to connected scatterplots [Haroz et al., 2016], etc. Our *false-friends* module aims to clarify these differences by placing the 'false-friend' beside the actual visualization, separated by a '$\neq$' sign. For each 'false-friend', we add one sentence to introduce the applications and annotated the different components.

4.3.2 User Evaluation

We conducted a readability study in which participants used cheatsheets to answer questions about three different complex visualisations (*boxplots*, parallel coordinates (*PCP*) and adjacency matrices (*matrix*)). 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. We recruited 11 students with background including cognitive science, psychology, philosophy, linguistics, art, and university administration staff. All of the participants indicated low familiarity with data visualizations.

The main findings of the study include.

- Liker scale ratings showed showed consistently **high and very high rankings for understandability and usefulness**.
- participants found cheat sheets useful for **developing understanding of complex visualisations**.
- Most participants **scored well on the quizzes**. Correctness-rates were 81% for boxplot (distribution of correct answers, 77% for PCP, and 75% for matrix.
- **Usage was highest for the visual patterns type** which was used by 90% of the participants, followed by pitfalls (75%) and anatomy (60%).

- **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.

4.3.3 Design Workshops

We ran a workshop with three educators in visualization to see if our cheat sheet types and examples could serve as blueprints to create more cheat sheets. For the workshop, we prepared design guidelines and visual instructions for creating cheat sheets (we called them '*meta* cheat-sheets'). 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 to create cheat sheets for their classes and any visualization type they would like.

The main findings from the workshops include:

- Using our guidelines, participants **created new cheat sheets** for timelines, connected scatter plots, word clouds, bar charts and Sankey diagrams.
- **Challenges** related to using a limited space efficiently, preparing material for specific audiences (no cheat-sheet to rule them all), and creating construction sheets for abstract concepts such as word clouds
- the process of creating cheat sheets helped them **learn and think** about a visualization technique,

Summary

Visualization cheat sheets are an intervention to address challenges in visualization education, especially around teaching 'remote' audiences (i.e., outside the classroom) and teaching visualization techniques. We present cheat sheets for 6 visualization techniques while showing that more can be created based on our meta-cheat-sheets. Cheat sheets can support independent learning, e.g., the *introduction* introduces basic knowledge about a visualization with comics. Besides as literal sheets, visualziation cheat sheets can be integrated into other forms of presentations such as slide shows (for teaching), videos, infographics or actual visualization applications.

From our studies, we were also able to compile some guidelines for designing visualization cheat sheets:

- **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.

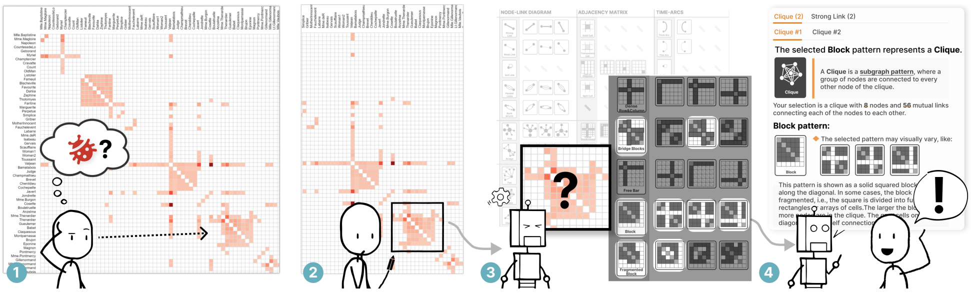


Figure 4.6: *NetworkNarratives helps analysts unfamiliar with network visualizations learn about visual patterns in the representation of their data.* 1 *Looking at the visualization, a user spots a visual pattern of interest, e.g. a "bug"-looking pattern in the matrix. To inquire about whether this pattern is meaningful, the user* 2 *selects the area.* 3 *NetworkNarratives then automatically mines the selection, against a dictionary of network motifs, and* 4 *provides the user with explanations of what underlying network patterns the visual pattern reveals.*

- Use **complementary terminology** to help linking concepts from visual space, abstract data space, and concrete application space.

Besides material for education, we envisioned cheat sheets to support actual exploration and analysis while an expert is using a visualization to explore their data, e.g., in environments such as the Vistorian. For example, in adjacency matrices, *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. This can lead to problems when dealing with *actual* data as it requires mapping form the abstract examples in the cheat sheets to the visual patterns and specific visual encodings in the visualization an expert is investigating.

The next section proposes a specific solution for this scenario problems: *a*) explaining more complex patterns in adjacency matrices, and *b*) assessing cheat-sheets in a more controlled user study.

4.4 Interactive Pattern Explanation for Network Visualizations

This section focuses on learning and visual patterns in network visualizations, such as those used in the Vistorian tool (in particular, adjacency matrices, time-arc diagrams, node-link diagrams). To support onboarding and learning for such visualizations, we propose *interactive pattern explanation*, an approach that explains visual patterns and their corresponding topological network motifs in a user-defined part of the visualization. As shown in Figure 4.6, when exploring an unfamiliar network visualization, 1 a user spots a visual pattern that they do not know how to interpret. The user 2 selects the area of visual interest in that visualization. Our approach automatically 3 mines and retrieves all the underlying network motifs, and 4 presents visual-textual explanations supporting the user's understanding and learning of network patterns in their data. The explanations are similar to visualization cheat sheets, but cheat sheets still require analysts' cognitive efforts to match canonical examples on the sheets with specific cases in the actual visualization. Our premise is that *visualizations are best learned experientially*, i.e. by example and on-the-fly through an analyst exploring and analyzing their own data.

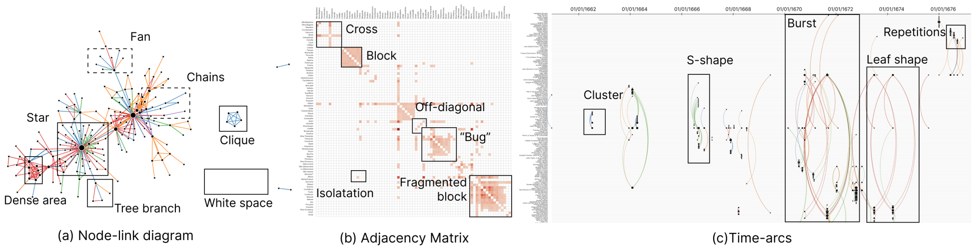


Figure 4.7: *Network visualizations considered in this work and their particular designs. The solid rectangles show some portions participants annotated, while the text around was added by the authors for clarity in the paper, not visible to the participants. The dashed rectangles show some visual patterns the authors annotated. The designs in these visualizations are optimized by the authors of this paper for readability. (a) and (c) use the Marie Boucher Trade network dataset [Dufournaud et al., 2017], and (b) uses the 'Les Misérables' Co-occurrence network dataset [Rossi and Ahmed, 2015].*

For this work, we focused on three representative network visualizations (node-link diagrams, adjacency matrices, and time-arcs) as a first step to study interactive pattern explanations. We designed and implemented a proof-of-concept tool, NetworkNarratives, which can explain a repertoire of 34 visual patterns, across 3 visualization techniques, matching 11 common topological network motifs such as clusters, cliques, or hubs and 2 temporal network motifs. Our 34 patterns were informed by a scoping study with four participants. We compared NetworkNarratives to static cheat sheets in two qualitative/quantitative studies.

4.4.1 Patterns in Network Visualization

We define a **visual pattern** as a *spatially contained, salient configuration of visual marks in a visualization that attracts an observer's attention*. This definition tells us *what* an observer might find interesting (a region in the visualization), independently from *why* they find it interesting (beyond the scope of this paper). This definition shares with other definitions [Koffka, 1922, Andrienko et al., 2022, Collins et al., 2018, Boy et al., 2014b] that there are salient parts in a visualization that somehow stick out and catch a viewer's attention. For example, a set of close points in a node-link diagram can be perceived as a visual pattern (Figure 4.7a) or a set of squares in a matrix visualization (Figure 4.7b) might result in a block. Generally, visual patterns are the opposite of randomness and entropy: they imply structure and information. Often, visual patterns recur within the same or across different datasets, implying some meaning for an analysis.

Importantly, when mapping visual patterns to data patterns (network patterns, in our case), several things can happen:

- In a **one-to-one mapping** where a visual pattern (VP) is mapped to a network pattern (NP) and vice versa: $VP \leftrightarrow NP$. That would mean that the presence of a visual pattern necessarily indicates the presence of a network pattern (and vice versa). Properties of the visual patterns (location, size, structural characteristics) could then be used as a proxy to interpret the data pattern.
- A visual pattern can also be an **artifact** of the layout algorithm or visual encoding and not map to any meaningful pattern in the data: $VP \dashv NP$. Examples of artifacts in matrices include band patterns or the "bugs" that can occur in specific seriation methods but do not relate to any particular network motif [Behrisch et al., 2016a]. Likewise, in Figure 4.7c we can see repetitive dot patterns that again result from node seriation, but do not translate to a meaningful topological feature.

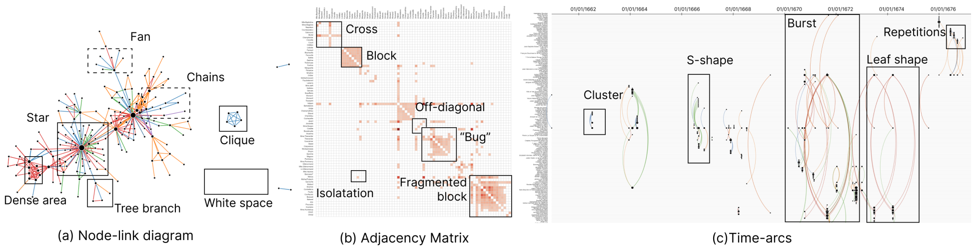


Figure 4.8: *Network visualizations considered in the work and their particular designs. The solid rectangles show some portions participants annotated, while the text around was added by the authors for clarity. The dashed rectangles show some visual patterns the authors annotated. The designs in these visualizations are optimized by the authors of this paper for readability. (a) and (c) use Marie Boucher Trade network dataset [Dufournaud et al., 2017], and (b) uses the 'Les Misérables' Co-occurrence network dataset [Rossi and Ahmed, 2015].*

- Likewise, a visualization can **obscure** a pattern present in the data, i.e., not resulting in any easily perceivable visual pattern $VP \vdash NP$. Examples of such obscurations include cliques or clusters in matrices that happen to not appear as block patterns but instead the individual cells are spread over the matrix (Figure 4.9 5-Clique #1).
- A special case of artifacts are **confusers** where one visual pattern implies multiple possible patterns (explanations) in the data $VP \leftrightarrow \{NP_1, NP_2, ...\}$, including wrong ones. Examples of such cases include links overlapping nodes in a node-link diagram but not connecting to these nodes [Wong and Carpendale, 2007]; or, dense areas in node-link diagrams that originate from incidentally overlapping nodes and links but do not form a topological cluster.
- **Hallucinators** are seemingly different visual patterns that map to the same network pattern $\{VP_1, VP_2, ...\} \leftrightarrow NP$. Examples include any visual pattern impacted by layout, especially orderings in time-arcs where some arcs are short, others are long; or a block in a matrix becoming a split block with a mere reordering of the rows.

In this study, we focused on the three representative network visualization techniques shown in Figure 4.7 and whose designs we have carefully developed to clearly show visual patterns: node-link diagrams, adjacency matrices, and time arcs.

- **Node-link diagrams** (Figure 4.7a): Common visual patterns in node-link diagrams include stars, fans, chains, dense areas. Link color and thickness in our designs represent link types and link weight, respectively. The saliency of many of these visual patterns depends on the node positions in the respective graph layout. In our study, we use the default WebCoLa [Dwyer et al., 2006] layout to minimize the distance between connected nodes and ensure nodes do not overlap and are spaced apart appropriately.
- **Adjacency matrices** (Figure 4.7b) are likely to be less familiar to many people [Ghoniem et al., 2004]. Matrices can create very peculiar visual patterns such as crosses, blocks, fragmented blocks, outliers, lines, or all sorts of insect-like looking patterns along the diagonal (Figure 4.6). Like in node-link diagrams, these visual patterns depend on the algorithm dictating the layout of rows and columns, called the matrix seriation method [Behrisch et al., 2016b, Behrisch et al., 2016a]. In our design, we use the Barycenter heuristic implemented in Reorder.js [Fekete, 2015] to order rows and columns. This algorithm optimizes the ordering of the nodes so that every node is the closest possible to its topological neighbors in the matrix ordering. Cell shading (darkness) maps to link weight.

- **Time-arcs** (Figure 4.7c) visualize temporal networks [Dang et al., 2016] in a Cartesian coordinate system with a time axis (x-axis, left-to-right) and a list of nodes (y-axis). A connection between two nodes in time is shown as an arc connecting two circles placed at the intersection of the node and the point in time when the connection happens. Arcs encode link direction, counter-clockwise from the start to the end node. Visual patterns in time arcs can include overlapping arcs, recurring arcs, rows with lots of dots, sets of arcs resembling a S or leaf shape, clustered arcs, etc. Similar to matrices, the saliency and specific appearance of visual patterns in time-arcs depends on the ordering of nodes on the y-axis. This ordering can make arcs spanning the whole height of the visualization and flip the direction of an arc depending on the link direction. To order nodes along the vertical dimension, we used the same Barycenter seriation method as that for the matrix visualization.

4.4.2 The Pattern Explainer Tool

The pattern explainer interface (Figure 4.9) contains a network visualization and can in theory be wrapped around any network visualization application. Interactive pattern explanation with our interface works as follows (blue-circled labels refer to parts of Figure 4.9a):

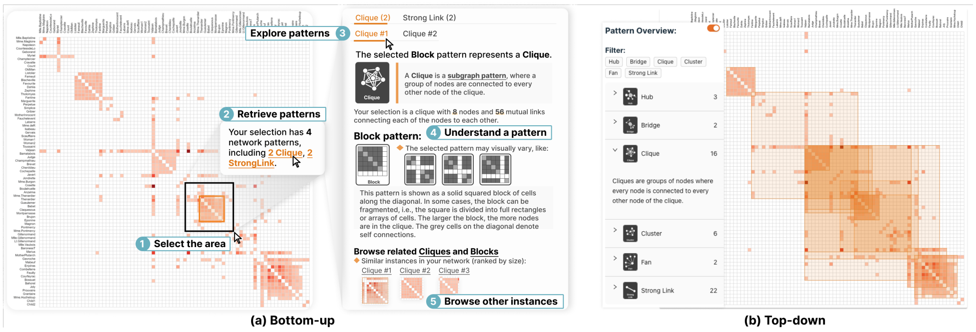


Figure 4.9: *The NetworkNarratives idea. (a) In the bottom-up explanation, a user can select an arbitrary region of interest in a network visualization. Our system retrieves all the underlying network patterns in the user selection, backed by a pattern repository and a set of heuristics, and pops up an overview. After a user selects a pattern for exploration, the pop-up provides visual-textual explanations of the network and visual patterns, and lists other instances for browsing. (b) In the top-down explanation, the user can browse all the found instances according to pattern types.*

- As part of an analysis task, a user explores the visualization and spots a visual pattern which they do not know how to interpret.
- The user then selects the respective part of the visualization through a rectangle or lasso selection. This selection essentially defines a subgraph containing the nodes and/or links represented by the visual marks in the selection.
- Then, NetworkNarratives looks for all possible network patterns in this subgraph, using a pattern repository (see paper for details) and a set of heuristics when a motif qualifies as such (see paper for details). It may happen that the visual pattern selected by the user refers to multiple network patterns.
- The **selector pop-up** shows up with the list of network patterns found, indicating, e.g., Your selection has 4 network patterns, including 2 cliques and 2 strong links. Likely, if no patterns are found (artifacts), a message informs the user, explaining that the selected visual pattern is most likely an artifact.
- Clicking on either of the patterns in the pop-up brings up a second pop-up, the **explainer pop-up**. The explainer pop-up explains the chosen network pattern

as well as the corresponding visual pattern through corresponding visual and textual descriptions.

- After reading the information in the explainer pop-up, a user can continue either by *i*) exploring other instances of the same network pattern in the network, *ii*) choosing another network pattern from the selector pop-up, or *iii*) selecting another visual patterns in the visualization.

We call this process *bottom-up explanation*, because a user starts from a single particular visual pattern in the visualization. NetworkNarratives can also support *top-down explanation* in which NetworkNarrativesfirst retrieves *all* network patterns in the entire visualization, and highlights them alongside a summary of network patterns and their quantities (Figure 4.9b). In the top-down explanation, a user can immediately start browsing patterns, complementary to bottom-up explanation. A toggle button in the interface switches between both explanation modes.

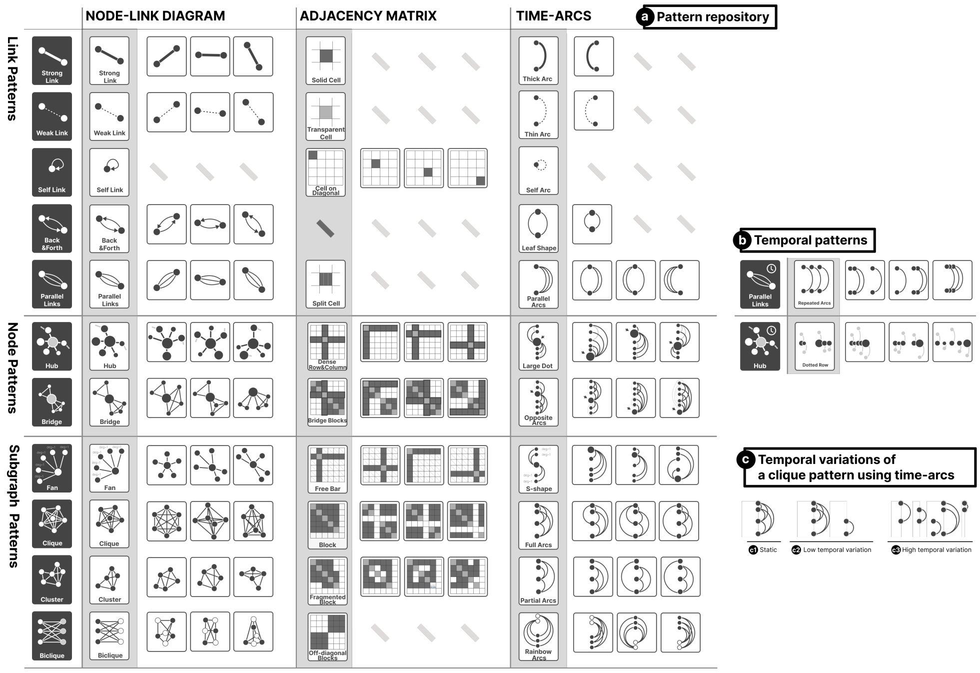


Figure 4.10: *An overview of the pattern repository (a, b). Network patterns (black background) are organized vertically, while corresponding visual patterns (white background) are listed horizontally in each of the three visualizations. Icons for visual patterns include one lead icon (gray background) and several smaller versions of visual variations. (c) illustrates temporal variations of a clique pattern in time-arcs.*

4.4.3 Pattern detection

NetworkNarrativescurrently can find 11 network patterns and 34 visual patterns across node-link diagrams, matrices, and time-arcs (Figure 4.10). Those patterns are mined by an underlying system in the back-end of NetworkNarratives, using state-of-the art graph mining algorithms and heuristics. For example, if the user selects a dense area in the matrix, it its density is calculated. If the density is higher than a threshold we set, it qualifies as a cluster. If the area is complete, it is returned as a clique. Likewise, we do with all the other patterns in Figure 4.10. If a visual pattern cannot be matched to a network pattern, a message alerts the user. If a pattern is matched to multiple network patterns (confusers), the user is informed and can chose which ones they are are interested in. The same happens if a selection contains multiple sub-patterns, e.g., two cliques.

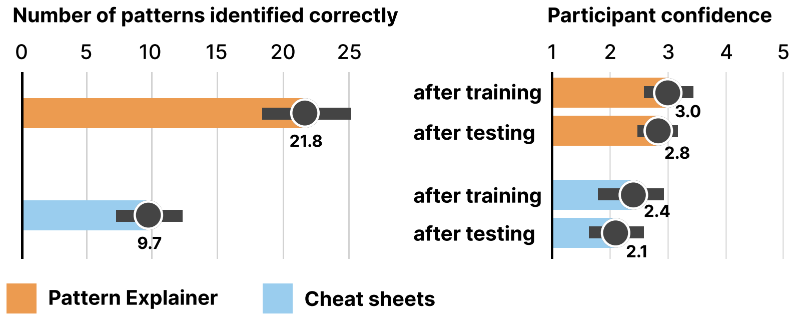


Figure 4.11: *Results from the quantitative controlled study with the number of patterns correctly identified (left, means) as well as confidence after training and after testing (right). Error bars indicate 95% confidence intervals.*

Further details on the user interface design, rationale, the creation and selection of patterns and their computation are found in the paper.

4.4.4 Qualitative and quantitative studies

We conducted two complementary user studies to collect qualitative and quantitative data about how participants use pattern explanation to learn network visualizations. We hypothesized interactive explanation increases the learning of visualizations and network patterns, compared to baseline techniques such as text-only explanations or static textual-visual explanations on visualization cheatsheets.

In the **qualitative study (study 1)**, we asked 12 participants to answer questions on visualizations while being trained with three different conditions: *a*) textual explanations of visual encoding, *b*) cheat sheets, and *c*) Pattern Explainer. We only provide the bottom-up mode (Figure 4.9a) of the system in the study to understand the effects of our interactive by-example and on-the-fly explanations, in comparison with static and premeditated cheat sheets. Each participant were asked to investigate each visualization with a different explanation technique and to compare their experience.

In the **controlled user study (study-2)**, we measured how many patterns people can accurately identify using either cheat sheets or interactive pattern explanations (between-subjects design). We removed the textual description condition as they were deemed less effective in the qualitative study 1 and focused on adjacency matrices only since they were a little more complicated than node-link diagrams. 20 participants were chosen having stated least familiarity with the three visualizations in a pre-study questionnaire. After training, we then took away the learning intervention and asked participants to identify and annotate *as many patterns as possible* on a shared digital whiteboard (Figma) in three other matrices, for 2 minutes each. We asked participants about their confidence and measured (*1*) the number of patterns participants identified correctly (coded by two of the authors) after training, and (*2*) their pre- and post-confidence.

Across both studies, we found:

- with interactive Pattern Explainer, participants correctly found an average of 21.8 patterns; with cheat sheets it was only around 9.7 patterns (Figure 4.11).
- interactive Pattern Explainer was praised for filling the gap between abstract concepts and practical examples of patterns.
- non-experts appreciated the in-context and on-the-fly explanations.
- abstract pattern ideas would have a terminology and could be applied to the actual data.

- explainer helped interpret visually cluttered areas, and (in)validated assumptions, and explore similar patterns..
- Cheat sheets, in contrast, were perceived as a more comprehensive list of patterns to look up and gain an overview over the dictionary of patterns, their meaning, and their complexity.
- many of our participants reported being overwhelmed by the information shown on cheat sheets and voiced difficulties mapping visual patterns they perceived in the visualization to the abstract concepts on the cheat sheets
- A concern raised by the participants was potential over-reliance on the automated pattern explanation.

Summary

- both approaches—**interactive pattern explanation and cheat sheets can be complementary techniques**, and we should explore ways to combine both in interfaces for visualization exploration. For example, NetworkNarratives's top-down mode could be combined with cheat sheets to provide an overview of patterns and information in a visualization.
- Composing explanations for patterns should consider users' background knowledge and diverse domains, e.g., their specific network patterns, terminologies, and examples.
- we should think of **more guided and progressive tutorials** to visual patterns to guide people through reading a visualization properly, learn to spot and interpret visual patterns.
- we strongly believe that interactive pattern explanation has **potential beyond network visualizations**. So far, we have seen patterns explicitly discussed only for geographic data and map visualizations [Andrienko et al., 2022].

Visual patterns—a critique:

While seemingly a powerful thinking and analysis tool, there is a risk in overemphasizing the use and training for (visual) patterns since there might be the **danger of developing a narrow mind** and **over-reliance on the system**. There might be good reasons why the literature on visualization has not been discussing patterns in a more formal sense and why it has been restrained from defining pattern repositories. The open and fuzzy nature of patterns is what might make them useful—as a thinking tool and conceptual aid to approach teaching visualization techniques and thinking with visualization—rather than a formal dogmatic way of parsing visualizations. Because visualizations are hard to describe by composition and dissection: they need to be open in the data patterns they want to show instead of prescribing what to show; they need to give *some* perspectives on the data, show messiness where there is in the data and invite for human interpretation and inquiry—new visual patterns might emerge from the rules of the visual design and people will realize those patterns exist. In the best case, humans, after learning these initial patterns with NetworkNarratives, can question and refine their mental model for visualization, thinking and reasoning.

In conclusion, we **could ponder a more comprehensive 'critical theory of patterns'** in visualization. Not a theory that meticulously wants to cartograph any existing pattern or create a dogmatic framework. Rather, a theory that facilitates to describe, analyze, evaluate, and communicate visual patterns in visualization and provides the necessary terminology and methodological tools. A theory that describes visual patterns as essential building blocks of (complex) visualizations and offers keys to their interpretation. We could use such a concept of patterns to scrutinize a visualization (design) and algorithm and assess both visual artifacts and hallucinators; we could think about ways to make specific patterns more salient in a given design by tweaking its design; we could suggest visualization designs that are best suited to show specific patterns; and we could think about guiding novice users in understanding visualizations and interactive visual analysis. Such a theory could become a structured way to support engagement with visualization design and visualization education in a **critical and creative way** and to show the potentials and pitfalls of data visualization more generally.

4.5 Chapter Summary

This chapter started by discussing challenges in understanding visualizations and then presented two possible interventions to overcome some of these challenges.

We can see that there is an enormous space of unsolved challenges that risks being ignored in the search for new techniques, algorithms, and interfaces. Yet, as HCI researchers, we need to understand these barriers and built solutions into our interfaces. We need to understand how to best explain our techniques and how to communicate the workflows and strategies they imply. It is true that the motivated person will make the effort to master any technology (e.g., learn programming on their own), but we can only solve make progress in visualization and solve more complex problems if we work together with them.

Yet, in-the-wild research is inherently challenging [Shneiderman and Plaisant, 2006, Bonsignore et al., 2009]. Our findings are situated in the specifics of our visualization course (Section 4.1) and the Vistorian tool, as well as those participants respective backgrounds. For example analysts may have enrolled because they had particular conceptual challenges in network visualization and hence we cannot say how prominent our observed barriers are in other settings (e.g., among analysts with different backgrounds and training). Similarly, those participants who engaged in discussions and interviews might not be those who fail *silently* for reasons we were not able to capture. We need better methods to collect data about the use of visualization tools and interventions (Sections 4.3 and 4.4) to capture thinking processes and problems.

Still, visualization is lacking a theory of education and an established body of knowledge of its own. As shown in Section 4.2, we need to work on foundations for how to best teach and learn visualization, taken into account the many contexts in which visualization appears and the diverse skills it implies: teaching designers will require different approaches than teaching engineers, researchers, or business analysts. We also need assessment tests capturing the wide range of skills involved in visualization. While we have visualization tests for chart reading [Lee et al., 2017, Ge et al., 2023], we lack *a*) tests that capture things like 'creative problem solving', 'critical understanding' or 'ability for interactive exploration'. Likewise, we lack tests *b*) that can be deployed remotely to assess how someone is using a tool (see Section

4.1) and where they fail.

And yet, I think technology can help to guide people along their learning and understanding and critical thinking. *How could personal tutoring systems look for visualization? For design, exploration, analysis, communication? Which guidelines and lessons would they convey? How would they communicate with the people to explain visual patterns, interactions, fallacies? What would they base their assessment on?* Again, answering these questions, will require a theory of education alongside guidelines, assessment tests, and benchmark data and visualizations.

Chapter 5

Discussion and Directions

I set out this thesis with the two questions *How to support people present and explain complex data?* and *How to help them create and use visualization effectively?* I motivated these questions from two representative use cases of applied research that I worked on in the past years and that have strongly influenced my research questions through with observations, hands-on experiences and discussions. These use cases—network visualization and the visualization of peace and conflict data (Section 1.2)—can be described as wicked problems because they involve a multitude of goals, tasks, users, approaches as well as large data sets about complex systems. Visualization is crucial to help address the respective problems in understanding complex networks and highly dynamic social processes. The individual contributions in this thesis present solutions and explorations to address challenges around such wicked problems and some of the Nine E's I sketched out in the introduction: *extensiveness, explanation, exploration, evolution, expertise, expectations, exceptions, embargoes*, and *education*. The remainder of this section reflects on those findings and sketches future avenues of research.

5.1 Novel forms for storytelling

Each form describe in Chapter 2—data comics (Section 2.1), dashboards (Section 2.2), visualization atlases (Section 2.3) presents a unique combination of means for narration, interaction, the structure of information, and the media. All of these forms are complementary in that they present distinct strategies for exploration and explanation of extensive data; comics focus on explanation and provide a surprisingly rich design space for linear and non-linear, visual and textual narration. They can make use of illustrations and adapted to more educational or more exploratory scenarios; they can explain expectations and specific data contexts in detail, including reasons for not revealing certain information (embargoes). Dashboards address concerns about extensiveness by aggregating and abstracting data into KPIs, presented in ways that engage overview, exploration as well as contextualize evolving data. Data articles are in-depth discussions that can deal with evolving data, cater to specific expertise and help education in a similar form than data comics. Eventually, Atlases form an umbrella structure to manage different other forms of storytelling. In that, they provide navigation and structure to manage extensive data sets, combine explanation with exploration, account for the evolution of data and different expertise and expectations of their audiences within educational or non-educational scenarios.

As technology and application scenarios progress, new forms for visualization, narration, and engagement will be necessary, leveraging the power of visualization, interaction, storytelling, time, content, tangibility and virtual and real space in new ways to create engaging experiences. Some directions could be:

- **Narrative dashboards** would integrate explanation into dashboards aiming to display and communicate data to a wide audiences. Narration is required for understanding the information shown and to translate patterns in the data into meaningful information. The Covid-19 pandemic has brought us dashboards in the public, but those dashboards were accompanied by numerous analysis, stories and explanations. Narrative dashboards would integrate these stories into the context of the dashboards, explaining the data, its origin and analysis, emerging patterns as a situation progresses. Narrative dashboards could have a great potential to monitor the environmental crisis and make the respective data, measures, and observations part of a common culture as it happened with Covid-19. Other application scenarios could including personal and public spending's, opinions and polls, and conflicts. While dashboards exist for each of these scenarios, they lack explanations and onboarding and the challenge with narrative dashboards is to explain *for* exploration—similar to atlases—with the right balance of explanation and exploration but with the intrinsic goal to allow their audience to monitor and overview an evolving situation. Some research questions include *How we can integrate sequential narration into otherwise non-sequential dashboards? How do guide people's attention and create affordances for people to explore the data? How do they encourage exploration and enable personalization?*
- **Narrative Hyperwalls**. Hyperwalls have numerous applications for interactive exploration[Belkacem et al., 2022] but surprisingly little to no work exists exploring their potential for communication and education. However, NASA's Earth Information Dashboard[1] aims at monitoring and presenting data about the earth's environment to a general public for explanation and education. A large display space makes *a*) a lot of content available at the same time, and *b*) creates a almost intimate space of co-knowledge for the people observing the hyperwall at the same time. Other than online dashboards, hyperwalls create situated and synchronous experiences for the their audience. With the help of techniques from narrative dashboards, they can invite to come close, co-explore, discuss and share what they show *How can these experiences of 2-dimensional display space and social interaction space be leveraged for communication and understanding? How do hyperwalls use their enormous space for storytelling? How do they integrate textual or auditory annotations? What can we learn from existing storytelling forms for the design of narrative hyperwalls and its interaction techniques?*
- **Immersive Storytelling**. While hyperwalls remain 2-dimensional and constraint to the wall they are set up, virtual and mixed reality provide further yet unexplored potential for storytelling. They create three-dimensional experiences immersing a person, showing three-dimensional content, allowing for content to be projected into the real world and augment displays, spaces, and tangible artifacts. To date, there are no holistic and structured approaches to explore immersive data-driven storytelling, but rather those are reflections on how storytelling can be made more immersive in the sense of engagement [Isenberg et al., 2018]. Approaches specific to virtual and mixed reality include data visceralization [Lee et al., 2020] in VR or showing waste accumulation in mixed reality [Assor et al., 2024]. *What data would we show? How would we design narrations through text or and audio? How to walk a person through the virtual space and direct their attention to the explanations?*

[1] https://svs.gsfc.nasa.gov/gallery/earth-now-dashboard

5.2 Story authoring and visualization design

Chapter 3 discussed the authoring tools to create and democratize these forms. Design tools are needed to support more people create visualizations and include their expertise in wicked problems into these visualizations. There are still too few tools for effective storytelling in the form of comics, videos, dashboards or even visualization atlases. Tools are also missing for designing exploration in the form of more complex visualizations, interactions, and complementary views. NetPanorama aims to close this gap for network visualizations. How would we need to design these future tools?

Interoperability: Designing authoring tools for one visualization form at a time is a necessity to remain expressive and usable. However, the next generation of visualization authoring tools actually has to be both authoring platforms providing access to lots of different visualization forms as well as become interoperable with exploratory analytics systems. The same data story often needs to be told to different people, different contexts (e.g., mobile, hyperwall), and different purposes (e.g., overview, monitoring, exploration). Ideally, we'd convert between comics and videos or dashboards and infographics either at the time of authoring or the stage of reading. We also need to make sure presentation forms and stories remain updatable when data updates or remain easily amenadable with new facts.

Personalization is key to effective stories and interfaces; they adapt to the audience to increase relevance and understanding. Personalization in storytelling would be able to deliver stories, automatically generated, to individuals according to their interest and task. In fact, there probably is a gradual continuum between storytelling and exploration in this case, since to ability to craft a potentially huge amount of alternative stories is akin to exploring different aspects of the same dataset.

Guided design could mean the guidance of an author through the process of designing a visualization. A tool could support the user in exploring possibilities, making informed suggestions, and rationalizing its decisions. A cornerstone of such systems would be formalizable guidelines that can be interpreted and applied in a given context. For data-driven storytelling and the more complex forms such as dashboards and comics, guidelines would have to be more high-level and sophisticated than in current systems [Moritz et al., 2019]. One of the main problems here is that too little empirical research exists on data-driven storyteling interfaces, in order establish such guidelines in the first place.

Generative AI will deliver many possibilities to improve productivity for designing and creating visualizations and stories; either by creating code, analyzing data, creating examples, reading and interpreting visualizations, or advising on design guidelines [Kim et al., 2024]. However, if there is one achievement AI is making for visualization, I think, it is that we acknowledge the truly human nature of visualization, the creativity, problem solving, interpretation, questioning, decision-making, participation, communication, and collaboration.

5.3 Visualization Education

The challenges described in section 4.2 paint a picture of the research necessary to provide for a theory of education in visualization (Section 4.2): learning goals, materials, assessment, tools, the role of AI, etc.

- **Intelligent tutoring systems.** I see a huge potential for intelligent tutoring systems that help personalize learning and guide their users. Interactive pattern explanation (Section 4.4) is a simple method to personalize learning but open questions concern how we can personalize the experience with entire systems and tools, such as the Vistorian, or workflows full of barriers such as described in section 4.1? We would need to understand what a person is wanting to achieve, understand their their pre-knowledge, explain strategies explain user interfaces and visualizations and monitor progress and guide. In tools like the Vistorian, we could start with a constraint set of visualizations, then gradually extend the set of visualizations as users learn the basic ones. Or, we can help chose visualizations based on their task at hand. Likewise, we can leverage storytelling forms for explaining algorithms for analysis and visualization such as graph layouts and matrix 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 practice will help to draw out commonalities.

- **Explorable explanations** are a concept defined by Brett Victor (`https://worrydream.com/ExplorableExplanations`) and which have more recently been cited as inspirations for interactive documents [Conlen and Heer, 2018] and explorable statistical analyses [Dragicevic et al., 2019]. I see a huge potential for explorable explanations in visualization education, e.g., to explore visualization and interaction techniques.

5.4 The next wicked problem(s)

Wicked problems are fascinating for (visualization) research because there are no simple solutions. Instead, they offer plenty of opportunity for creativity and tinkering to hopefully make some impact. Climate change and climate action is prime example of a wicked problem [Perry, 2015], the field already benefits from visualization through storytelling,[2] dashboards,[3] a range of iconic images (e.g. climate stripes[4], the climate spiral[5], the hockey-stick chart [Mann et al., 1999] and heaps of public visualizations such as the 6th IPCC report [Pörtner et al., 2022], news,[6] art[7] and scientific communicative reports [Pörtner et al., 2022]. Likewise, NASA and MIT are creating extensive museum-like platforms and experiences for their Earth Now dashboards [Kostis et al., 2022][8] and the Earth Mission Control project [9] lead by former NASA deputy administrator Dava Newman. In research, most of the work has been scattered across many different domains such as climate and environmental research, science communication [Windhager et al., 2019], domain specific applications, scientific storytelling [Ma et al., 2011], and physicalization [Sauvé et al., 2023]. Yet, there are many challenges related to the works discussed in this thesis and the Nine E's: *How to balance visual complexity, depth of information, and visual/data literacy? How to understand abstract and widely unfamiliar scales of time, space, and numbers [Chevalier et al., 2013]? How to make model projections more accessible and actionable for the general public? How to work with policymakers and communities at risk? How to support decision making on a personal as well as collaborative level? How to train the creators of visualization?*

[2] `https://ig.ft.com/sites/climate-change-calculator`
[3] `https://www.zeit.de/wirtschaft/energiemonitor-strompreis-gaspreis-erneuerbare-energien-ausbau`
[4] `https://showyourstripes.info`
[5] `https://svs.gsfc.nasa.gov/5190`
[6] `https://www.climatecentral.org`
[7] `https://art21.org/read/inigo-manglano-ovalle-climate`
[8] `https://svs.gsfc.nasa.gov/gallery/earth-now-dashboard`
[9] `https://www.media.mit.edu/projects/earth-mission-control/overview`

Our community is yet to develop an overarching theory of practice and research agenda around climate visualization. To achieve this, we co-organize the first IEEE VIS Workshop on Visualization for Climate Action and Sustainability in 2024 [Bach et al., 2024] to examine what exactly is specific about data visualization for climate action and sustainability and how working in this field can form and impact our research agendas. I am extremely excited about the event and the discussions we will have and how this will shape the future of both our approach to climate action and our approach to using visualization for the good of the people and the planet.

Bibliography

[ide, 2016] (2016). Ideo Cards. online: `https://www.ideo.com/work/method-cards`. [last visited: Nov. 29, 2016].

[Adar and Lee-Robbins, 2023] Adar, E. and Lee-Robbins, E. (2023). Roboviz: A game-centered project for information visualization education. *IEEE TVCG*, 29(1):268–277.

[Ahn et al., 2013] Ahn, J.-w., Plaisant, C., and Shneiderman, B. (2013). A task taxonomy for network evolution analysis. 20(3):365–376.

[Aleixo and Sumner, 2017] Aleixo, P. A. and Sumner, K. (2017). Memory for biopsychology material presented in comic book format. *Journal of Graphic Novels and Comics*, 8(1):79–88.

[AlKadi et al., 2023] AlKadi, M., Serrano, V., Scott-Brown, J., Plaisant, C., Fekete, J., Hinrichs, U., and Bach, B. (2023). Understanding barriers to network exploration with visualization: A report from the trenches. *IEEE Trans. Vis. Comput. Graph.*, 29(1):907–917.

[Allison et al., 2021] Allison, J., Badanjak, S., Bach, B., Bell, C., Bhattacharya, D., Knaussel, F., and Wise, L. (2021). An interactive tracker for ceasefires in the time of covid-19. *The Lancet Infectious Diseases*, 21(6):764–765.

[Alper et al., 2013] Alper, B., Bach, B., Henry Riche, N., Isenberg, T., and Fekete, J.-D. (2013). Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems*, pages 483–492. ACM.

[Alsallakh et al., 2015] Alsallakh, B., Micallef, L., Aigner, W., Hauser, H., Miksch, S., and Rodgers, P. (2015). The state-of-the-art of set visualization. *CGF*, 35(1):234–260.

[Amini et al., 2015] Amini, F., Henry Riche, N., Lee, B., Hurter, C., and Irani, P. (2015). Understanding Data Videos: Looking at Narrative Visualization Through the Cinematography Lens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1459–1468.

[Amini et al., 2016] Amini, F., Riche, N. H., Lee, B., Monroy-Hernandez, A., and Irani, P. (2016). Authoring data-driven videos with dataclips. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):501–510.

[Andrienko et al., 2022] Andrienko, N., Andrienko, G., Chen, S., and Fisher, B. (2022). Seeking patterns of visual pattern discovery for knowledge building. 41(6):124–148.

[Apache ECharts,] Apache ECharts. Apache ECharts. `https://echarts.apache.org/`. Accessed: 2020-12-04.

[Asimov, 1985] Asimov, D. (1985). The grand tour: A tool for viewing multidimensional data. *SIAM Journal on Scientific and Statistical Computing*, 6(1):128–143.

[Assor et al., 2024] Assor, A., Prouzeau, A., Dragicevic, P., and Hachet, M. (2024). Augmented reality waste accumulation visualizations. *ACM Journal on Computing and Sustainable Societies*, 2(2):1–29.

[Auber et al., 2017] Auber, D., Archambault, D., Bourqui, R., Delest, M., Dubois, J., Lambert, A., Mary, P., Mathiaut, M., Melançon, G., Pinaud, B., Renoust, B., and Vallet, J. (2017). TULIP 5. In *Encyclopedia of Social Network Analysis and Mining*, pages 1–28. Springer.

[Bach et al., 2022a] Bach, B., Abdul-Rahman, A., and Diehl, A. (2022a). 4th IEEE workshop on visualization guidelines in research, design, and education. `https://visguides-workshop.github.io/`.

[Bach et al., 2023] Bach, B., Carpendale, S., Hinrichs, U., and Huron, S. (2023). Visualization empowerment: How to teach and learn data visualization (dagstuhl seminar 22261). In *Dagstuhl Reports*, volume 12. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[Bach et al., 2024] Bach, B., Chevalier, F., Kostis, H.-N., Subbaro, M., Jansen, Y., and Soden, R. (2024). Ieee vis workshop on visualization for climate action and sustainability. *arXiv preprint arXiv:2404.02743*.

[Bach et al., 2022b] Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., and Chen, M. (2022b). Dashboard design patterns. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):342–352.

[Bach et al., 2015a] Bach, B., Henry-Riche, N., Dwyer, T., Madhyastha, T., Fekete, J.-D., and Grabowski, T. (2015a). Small multipiles: Piling time to explore temporal patterns in dynamic networks. In *Computer Graphics Forum*, volume 34, pages 31–40.

[Bach et al., 2015b] Bach, B., Henry Riche, N., Fernandez, R., Giannisakis, E., Lee, B., and Fekete, J.-D. (2015b). NetworkCube: Bringing Dynamic Network Visualizations to Domain Scientists. Posters of InfoVis. Poster.

[Bach et al., 2021] Bach, B., Huron, S., Hinrichs, U., Roberts, J. C., and Carpendale, S. (2021). Special issue on visualization teaching and literacy. *IEEE Comput. Graph. Appl*, 41(06):13–14.

[Bach et al., 2016a] Bach, B., Kerracher, N., Hall, K. W., Carpendale, S., Kennedy, J., and Henry Riche, N. (2016a). Telling stories about dynamic networks with graph comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3670–3682. ACM.

[Bach et al., 2013] Bach, B., Pietriga, E., and Fekete, J.-D. (2013). GraphDiaries: Animated transitions andtemporal navigation for dynamic networks. 20(5):740–754.

[Bach et al., 2014] Bach, B., Pietriga, E., and Fekete, J.-D. (2014). Visualizing dynamic networks with matrix cubes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 877–886. ACM.

[Bach et al., 2011] Bach, B., Pietriga, E., Liccardi, I., and Legostaev, G. (2011). Ontotrix: a hybrid visualization for populated ontologies. In *Proceedings of the 20th international conference companion on World wide web*, WWW '11. ACM.

[Bach et al., 2017a] Bach, B., Riche, N. H., Carpendale, S., and Pfister, H. (2017a). The emerging genre of data comics. *IEEE computer graphics and applications*, 37(3):6–13.

[Bach et al., 2017b] Bach, B., Riche, N. H., Hurter, C., Marriott, K., and Dwyer, T. (2017b). Towards unambiguous edge bundling: Investigating confluent drawings for network visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):541–550.

[Bach et al., 2015c] Bach, B., Shi, C., Heulot, N., Madhyastha, T., Grabowski, T., and Dragicevic, P. (2015c). Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568.

[Bach et al., 2016b] Bach, B., Shi, C., Heulot, N., Madhyastha, T., Grabowski, T., and Dragicevic, P. (2016b). Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568.

[Bach et al., 2018a] Bach, B., Wang, Z., Farinella, M., Murray-Rust, D., and Henry Riche, N. (2018a). Design Patterns for Data Comics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 38:1–38:12.

[Bach et al., 2018b] Bach, B., Wang, Z., Nathalie Henry Riche, M. F., Murray-Rust, D., Carpendale, S., and Pfister, H. (2018b). online: retrieved from `http://datacomics.net`.

[Badam et al., 2018] Badam, S. K., Liu, Z., and Elmqvist, N. (2018). Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):661–671.

[Baddeley, 1997] Baddeley, A. D. (1997). *Human memory: Theory and practice*. Psychology Press.

[Ballentine, 2022] Ballentine, B. (2022). Digital humanities and technical communication pedagogy: a case and a course for cross-program opportunities. *Commun. Des. Q. Rev*, 10(1):24–37.

[Bartram et al., 2021] Bartram, L., Correll, M., and Tory, M. (2021). Untidy data: The unreasonable effectiveness of tables. *arXiv preprint arXiv:2106.15005*.

[Bastian et al., 2009] Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In *International AAAI Conference on Weblogs and Social Media*.

[Bateman et al., 2010] Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., and Brooks, C. (2010). Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In *Proc. CHI*, pages 2573–2582. ACM.

[Baur, 2008] Baur, M. (2008). *visone - Software for the Analysis and Visualization of Social Networks*. PhD thesis, Karlsruhe Institute of Technology.

[Beasley et al., 2021] Beasley, Z. J., Friedman, A., and Rosen, P. (2021). Through the looking glass: Insights into visualization pedagogy through sentiment analysis of peer review text. *IEEE Comput. Graph. Appl.*, 41(6):59–70.

[Beck et al., 2017] Beck, F., Burch, M., Diehl, S., and Weiskopf, D. (2017). A taxonomy and survey of dynamic graph visualization. In *Computer graphics forum*, volume 36, pages 133–159. Wiley Online Library.

[Behrisch et al., 2016a] Behrisch, M., Bach, B., Henry Riche, N., Schreck, T., and Fekete, J.-D. (2016a). Matrix reordering methods for table and network visualization. In *Computer Graphics Forum*, volume 35, pages 693–716. Wiley Online Library.

[Behrisch et al., 2016b] Behrisch, M., Bach, B., Hund, M., Delz, M., Von Rüden, L., Fekete, J.-D., and Schreck, T. (2016b). Magnostics: Image-based search of interesting matrix views for guided network exploration. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):31–40.

[Belkacem et al., 2022] Belkacem, I., Tominski, C., Médoc, N., Knudsen, S., Dachselt, R., and Ghoniem, M. (2022). Interactive visualization on large high-resolution displays: A survey. In *Computer Graphics Forum*, page e15001. Wiley Online Library.

[Bell, 2024] Bell, C. (2024). *PeaceTech: Digital Transformation to End Wars*. Springer Nature.

[Bernaerts, 2006] Bernaerts, A. (2006). Booklet on naval war changes climate.

[Bernard et al., 2018] Bernard, J., Sessler, D., Kohlhammer, J., and Ruddle, R. A. (2018). Using dashboard networks to visualize multiple patient histories: a design study on post-operative prostate cancer. *IEEE Transactions on Visualization and Computer Graphics*, 25(3):1615–1628.

[Bertin et al., 1983] Bertin, J., Berg, W. J., and Wainer, H. (1983). *Semiology of graphics: diagrams, networks, maps*, volume 1. University of Wisconsin press Madison.

[Bertling et al., 2021] Bertling, J. G., Hodge, L., and King, S. (2021). The case for data visualization in the art classroom. *Art Education*, 74(2):44–49.

[Bezerianos et al., 2010] Bezerianos, A., Chevalier, F., Dragicevic, P., Elmqvist, N., and Fekete, J.-D. (2010). Graphdice: A system for exploring multivariate social networks. *CGF*, 29(3):863–872.

[Bigelow et al., 2019] Bigelow, A., Nobre, C., Meyer, M., and Lex, A. (2019). Origraph: Interactive network wrangling. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE.

[Bigelow et al., 2020] Bigelow, A., Williams, K., and Isaacs, K. E. (2020). Guidelines for pursuing and revealing data abstractions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1503–1513.

[Biggs and Tang, 2011] Biggs, J. and Tang, C. (2011). *Teaching for Quality Learning at University*. Open University Press.

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

[Bonsignore et al., 2009] Bonsignore, E. M., Dunne, C., Rotman, D., Smith, M., Capone, T., Hansen, D. L., and Shneiderman, B. (2009). First steps to netviz nirvana: Evaluating social network analysis with NodeXL. In *Proceedings of the 2009 International Conference on Computational Science and Engineering*. IEEE.

[Borkin et al., 2013] Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., and Pfister, H. (2013). What makes a visualization memorable? *IEEE transactions on visualization and computer graphics*, 19(12):2306–2315.

[Borner, 2010] Borner, K. (2010). *Atlas of science: Visualizing what we know*. Mit Press.

[Börner et al., 2016] Börner, K., Maltese, A., Balliet, R. N., and Heimlich, J. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213.

[Bostock et al., 2011a] Bostock, M., Ogievetsky, V., and Heer, J. (2011a). D^3: Data-Driven Documents. *IEEE TVCG*, 17(12):2301–2309.

[Bostock et al., 2011b] Bostock, M., Ogievetsky, V., and Heer, J. (2011b). D^3 Data-Driven Documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309.

[Boucher et al., 2023] Boucher, M., Bach, B., Stoiber, C., Wang, Z., and Aigner, W. (2023). Educational data comics: What can comics do for education in visualization? In *2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)*, pages 34–40. IEEE.

[Boy et al., 2015] Boy, J., Eveillard, L., Detienne, F., and Fekete, J.-D. (2015). Suggested interactivity: Seeking perceived affordances for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):639–648.

[Boy et al., 2014a] Boy, J., Rensink, R. A., Bertini, E., and Fekete, J.-D. (2014a). A Principled Way of Assessing Visualization Literacy. *IEEE TVCG*, 20(12):1963–1972.

[Boy et al., 2014b] Boy, J., Rensink, R. A., Bertini, E., and Fekete, J.-D. (2014b). A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972.

[Brandes et al., 2003] Brandes, U., Dwyer, T., and Schreiber, F. (2003). Visualizing related metabolic pathways in two and a half dimensions. In *International Symposium on Graph Drawing*, pages 111–122. Springer.

[Brandes and Nick, 2011] Brandes, U. and Nick, B. (2011). Asymmetric relations in longitudinal social networks. *IEEE transactions on visualization and computer graphics*, 17(12):2283–2290.

[Brandt et al., 2014] Brandt, M. J., IJzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., et al. (2014). The replication recipe: What makes for a convincing replication? *J. of Experimental Social Psychology*, 50:217–224.

[Brewer and Harrower,] Brewer, C. and Harrower, M. Colorbrewer. `https://colorbrewer2.org/`. Accessed: 2020-12-04.

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

[Buono et al., 2021] Buono, P., Ceriani, M., and Costabile, M. F. (2021). Hypergraph data analysis with paohvis. In *SEBD*, pages 160–167.

[Burch and Diehl, 2008] Burch, M. and Diehl, S. (2008). TimeRadarTrees: Visualizing dynamic compound digraphs. *CGF*, 27(3):823–830.

[Burch et al., 2011] Burch, M., Vehlow, C., Beck, F., Diehl, S., and Weiskopf, D. (2011). Parallel edge splatting for scalable dynamic graph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2344–2353.

[Burns et al., 2023] Burns, A., Lee, C., Chawla, R., Peck, E., and Mahyar, N. (2023). Who do we mean when we talk about visualization novices? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–16.

[Cagle, 2019] Cagle, S. (2019). Humans have made 8.3bn tons of plastic since 1950. `https://www.theguardian.com/us-news/2019/jun/23/all-the-plastic-ever-made-study-comic`.

[Cairo, 2012] Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.

[Carpendale, 2003] Carpendale, M. S. T. (2003). Considering visual variables as a basis for information visualisation. Technical report, University of Calgary.

[Carr, 1999] Carr, D. (1999). Guidelines for designing information visualization applications. In *ECUE'99: 01/12/1999-03/12/1999*.

[Chandler and Sweller, 1991] Chandler, P. and Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, 8(4):293–332.

[Chang et al., 2000] Chang, B.-W., Mackinlay, J., and Zellweger, P. T. (2000). Fluidly revealing information in fluid documents. In *Proc. Smart Graphics AAAI Spring Symposium*.

[Chang et al., 2017] Chang, C., Bach, B., Dwyer, T., and Marriott, K. (2017). Evaluating perceptually complementary views for network exploration tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1397–1407. ACM.

[Charleer et al., 2016] Charleer, S., Klerkx, J., Duval, E., De Laet, T., and Verbert, K. (2016). Creating effective learning analytics dashboards: Lessons learnt. In *European Conference on Technology Enhanced Learning*, pages 42–56. Springer.

[Chen et al., 2020] Chen, Z., Tong, W., Wang, Q., Bach, B., and Qu, H. (2020). Augmenting static visualizations with paparvis designer. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12.

[Chevalier et al., 2013] Chevalier, F., Vuillemot, R., and Gali, G. (2013). Using concrete scales: A practical framework for effective visual depiction of complex measures. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2426–2435.

[Cisneros, 2011] Cisneros, M. (2011). Is that true? `https://public.tableau.com/profile/mikevizneros#!/vizhome/IsThatRight/IsThatTrue`. online, last accessed Sept., 24, 2015.

[Clark and Paivio, 1991] Clark, J. M. and Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, 3(3):149–210.

[Cohn, 2013] Cohn, N. (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images.* A&C Black.

[Cohn, 2014] Cohn, N. (2014). The architecture of visual narrative comprehension: the interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology*, 5.

[Collins et al., 2018] Collins, C., Andrienko, N., Schreck, T., Yang, J., Choo, J., Engelke, U., Jena, A., and Dwyer, T. (2018). Guidance in the human–machine analytics process. *Visual Informatics*, 2(3):166–180.

[Conlen and Heer, 2018] Conlen, M. and Heer, J. (2018). Idyll: A markup language for authoring and publishing interactive articles on the web. In *Proc. ACM Symposium on User Interface Software and Technology (UIST)*, pages 977–989.

[Cordeil et al., 2019] Cordeil, M., Cunningham, A., Bach, B., Hurter, C., Thomas, B. H., Marriott, K., and Dwyer, T. (2019). Iatk: An immersive analytics toolkit. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 200–209. IEEE.

[Dang et al., 2016] Dang, T. N., Pendar, N., and Forbes, A. G. (2016). Timearcs: Visualizing fluctuations in dynamic networks. 35(3):61–69.

[Data Illustrator,] Data Illustrator. Data Illustrator: Create infographics and data visualizations without programming. `http://data-illustrator.com/`. Accessed: 2020-12-04.

[Delp and Jones, 1996] Delp, C. and Jones, J. (1996). 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.

[Diamond et al., 2016] Diamond, J., McQuillan, J., Spiegel, A. N., Hill, P. W., Smith, R., West, J., and Wood, C. (2016). Viruses, Vaccines and the Public. *Museums & Social Issues*, 11(1):9–16.

[Diehl et al., 2018] Diehl, A., Abdul-Rahman, A., El-Assady, M., Bach, B., Keim, D. A., and Chen, M. (2018). Visguides: A forum for discussing visualization guidelines. In *EuroVis (Short Papers)*.

[Dong et al., 2020] Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534.

[Dragicevic et al., 2019] Dragicevic, P., Jansen, Y., Sarma, A., Kay, M., and Chevalier, F. (2019). Increasing the transparency of research papers with explorable multiverse analyses. In *proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–15.

[Drexler and Wimpissinger, 1934] Drexler, A. and Wimpissinger, R. (1934). *Atlas Linguisticus*. Kifadruck.

[Dufournaud et al., 2017] Dufournaud, N., Michon, B., Bach, B., and Cristofoli, P. (2017). L'analyse des réseaux, une aide à penser : réflexions sur les stratégies économique et sociale de Marie Boucher, marchande à Nantes au XVII e siècle. In *Réseaux de femmes, femmes en réseaux (XVIe-XXIe siècles)*, pages 109–137. Presses universitaires de Bordeaux.

[Dwyer, 2013] Dwyer, T. (2013). cola.js: Constraint-Based Layout in the Browser. `http://marvl.infotech.monash.edu/webcola/`. [Online; accessed 10-Mar-2019].

[Dwyer et al., 2006] Dwyer, T., Koren, Y., and Marriott, K. (2006). IPSep-CoLa: An incremental procedure for separation constraint layout of graphs. *IEEE TVCG*, 12(5):821–828.

[Dwyer et al., 2018] Dwyer, T., Marriott, K., Isenberg, T., Klein, K., Riche, N., Schreiber, F., Stuerzlinger, W., and Thomas, B. H. (2018). Immersive analytics: An introduction. *Immersive analytics*, pages 1–23.

[DXR,] DXR. Dxr: A toolkit for building immersive data visualizations. `https://sites.google.com/view/dxr-vis`. Accessed: 2020-12-04.

[Dykes et al., 2016] Dykes, T., Wallace, J., Blythe, M., and Thomas, J. (2016). Paper street view: A guided tour of design and making using comics. In *Proceedings of ACM Conference on Designing Interactive Systems*, pages 334–346. ACM.

[Eisner, 2008] Eisner, W. (2008). *Graphic storytelling and visual narrative*. WW Norton & Company.

[Elshehaly et al., 2020] Elshehaly, M., Randell, R., Brehmer, M., McVey, L., Alvarado, N., Gale, C. P., and Ruddle, R. A. (2020). Qualdash: Adaptable generation of visualisation dashboards for healthcare quality improvement. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):689–699.

[Ens et al., 2021] Ens, B., Bach, B., Cordeil, M., Engelke, U., Serrano, M., et al. (2021). Grand challenges in immersive analytics. In *Proc. CHI*. ACM.

[Farinella and Ros, 2014] Farinella, M. and Ros, H. (2014). *Neurocomic*. Nobrow Press.

[Fekete, 2015] Fekete, J.-D. (2015). Reorder.js: A javascript library to reorder tables and networks. In *IEEE Visualization Conference (VIS)*.

[Few, 2006] Few, S. (2006). *Information dashboard design: The effective visual communication of data*, volume 2. O'reilly Sebastopol, CA.

[Few and Edge, 2007] Few, S. and Edge, P. (2007). Dashboard confusion revisited. *Perceptual Edge*, pages 1–6.

[Figma,] Figma. Figma Charts Infographics UI kit. `https://www.figma.com/community/file/855517047816771255`. Accessed: 2020-12-04.

[Flourish,] Flourish. Flourish: Beautiful and easy data visualization and storytelling. `https://flourish.studio/`. Accessed: 2020-12-04.

[flowmaps.blue (based on data from Statistics Netherlands), 2020] flowmaps.blue (based on data from Statistics Netherlands) (2020). Commuters in the netherlands.

[Foulsham et al., 2016] Foulsham, T., Wybrow, D., and Cohn, N. (2016). Reading without words: Eye movements in the comprehension of comic strips. *Applied Cognitive Psychology*, 30(4):566–579.

[Franz et al., 2015] Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., and Bader, G. D. (2015). Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, page btv557.

[Ge et al., 2023] Ge, L. W., Cui, Y., and Kay, M. (2023). CALVI: Critical thinking assessment for literacy in visualizations. In *Proc. CHI*, pages 1–18. ACM.

[Gelman and Loken, 2013] Gelman, A. and Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*, 348(1-17):3.

[Ghani et al., 2011] Ghani, S., Riche, N. H., and Elmqvist, N. (2011). Dynamic insets for context-aware graph navigation. In *CGF*, volume 30, pages 861–870. Wiley Online Library.

[Ghoniem et al., 2004] Ghoniem, M., Fekete, J.-D., and Castagliola, P. (2004). A comparison of the readability of graphs using node-link and matrix-based representations. In *IEEE Symposium on Information Visualization*, pages 17–24.

[Goodwin et al., 2021] Goodwin, S., Dkhissi, Y., Wu, Q., Moyle, B., Freidin, K., and Liebman, A. (2021). Informed dashboard designs for microgrid electricity market operators. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems*, e-Energy '21, page 406–411, New York, NY, USA. Association for Computing Machinery.

[Goodyear, 2015] Goodyear, P. (2015). Teaching as design. *Herdsa review of higher education*, 2(2):27–50.

[Goodyear et al., 2021] Goodyear, P., Carvalho, L., and Yeoman, P. (2021). Activity-centred analysis and design (ACAD): Core purposes, distinctive qualities and current developments. *Educational Technology Research and Development*, 69:445–464.

[Graham, 2017] Graham, E. (2017). Introduction: data visualization and the humanities. *English Studies*, 98(5):449–458.

[Gramener, 2019] Gramener (2019). Comicgen. `https://gramener.com/comicgen/#`.

[Grammel et al., 2013] Grammel, L., Bennett, C., Tory, M., and Storey, M.-A. (2013). A Survey of Visualization Construction User Interfaces. In Hlawitschka, M. and Weinkauf, T., editors, *EuroVis - Short Papers*. The Eurographics Association.

[Green, 1989] Green, T. R. (1989). Cognitive dimensions of notations. *People and computers V*, pages 443–460.

[Groensteen, 2007] Groensteen, T. (2007). *The System of Comics*. Univ. Press of Mississippi.

[Groensteen, 2013] Groensteen, T. (2013). *Comics and Narration*. Univ. Press of Mississippi.

[H. Brown and D. Green, 2020] H. Brown, A. and D. Green, T. (2020). *The Essentials of Instructional Design: Connecting Fundamental Principles with Process and Practice*. Taylor & Francis, 4th edition.

[Hadlak et al., 2011] Hadlak, S., Schulz, H.-J., and Schumann, H. (2011). In situ exploration of large dynamic networks. *IEEE TVCG*, 17(12):2334–2343.

[Hans Rosling and Rönnlund,] Hans Rosling, O. R. and Rönnlund, A. R. online: `https://www.gapminder.org`, [last accessed: 17 Aug 2018].

[Hao et al., 2024] Hao, S., Wang, Z., Bach, B., and Pschetz, L. (2024). Design patterns for data-driven news articles. In *CHI'24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–23.

[Haroz et al., 2016] Haroz, S., Kosara, R., and Franconeri, S. L. (2016). The connected scatterplot for presenting paired time series. *IEEE Transactions on Visualization and Computer Graphics*, 22(9):2174–2186.

[Heer et al., 2008] Heer, J., Mackinlay, J. D., Stolte, C., and Agrawala, M. (2008). Graphical histories for visualization: Supporting analysis, communication, and evaluation. *IEEE Trans. Vis. Comput. Graph.*, 14(6):1189–1196.

[Henry and Fekete, 2006] Henry, N. and Fekete, J.-D. (2006). Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):677–684.

[Henry and Fekete, 2007] Henry, N. and Fekete, J.-D. (2007). MatLink: Enhanced Matrix Visualization for Analyzing Social Networks. In *Human-Computer Interaction – INTERACT*, pages 288–302.

[Henry et al., 2007] Henry, N., Fekete, J.-D., and McGuffin, M. J. (2007). NodeTrix: a hybrid visualization of social networks. *IEEE TVCG*, 13(6):1302–1309.

[Henshaw and Meinke, 2018] Henshaw, A. L. and Meinke, S. R. (2018). Data analysis and data visualization as active learning in political science. *J. of Political Science Education*, 14(4):423–439.

[Hinrichs et al., 2019] Hinrichs, U., Forlini, S., and Moynihan, B. (2019). In defense of sandcastles: Research thinking through visualization in digital humanities. *Digital Scholarship in the Humanities*, 34(Supplement_1):i80–i99.

[Holten et al., 2011] Holten, D., Isenberg, P., van Wijk, J. J., and Fekete, J.-D. (2011). An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs. In *2011 IEEE Pacific Visualization Symposium*. IEEE.

[Hosler and Boomer, 2011a] Hosler, J. and Boomer, K. (2011a). Are comic books an effective way to engage nonmajors in learning and appreciating science? *CBE—Life Sciences Education*, 10(3):309–317.

[Hosler and Boomer, 2011b] Hosler, J. and Boomer, K. B. (2011b). Are Comic Books an Effective Way to Engage Nonmajors in Learning and Appreciating Science? *CBE Life Sciences Education*, 10(3):309–317.

[Hullman et al., 2013] Hullman, J., Drucker, S., Riche, N. H., Lee, B., Fisher, D., and Adar, E. (2013). A deeper understanding of sequence in narrative visualization. *IEEE Transactions on visualization and computer graphics*, 19(12):2406–2415.

[Huron et al., 2020] Huron, S., Bach, B., Hinrichs, U., Keck, M., and Roberts, J. (2020). 1st IEEE VIS workshop on data vis activities to facilitate learning, reflecting, discussing, and designing. `https://visactivities.github.io/2020/`.

[Huron et al., 2021] Huron, S., Bach, B., Panagiotidou, G., Keck, M., Roberts, J., and Carpendale, S. (2021). 2nd IEEE VIS workshop on data vis activities to facilitate learning, reflecting, discussing, and designing. `https://visactivities.github.io/`.

[Isenberg et al., 2018] Isenberg, P., Lee, B., Qu, H., and Cordeil, M. (2018). Immersive visual data stories. *Immersive analytics*, pages 165–184.

[Jansen et al., 2015] Jansen, Y., Dragicevic, P., Isenberg, P., Alexander, J., Karnik, A., Kildal, J., Subramanian, S., and Hornbæk, K. (2015). Opportunities and challenges for data physicalization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 3227–3236, New York, NY, USA. Association for Computing Machinery.

[Jones, 2011] Jones, G. E. (2011). *How to lie with charts*. LaPuerta Books and Media.

[Kahn and Jiang, 2021] Kahn, J. and Jiang, S. (2021). Learning with large, complex data and visualizations: youth data wrangling in modeling family migration. *Learning, Media and Technology*, 46(2):128–143.

[Kasica et al., 2020] Kasica, S., Berret, C., and Munzner, T. (2020). Table scraps: an actionable framework for multi-table data wrangling from an artifact study of computational journalism. *IEEE Transactions on visualization and computer graphics*, 27(2):957–966.

[Kauer et al., 2024] Kauer, T., Akbaba, D., Dörk, M., and Bach, B. (2024). Discursive patinas: Anchoring discussions in data visualizations. *IEEE Transactions on Visualization and Computer Graphics*.

[Keck et al., 2023] Keck, M., Huron, S., Panagiotidou, G., Stoiber, C., Rajabiyazdi, F., et al. (2023). EduVis: Workshop on visualization education, literacy, and activities.

[Khairat et al., 2018] Khairat, S. S., Dukkipati, A., Lauria, H. A., Bice, T., Travers, D., and Carson, S. S. (2018). The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. *JMIR Human Factors*, 5(2):e9328.

[Kim et al., 2021] Kim, H., Moritz, D., and Hullman, J. (2021). Design patterns and trade-offs in responsive visualization for communication. In *Computer Graphics Forum*, volume 40, pages 459–470. Wiley Online Library.

[Kim et al., 2024] Kim, N. W., Ko, H.-K., Myers, G., and Bach, B. (2024). Chatgpt in data visualization education: A student perspective. *arXiv preprint arXiv:2405.00748*.

[Kim et al., 2019] Kim, N. W., Riche, N. H., Bach, B., Xu, G. A., Brehmer, M., Hinckley, K., Pahud, M., Xia, H., McGuffin, M., and Pfister, H. (2019). Datatoon: Drawing dynamic network comics with pen + touch interaction. In *Proc. of ACM Conference of Human Factors in Computing Systems (CHI)*.

[Kim et al., 2016] Kim, N. W., Schweickart, E., Liu, Z., Dontcheva, M., Li, W., Popovic, J., and Pfister, H. (2016). Data-driven guides: Supporting expressive design for information graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):491–500.

[Kim et al., 2017a] Kim, Y., Wongsuphasawat, K., Hullman, J., and Heer, J. (2017a). Graphscape: A model for automated reasoning about visualization similarity and sequencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, May 06-11, 2017*, pages 2628–2638. ACM.

[Kim et al., 2017b] Kim, Y.-S., Reinecke, K., and Hullman, J. (2017b). Explaining the gap: Visualizing one's predictions improves recall and comprehension of data. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1375–1386.

[Kirk,] Kirk, A. Visualising Data – Resources. `https://www.visualisingdata.com/resources/`. Accessed: 2020-07-17.

[Kitchin et al., 2015] Kitchin, R., Lauriault, T. P., and McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, 2(1):6–28.

[Koffka, 1922] Koffka, K. (1922). Perception: An introduction to the gestalt-theorie. *Psychological bulletin*, 19(10):531–585.

[Kong et al., 2018] Kong, H.-K., Liu, Z., and Karahalios, K. (2018). Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, page 438. ACM.

[Kostis et al., 2022] Kostis, H.-N., Subbarao, M., Silva, B. L., Gurvich, A., Lentz, M., Kim, D., Ott, L., Ardizzone, J., Boller, R., Wong, M., Weiler, R., Li, A., Chevalier, F., Bach, B., Elkins, K., Christensen, A., Garrison, M., Kekesi, A., Schindler, T., Shirah, G., Chyatte, M., Jones, I., Kaplan, E., Schuler, L., Bridgman, T., Mitchell, H., Perkins, L., Starr, C., and Wright, E. (2022). Nasa's earth information dashbaord. *in Proc/ of Visualization in Science and Education*.

[Kraft et al., 2016] Kraft, S. A., Constantine, M., Magnus, D., Porter, K. M., Lee, S. S.-J., Green, M., Kass, N. E., Wilfond, B. S., and Cho, M. K. (2016). A randomized study of multimedia informational aids for research on medical practices: Implications for informed consent. *Clinical Trials*, 14(1):94–102.

[Krzywinski et al., 2011] Krzywinski, M., Birol, I., Jones, S. J., and Marra, M. A. (2011). Hive plots—rational approach to visualizing networks. *Briefings in Bioinformatics*, 13(5):627–644.

[Lam et al., 2012] Lam, H., Bertini, E., Isenberg, P., Plaisant, C., and Carpendale, S. (2012). Empirical studies in information visualization: Seven scenarios. *IEEE TVCG*, 18(9):1520–1536.

[Lau and Guo, 2020] Lau, S. and Guo, P. J. (2020). Data Theater: A live programming environment for prototyping data-driven explorable explanations. In *Workshop on Live Programming (LIVE)*.

[Lechner and Fruhling, 2014] Lechner, B. and Fruhling, A. (2014). Towards public health dashboard design guidelines. In Nah, F. F.-H., editor, *HCI in Business*, pages 49–59, Cham. Springer International Publishing.

[Lee et al., 2020] Lee, B., Brown, D., Lee, B., Hurter, C., Drucker, S., and Dwyer, T. (2020). Data visceralization: Enabling deeper understanding of data using virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1095–1105.

[Lee et al., 2006] Lee, B., Plaisant, C., Parr, C. S., Fekete, J.-D., and Henry, N. (2006). Task taxonomy for graph visualization. In *Proceedings of the AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*, page 1–5.

[Lee et al., 2015a] Lee, B., Riche, N. H., Isenberg, P., and Carpendale, S. (2015a). More than telling a story: Transforming data into visually shared stories. *Computer Graphics and Applications (CGA)*, 35(5):84–90.

[Lee et al., 2023] Lee, B., Sedlmair, M., and Schmalstieg, D. (2023). Design patterns for situated visualization in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*.

[Lee et al., 2015b] Lee, D., Felix, J. R. A., He, S., Offenhuber, D., and Ratti, C. (2015b). Cityeye: Real-time visual dashboard for managing urban services and citizen feedback loops. In *Proceedings of the 14th International Conference on Computing in Urban Planning and Urban Management (CUPUM), Cambridge, MA, USA*, pages 7–10.

[Lee et al., 2019] Lee, D. J. L., Dev, H., Hu, H., Elmeleegy, H., and Parameswaran, A. G. (2019). Avoiding drill-down fallacies with *VisPilot*: assisted exploration of data subsets. In *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI 2019, Marina del Ray, CA, USA, March 17-20, 2019*, pages 186–196. ACM.

[Lee et al., 2016] Lee, S., Kim, S.-H., Hung, Y.-H., Lam, H., Kang, Y.-a., and Yi, J. S. (2016). 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.

[Lee et al., 2017] Lee, S., Kim, S.-H., and Kwon, B. C. (2017). Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):551–560.

[Lee and Wilkerson, 2018] Lee, V. and Wilkerson, M. (2018). Data use by middle and secondary students in the digital age: a status report and future prospects. *Instructional Technology and Learning Sciences Faculty Public.*, pages 1–43.

[Lekschas et al., 2017] Lekschas, F., Bach, B., Kerpedjiev, P., Gehlenborg, N., and Pfister, H. (2017). Hipiler: visual exploration of large genome interaction matrices with interactive small multiples. *IEEE transactions on visualization and computer graphics*, 24(1):522–531.

[Lekschas et al., 2019] Lekschas, F., Behrisch, M., Bach, B., Kerpedjiev, P., Gehlenborg, N., and Pfister, H. (2019). Pattern-driven navigation in 2d multiscale visualizations with scalable insets. *IEEE transactions on visualization and computer graphics*, 26(1):611–621.

[Lekschas et al., 2020] Lekschas, F., Zhou, X., Chen, W., Gehlenborg, N., Bach, B., and Pfister, H. (2020). A generic framework and library for exploration of small multiples through interactive piling. *IEEE Transactions on Visualization and Computer Graphics*.

[Levie and Lentz, 1982] Levie, W. H. and Lentz, R. (1982). Effects of text illustrations: A review of research. *ECTJ*, 30(4):195–232.

[Levin et al., 2012] Levin, K., Cashore, B., Bernstein, S., and Auld, G. (2012). Overcoming the tragedy of super wicked problems: constraining our future selves to ameliorate global climate change. *Policy sciences*, 45(2):123–152.

[Li et al., 2023] Li, W., Schöttler, S., Scott-Brown, J., Wang, Y., Chen, S., Qu, H., and Bach, B. (2023). Networknarratives: Data tours for visual network exploration and analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–15.

[Lima, 2014] Lima, M. (2014). *The Book of Trees: Visualizing Branches of Knowledge*. Chronicle Books.

[Lima, 2017] Lima, M. (2017). *The book of circles: Visualizing spheres of knowledge*. Chronicle Books.

[Lin and Chen, 2007] Lin, H. and Chen, T. (2007). Reading authentic efl text using visualization and advance organizers in a multimedia learning environment.

[Liu et al., 2023] Liu, X., Alharbi, M. S., Chen, J., Diehl, A., Rees, D., et al. (2023). Visualization Resources: A Survey. *Information Visualization*, 22(1).

[Liu et al., 2018] Liu, Z., Thompson, J., Wilson, A., Dontcheva, M., Delorey, J., Grigg, S., Kerr, B., and Stasko, J. (2018). Data Illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1–13.

[Loy et al., 2019] Loy, A., Kuiper, S., and Chihara, L. (2019). Supporting data science in the statistics curriculum. *Journal of Statistics Education*, 27(1):2–11.

[Ludwig, 2001] Ludwig, D. (2001). The era of management is over. *Ecosystems*, 4:758–764.

[Ma et al., 2011] Ma, K.-L., Liao, I., Frazier, J., Hauser, H., and Kostis, H.-N. (2011). Scientific storytelling using visualization. *IEEE Computer Graphics and Applications*, 32(1):12–19.

[MacAskill, 2022] MacAskill, W. (2022). *What We Owe The Future: The Sunday Times Bestseller*. Simon and Schuster.

[Madden, 99] Madden, M. (99). ways to tell a story: Exercises in style. *New York, NY: Chamberlain Bros*.

[Maltese et al., 2015] Maltese, A. V., Harsh, J. A., and Svetina, D. (2015). Data visualization literacy: Investigating data interpretation along the novice—expert continuum. *Journal of College Science Teaching*, 45(1):84–90.

[Mann et al., 1999] Mann, M. E., Bradley, R. S., and Hughes, M. K. (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophysical Research Letters*, 26(6):759–762.

[Maps, 2016] Maps, C. (2016). *Collins School Atlas*. Collins.

[Mauri et al., 2017] Mauri, M., Elli, T., Caviglia, G., Uboldi, G., and Azzi, M. (2017). Rawgraphs: a visualisation platform to create open outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, page 28. ACM.

[Mayer and Gallini, 1990] Mayer, R. E. and Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82(4):715–726.

[Mayr et al., 2019] Mayr, E., Hynek, N., Salisu, S., and Windhager, F. (2019). Trust in Information Visualization. In *Proceedings of the EuroVis Workshop on Trustworthy Visualization (TrustVis)*. The Eurographics Association.

[McCloud, 1993] McCloud, S. (1993). Understanding comics: The invisible art. *Northampton, Mass*.

[McCloud and Manning, 1998] McCloud, S. and Manning, A. (1998). Understanding comics: The invisible art. *IEEE Transactions on Professional Communications*, 41(1):66–69.

[McGee et al., 2019] McGee, F., Ghoniem, M., Melançon, G., Otjacques, B., and Pinaud, B. (2019). The state of the art in multilayer network visualization. *CGF*, 38(6):125–149.

[McKenna et al., 2017] McKenna, S., Henry Riche, N., Lee, B., Boy, J., and Meyer, M. (2017). Visual narrative flow: Exploring factors shaping data visualization story reading experiences. In *Computer Graphics Forum*, volume 36, pages 377–387. Wiley Online Library.

[McNicol, 2017] McNicol, S. (2017). The potential of educational comics as a health information medium. *Health Information & Libraries Journal*, 34(1):20–31.

[McNutt, 2023] McNutt, A. (2023). No grammar to rule them all: A survey of JSON-style DSLs for visualization. *IEEE Transactions on Computer Graphics and Visualization (TVCG)*.

[Mehta et al., 2017] Mehta, H., Chalbi, A., Chevalier, F., and Collins, C. (2017). Datatours: A data narratives framework. In *IEEE InfoVis 2017-IEEE Information Visualization conference*, pages 1–2.

[Meirelles, 2013] Meirelles, I. (2013). *Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations*. Rockport publishers.

[Melton, 2014] Melton, R. (2014). *Objectives, Competencies and Learning Outcomes: Developing Instructional Materials in Open and Distance Learning*. Routledge.

[Mlaver et al., 2017] Mlaver, E., Schnipper, J. L., Boxer, R. B., Breuer, D. J., Gershanik, E. F., Dykes, P. C., Massaro, A. F., Benneyan, J., Bates, D. W., and Lehmann, L. S. (2017). User-centered collaborative design and development of an inpatient safety dashboard. *The Joint Commission Journal on Quality and Patient Safety*, 43(12):676–685.

[Moere, 2007] Moere, A. V. (2007). Aesthetic data visualization as a resource for educating creative design. In Dong, A., Moere, A. V., and Gero, J. S., editors, *Computer-Aided Architectural Design Futures 2007*, pages 71–84. Springer Netherlands, Dordrecht.

[Monastersky and Sousanis, 2015] Monastersky, R. and Sousanis, N. (2015). The fragile framework. *Nature News*, 527(7579):427.

[Monmonier, 2018] Monmonier, M. (2018). *How to lie with maps*. University of Chicago Press.

[Moritz et al., 2019] Moritz, D., Wang, C., Nelson, G. L., Lin, H., Smith, A. M., Howe, B., and Heer, J. (2019). Formalizing visualization design knowledge as constraints: Actionable and extensible models in Draco. *IEEE TVCG*, 25(1):438–448.

[Muller et al., 2019] Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., and Erickson, T. (2019). How data science workers work with data: Discovery, capture, curation, design, creation. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15.

[Munroe, 2015] Munroe, R. (2015). *Thing explainer: complicated stuff in simple words*. Hachette UK.

[Munzner, 2009] Munzner, T. (2009). A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928.

[Munzner, 2014] Munzner, T. (2014). *Visualization analysis and design*. AK Peters/CRC Press.

[Newman, 2004] Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6).

[Nielsen et al., 2009] Nielsen, C., Jackman, S., Birol, I., and Jones, S. (2009). ABySS-explorer: Visualizing genome sequence assemblies. *IEEE TVCG*, 15(6):881–888.

[Nobre et al., 2019a] Nobre, C., Meyer, M., Streit, M., and Lex, A. (2019a). The state of the art in visualizing multivariate networks. *CGF*, 38(3):807–832.

[Nobre et al., 2019b] Nobre, C., Streit, M., Meyer, M., and Lex, A. (2019b). The state of the art in visualizing multivariate networks. *Computer Graphics Forum*, 38:807–832.

[Noonpakdee et al., 2018] Noonpakdee, W., Khunkornsiri, T., Phothichai, A., and Danaisawat, K. (2018). A framework for analyzing and developing dashboard templates for small and medium enterprises. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, pages 479–483. IEEE.

[of Economic and Affairs, 2023] of Economic, U. N. D. and Affairs, S. (2023). *The Sustainable Development Goals Report 2023: Special Edition*. UN.

[Ortelius, 1570] Ortelius, A. (1570). *Theatrum Orbis Terrarum*. Gilles Coppens de Diest.

[Paivio, 1990] Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

[Pandey et al., 2015] Pandey, A. V., Rall, K., Satterthwaite, M. L., Nov, O., and Bertini, E. (2015). How deceptive are deceptive visualizations? An empirical analysis of common distortion techniques. In *Proc. CHI*, page 1469–1478. ACM.

[Parsons, 2022] Parsons, P. (2022). Understanding data visualization design practice. *IEEE TVCG*, 28(1):665–675.

[Pereira et al., 2016] Pereira, D., Flores, M. A., and Niklasson, L. (2016). Assessment revisited: A review of research in assessment and evaluation in higher education. *Assessment & Evaluation in Higher Education*, 41(7):1008–1032.

[Perer and Shneiderman, 2009] Perer, A. and Shneiderman, B. (2009). Integrating statistics and visualization for exploratory power: From long-term case studies to design guidelines. *IEEE Computer Graphics and Applications*, 29(3):39–51.

[Perer and Sun, 2012] Perer, A. and Sun, J. (2012). Matrixflow: temporal network visual analytics to track symptom evolution during disease progression. In *AMIA annual symposium proceedings*, volume 2012, page 716. American Medical Informatics Association.

[Perry, 2015] Perry, J. (2015). Climate change adaptation in the world's best places: A wicked problem in need of immediate attention. *Landscape and Urban Planning*, 133:1–11.

[Pinker, 1990] Pinker, S. (1990). A theory of graph comprehension. *Artificial intelligence and the future of testing*, 73:126.

[Pirlea et al., 2023] Pirlea, A. F., Serajuddin, U., Wadhwa, D., and Welch, M. (2023). Atlas of sustainable development goals 2023. World Bank. License: Creative Commons Attribution CC BY 3.0 IGO.

[Pirlea et al., 2020] Pirlea, A. F., Serajuddin, U., Wadhwa, D., Welch, M., and Whitby, A. (2020). *Atlas of the Sustainable Development Goals 2020: From World Development Indicators*. World Bank, Washington, DC. License: Creative Commons Attribution CC BY 3.0 IGO.

[Pörtner et al., 2022] Pörtner, H.-O., Roberts, D., Tignor, M., Poloczanska, E., Mintenbeck, K., et al. (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. IPCC Geneva, Switzerland.

[Qu and Hullman, 2018] Qu, Z. and Hullman, J. (2018). Keeping multiple views consistent: Constraints, validations, and exceptions in visualization authoring. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 24(1):468–477.

[Rasmussen et al., 2009] Rasmussen, N. H., Bansal, M., and Chen, C. Y. (2009). *Business dashboards: a visual catalog for design and deployment*. John Wiley & Sons.

[Reas and Fry, 2006] Reas, C. and Fry, B. (2006). Processing: programming for the media arts. *AI & SOCIETY*, 20(4):526–538.

[Ren et al., 2017] Ren, D., Brehmer, M., Lee, B., Höllerer, T., and Choe, E. K. (2017). Chartaccent: Annotation for data-driven storytelling. In *Proc. IEEE Pacific Visualization Symposium (PacificVis)*, pages 230–239.

[Ren et al., 2018] Ren, D., Lee, B., and Brehmer, M. (2018). Charticulator: Interactive construction of bespoke chart layouts. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):789–799.

[Rendgen, 2018] Rendgen, S. (2018). *The minard system: the complete statistical graphics of Charles-Joseph Minard*. Chronicle Books.

[Riche et al., 2012] Riche, N. H., Dwyer, T., Lee, B., and Carpendale, S. (2012). Exploring the design space of interactive link curvature in network diagrams. In *Proc. AVI*. ACM Press.

[Riche et al., 2018] Riche, N. H., Hurter, C., Diakopoulos, N., and Carpendale, S. (2018). *Data-driven Storytelling*. CRC Press.

[Riche et al., 2014] Riche, N. H., Riche, Y., Roussel, N., Carpendale, S., Madhyastha, T., and Grabowski, T. J. (2014). Linkwave: A visual adjacency list for dynamic weighted networks. In *Proceedings of the 26th Conference on l'Interaction Homme-Machine*, pages 113–122.

[Rittel and Webber, 1973] Rittel, H. W. and Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169.

[Roberts, 2007] Roberts, J. C. (2007). State of the art: Coordinated & multiple views in exploratory visualization. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, pages 61–71. IEEE.

[Romat et al., 2020] Romat, H., Henry Riche, N., Hurter, C., Drucker, S., Amini, F., and Hinckley, K. (2020). Dear pictograph: Investigating the role of personalization and immersion for consuming and enjoying visualizations. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1–13.

[Rop et al., 2018] Rop, G., Schüler, A., Verkoeijen, P. P., Scheiter, K., and Gog, T. V. (2018). The effect of layout and pacing on learning from diagrams with unnecessary text. *Applied Cognitive Psychology*.

[Rosenberg and Grafton, 2013] Rosenberg, D. and Grafton, A. (2013). *Cartographies of time: A history of the timeline*. Princeton Architectural Press.

[Roser, nd] Roser, M. (n.d.). Our world in data. `https://ourworldindata.org/`. Accessed: 2024-02-20.

[Rossi and Ahmed, 2015] Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *AAAI*.

[Roy and Warren, 2019] Roy, R. and Warren, J. P. (2019). Card-based design tools: A review and analysis of 155 card decks for designers and designing. *Design Studies*, 63:125–154.

[Rushmeier, 2006] Rushmeier, H. (2006). Workshop: Visualization Education for Non-Technical Majors. In *Proc. IEEE Visualization*.

[Salton and McGill, 1984] Salton, G. and McGill, M. (1984). *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company.

[Sarikaya et al., 2019] Sarikaya, A., Correll, M., Bartram, L., Tory, M., and Fisher, D. (2019). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1):682–692.

[Satyanarayan and Heer, 2014a] Satyanarayan, A. and Heer, J. (2014a). Authoring narrative visualizations with ellipsis. In *Computer Graphics Forum*, volume 33, pages 361–370.

[Satyanarayan and Heer, 2014b] Satyanarayan, A. and Heer, J. (2014b). Lyra: An Interactive Visualization Design Environment. *Computer Graphics Forum*, 33(3):351–360. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12391.

[Satyanarayan et al., 2019] Satyanarayan, A., Lee, B., Ren, D., Heer, J., Stasko, J., Thompson, J., Brehmer, M., and Liu, Z. (2019). Critical reflections on visualization authoring systems. *IEEE transactions on visualization and computer graphics*, 26(1):461–471.

[Satyanarayan et al., 2016] Satyanarayan, A., Moritz, D., Wongsuphasawat, K., and Heer, J. (2016). Vega-lite: A grammar of interactive graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350.

[Sauvé et al., 2023] Sauvé, K., Dragicevic, P., and Jansen, Y. (2023). Edo: A participatory data physicalization on the climate impact of dietary choices. In *Proceedings of the Seventeenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–13.

[Sayama et al., 2016] Sayama, H., Cramer, C., Porter, M. A., Sheetz, L., and Uzzo, S. (2016). What are essential concepts about networks? *Journal of Complex Networks*, 4(3):457–474.

[Schöttler et al., 2024] Schöttler, S., Dykes, J., Wood, J., Hinrichs, U., and Bach, B. (2024). Constraint-based breakpoints for responsive visualization design and development. *IEEE Transactions on Visualization and Computer Graphics*.

[Schöttler et al., 2021] Schöttler, S., Yang, Y., Pfister, H., and Bach, B. (2021). Visualizing and interacting with geospatial networks: A survey and design space. *Comput. Graph. Forum*, 40(6):5–33.

[Sciences-Po médialab and OuestWare, 2013] Sciences-Po médialab and OuestWare (2013). Sigma.js — sigmajs.org. `https://www.sigmajs.org/`. [Accessed 06-Sep-2022].

[Sedig and Parsons, 2016] Sedig, K. and Parsons, P. (2016). Design of visualizations for human-information interaction: A pattern-based framework. *Synthesis Lectures on Visualization*, 4(1):1–185.

[Segel and Heer, 2010] Segel, E. and Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148.

[Serrano et al., 2020] Serrano, J. C. M., Papakyriakopoulos, O., Shahrezaye, M., and Hegelich, S. (2020). The political dashboard: A tool for online political transparency. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 983–985.

[Setiawan and Sukoco, 2021] Setiawan, E. P. and Sukoco, H. (2021). Exploring first year university students' statistical literacy: a case on describing and visualizing data. *J. Math. Educ.*, 12(3):427–448.

[Shen et al., 2023] Shen, L., Zhang, Y., Zhang, H., and Wang, Y. (2023). Data player: Automatic generation of data videos with narration-animation interplay. *IEEE Transactions on Visualization and Computer Graphics*.

[Shi et al., 2021] Shi, D., Xu, X., Sun, F., Shi, Y., and Cao, N. (2021). Calliope: Automatic visual data story generation from a spreadsheet. *IEEE Trans. Vis. Comput. Graph.*, 27(2):453–463.

[Shneiderman and Aris, 2006] Shneiderman, B. and Aris, A. (2006). Network visualization by semantic substrates. *IEEE TVCG*, 12(5):733–740.

[Shneiderman and Plaisant, 2006] Shneiderman, B. and Plaisant, C. (2006). Strategies for evaluating information visualization tools. In *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for information visualization - BELIV '06*. ACM Press.

[Shores, 2016] Shores, C. (2016). 'ragged time'in intra-panel comics rhythms. *The Comics Grid: Journal of Comics Scholarship*, 6.

[Short et al., 2013a] Short, J. C., Randolph-Seng, B., and McKenny, A. F. (2013a). Graphic Presentation An Empirical Examination of the Graphic Novel Approach to Communicate Business Concepts. *Business Communication Quarterly*, 76(3):273–303.

[Short et al., 2013b] Short, J. C., Randolph-Seng, B., and McKenny, A. F. (2013b). Graphic Presentation An Empirical Examination of the Graphic Novel Approach to Communicate Business Concepts. *Business Communication Quarterly*, 76(3):273–303.

[Shu et al., 2024] Shu, X., Pister, A., Tang, J., Chevalier, F., and Bach, B. (2024). Does this have a particular meaning? interactive pattern explanation for network visualizations.

[Shu et al., 2020] Shu, X., Wu, A., Tang, J., Bach, B., Wu, Y., and Qu, H. (2020). What makes a data-gif understandable? *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1492–1502.

[Smith et al., 2010] Smith, M., Ceni, A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., and Dunne, C. (2010). NodeXL: a free and open network overview, discovery and exploration add-in for excel 2007/2010/2013/2016 from the social media research foundation.

[Smith et al., 2009] Smith, M. A., Shneiderman, B., Milic-Frayling, N., Rodrigues, E. M., Barash, V., Dunne, C., Capone, T., Perer, A., and Gleave, E. (2009). Analyzing (social media) networks with NodeXL. In *Proceedings of the fourth international conference on Communities and technologies - C&T '09*. ACM Press.

[Sousanis, 2015] Sousanis, N. (2015). *Unflattening*. Harvard University Press, Cambridge, Massachusetts.

[Spiegel et al., 2013a] Spiegel, A. N., McQuillan, J., Halpin, P., Matuk, C., and Diamond, J. (2013a). Engaging teenagers with science through comics. *Research in science education*, 43(6):2309–2326.

[Spiegel et al., 2013b] Spiegel, A. N., McQuillan, J., Halpin, P., Matuk, C., and Diamond, J. (2013b). Engaging Teenagers with Science Through Comics. *Research in science education*, 43(6).

[Stanford University Humanities + Design lab,] Stanford University Humanities + Design lab. Palladio: Visualize complex historical data with ease.

[Stasko et al., 2008] Stasko, J., Görg, C., and Liu, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132.

[Stein et al., 2010] Stein, K., Wegener, R., and Schlieder, C. (2010). Pixel-oriented visualization of change in social networks. In *2010 International Conference on Advances in Social Networks Analysis and Mining*, pages 233–240. IEEE.

[Stolper et al., 2014] Stolper, C. D., Kahng, M., Lin, Z., Foerster, F., Goel, A., Stasko, J., and Chau, D. H. (2014). GLO-STIX: Graph-level operations for specifying techniques and interactive eXploration. *IEEE TVCG*, 20(12):2320–2328.

[Sudmant et al., 2024] Sudmant, A., Boyle, D., Higgins-Lavery, R., Gouldson, A., Boyle, A., Fulker, J., and Brogan, J. (2024). Climate policy as social policy? a comprehensive assessment of the economic impact of climate action in the uk. *Journal of Environmental Studies and Sciences*, pages 1–15.

[Sultanum et al., 2018] Sultanum, N., Brudno, M., Wigdor, D., and Chevalier, F. (2018). 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 422. ACM.

[Sultanum et al., 2021] Sultanum, N., Chevalier, F., Bylinkskii, Z., and Liu, Z. (2021). Leveraging text-chart links to support authoring of data-driven articles with vizflow. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI)*, page In press.

[Tableau,] Tableau. Tableau. `https://www.tableau.com/`. Accessed: 2020-12-04.

[Tan, 1982] Tan, Q. (1982). *The Historical Atlas of China*. China Cartographic Publishing House.

[Tatalovic, 2009] Tatalovic, M. (2009). Science comics as tools for science education and communication: a brief, exploratory study. *The Journal of Science Communication (JCOM)*.

[Tedre and Pajunen, 2022] Tedre, M. and Pajunen, J. (2022). Grand theories or design guidelines? perspectives on the role of theory in computing education research. *ACM Trans. on Computing Education*, 23(1):1–20.

[Tekle-Haimanot et al., 2016] Tekle-Haimanot, R., Pierre-Marie, P., Daniel, G., Worku, D. K., Belay, H. D., and Gebrewold, M. A. (2016). Impact of an educational comic book on epilepsy-related knowledge, awareness, and attitudes among school children in Ethiopia. *Epilepsy & Behavior: E&B*, 61:218–223.

[The Growth Lab at Harvard University,] The Growth Lab at Harvard University. The atlas of economic complexity. `http://www.atlas.cid.harvard.edu`. Last accessed: 2024-03-13.

[Tukey, 1977] Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Springer.

[Tyner et al., 2017] Tyner, S., Briatte, F., and Hofmann, H. (2017). Network visualization with ggplot2. *The R Journal*, 9(1):27.

[van den Elzen and van Wijk, 2013] van den Elzen, S. and van Wijk, J. J. (2013). Small multiples, large singles: A new approach for visual data exploration. *Comput. Graph. Forum*, 32(3):191–200.

[Victor, 2011] Victor, B. (2011). Explorable explanations. Online. `http://worrydream.com/ExplorableExplanations/`.

[Victor, 2013] Victor, B. (2013). Tangle: explorable explanations made easy. `http://worrydream.com/Tangle`.

[Viegas et al., 2007] Viegas, F. B., Wattenberg, M., Ham, F. v., Kriss, J., and McKeon, M. (2007). ManyEyes: a Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128.

[Vogogias et al., 2020] Vogogias, A., Archambault, D., Bach, B., and Kennedy, J. (2020). Visual encodings for networks with multiple edge types. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 1–9.

[Von Landesberger et al., 2011] Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., Van Wijk, J., Fekete, J.-D., and Fellner, D. (2011). Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges. *Computer Graphics Forum*, 30(6):1719–1749.

[Walny et al., 2019] Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S., and Willett, W. (2019). Data Changes Everything: Challenges and Opportunities in Data Visualization Design Handoff. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):12–22.

[Wang et al., 2023] Wang, J., AlKadi, M., and Bach, B. (2023). Show me my users: A dashboard visualizing user interaction logs. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pages 156–160. IEEE.

[Wang et al., 2018] Wang, Y., Zhang, H., Huang, H., Chen, X., Yin, Q., Hou, Z., Zhang, D., Luo, Q., and Qu, H. (2018). Infonice: Easy creation of information graphics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018, Montreal, QC, Canada, April 21-26, 2018*, page 335. ACM.

[Wang et al., 2019a] Wang, Z., Dingwall, H., and Bach, B. (2019a). Teaching data visualization and storytelling with data comic workshops. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–9.

[Wang et al., 2020] 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*, 27(2):967–977.

[Wang et al., 2021] Wang, Z., Romat, H., Chevalier, F., Riche, N. H., Murray-Rust, D., and Bach, B. (2021). Interactive data comics. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):944–954.

[Wang et al., 2019b] Wang, Z., Wang, S., Farinella, M., Murray-Rust, D., Henry Riche, N., and Bach, B. (2019b). Comparing effectiveness and engagement of data comics and infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12.

[Ware, 2019] Ware, C. (2019). *Information visualization: perception for design*. Morgan Kaufmann.

[Wattenberg, 2002] Wattenberg, M. (2002). Arc diagrams: Visualizing structure in strings. In *Proc. InfoVis*, pages 110–116. IEEE.

[Wattenberg, 2006] Wattenberg, M. (2006). Visual exploration of multivariate graphs. In *Proc. CHI*, pages 811–819.

[Wilbanks and Langford, 2014] Wilbanks, B. A. and Langford, P. A. (2014). A review of dashboards for data analytics in nursing. *CIN: Computers, Informatics, Nursing*, 32(11):545–549.

[William JR, 2012] William JR, L. (2012). Combing the hairball with biofabric: a new approach for visualization of large networks. *BMC Bioinformatics*.

[Williams, 2012] Williams, I. C. (2012). Graphic medicine: comics as medical narrative. *Medical Humanities*, 38(1):21–27.

[Windhager et al., 2019] Windhager, F., Schreder, G., and Mayr, E. (2019). On inconvenient images: Exploring the design space of engaging climate change visualizations for public audiences. In *EnvirVis@ EuroVis*, pages 1–8.

[Wong and Carpendale, 2007] Wong, N. and Carpendale, S. (2007). Supporting interactive graph exploration using edge plucking. In *Visualization and Data Analysis*, volume 6495, pages 76–87.

[World Bank, 2017] World Bank (2017). *Atlas of Sustainable Development Goals 2017: World Development Indicators*. World Bank, Washington, DC. License: Creative Commons Attribution CC BY 3.0 IGO.

[World Bank, 2018] World Bank (2018). *Atlas of Sustainable Development Goals 2018: From World Development Indicators*. World Bank, Washington, DC. License: CC BY 3.0 IGO.

[Xia et al., 2018] Xia, H., Henry Riche, N., Chevalier, F., De Araujo, B., and Wigdor, D. (2018). DataInk: Direct and creative data-oriented drawing. In *Proc. SIGCHI Conference on Human Factors in Computing Systems* (*CHI*), pages 1–13.

[Yang, 2008] Yang, G. (2008). Graphic novels in the classroom. *Language Arts*, 85(3):185.

[Yarnes, 2013] Yarnes, S. C. (2013). Introduction to randomization and layout. Online. `https://pbgworks.org/index.php?q=node/1534`.

[Yi et al., 2007] Yi, J. S., ah Kang, Y., Stasko, J., and Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231.

[Yi et al., 2010] Yi, J. S., Elmqvist, N., and Lee, S. (2010). Timematrix: Analyzing temporal social networks using interactive matrix-based visualizations. *Intl. Journal of Human–Computer Interaction*, 26(11-12):1031–1051.

[Yigitbasioglu and Velcu, 2012] Yigitbasioglu, O. M. and Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, 13(1):41–59.

[Yoghourdjian et al., 2021] Yoghourdjian, V., Yang, Y., Dwyer, T., Lee, L., Wybrow, M., and Marriott, K. (2021). Scalability of network visualisation from a cognitive load perspective. *IEEE Trans. Vis. Comput. Graph.*, 27(2):1677–1687.

[Zhang et al., 2020] Zhang, J. E., Sultanum, N., Bezerianos, A., and Chevalier, F. (2020). DataQuilt: Extracting visual elements from images to craft pictorial visualizations. In *Proc. SIGCHI Conference on Human Factors in Computing Systems* (*CHI*), pages 1–13.

[Zhang et al., 2021] Zhang, Y., Sun, Y., Padilla, L., Barua, S., Bertini, E., and Parker, A. G. (2021). Mapping the landscape of covid-19 crisis visualizations. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–23.

[Zhao and Elmqvist, 2022] Zhao, Z. and Elmqvist, N. (2022). Datatv: Streaming data videos for storytelling. *arXiv preprint arXiv:2210.08175*.

[Zhao et al., 2015] Zhao, Z., Marr, R., and Elmqvist, N. (2015). Data comics: Sequential art for data-driven storytelling. *tech. report*.

[Zhao et al., 2019] Zhao, Z., Marr, R., Shaffer, J., and Elmqvist, N. (2019). Understanding partitioning and sequence in data-driven storytelling. In *Information in Contemporary Society: 14th International Conference, iConference 2019, Washington, DC, USA, March 31–April 3, 2019, Proceedings 14*, pages 327–338. Springer.

[Zhu et al., 2020] Zhu, S., Sun, G., Jiang, Q., Zha, M., and Liang, R. (2020). A survey on automatic infographics and visualization recommendations. *Visual Informatics*, 4(3):24–40.