Report from Dagstuhl Seminar 16061

Data-Driven Storytelling

Edited by

Sheelagh Carpendale[1], Nicholas Diakopoulos[2], Nathalie Henry Riche[3], and Christophe Hurter[4]

1   **University of Calgary, Calgary, CA,** `sheelagh@ucalgary.ca`
2   **University of Maryland – College Park, US,** `nicholas.diakopoulos@gmail.com`
3   **Microsoft Research – Redmond, US,** `nathalie.henry@microsoft.com`
4   **ENAC – Toulouse, FR,** `christophe.hurter@enac.fr`

──── **Abstract** ────

This report documents the program and the outcomes of Dagstuhl Seminar 16061 "Data-Driven Storytelling". Close to forty researchers and practitioners descended on Schloss Dagstuhl to forge an interdisciplinary agenda on the topic of data-driven storytelling using visualization in early February, 2016. With burgeoning research interest in understanding what makes visualization effective for communication, and with practitioners pushing the envelope of the craft of visual communication, the meeting put different modes of thinking between computer science researchers and data visualization practitioners in close proximity for a week.

1   Summary

Sheelagh Carpendale
Nicholas Diakopoulos
Nathalie Henry Riche
Christophe Hurter

Data visualization is the "use of computer-supported, interactive, visual representations of data to amplify cognition" [5]. Visualization can play a crucial role for exploring data and for communicating information as "a picture is worth a thousand words". Early research in this field focused on producing static images and quantifying the perception of different visual encodings [6] in these visual representations. The vast majority of research since then focused on designing and implementing novel interfaces and interactive techniques to enable data exploration. Major advances in visual analytics and big data initiatives concentrated on integrating machine learning and analysis methods with visual representations to enable powerful exploratory analysis and data mining [10]. As interactive visualizations play an increasing role in data analysis scenarios, they also started to appear as a powerful vector for communicating information. Stories supported by facts extracted from data analysis
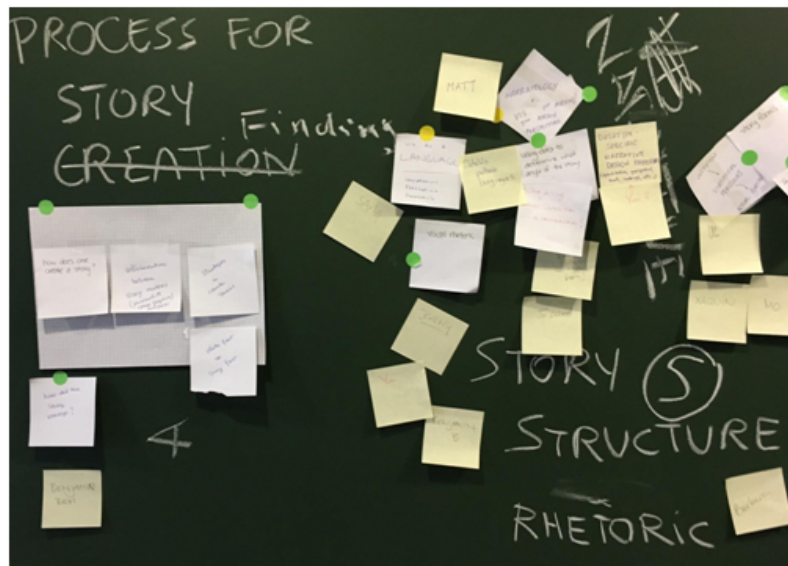
proliferate in many different forms from animated infographics and videos [2] to interactive online visualizations on news media outlets. We argue that it is now time for the visualization research community to understand how these powerful interactive visualizations play a role in communicating information. We define this line of research as data-driven storytelling.

The popularity of javascript web technology and the availability of the D3 toolkit [3] enabled a wider range of people to create data visualizations. Being able to easily share interactive data visualizations on the web also increased the democratization of interactive visualizations. Coupled with the emphasis on data science, these advances raise new practices such as data journalism. Data journalists gather and explore available datasets to extract relevant insights, often conveying their stories via interactive data visualizations [1, 9]. The popularity of data-driven stories on New York Times especially, revealed the potential of interactive visualizations as a powerful communication tool [7].

Central to our vision of the convening was that the vast majority of research on data visualization to date has focused on designing and implementing novel interfaces and interactive techniques to enable data exploration. Major advances in visual analytics and big data initiatives have concentrated on integrating machine learning and analysis methods with visual representations to enable powerful exploratory analysis and data mining. But just as interactive visualization plays an important role in data analysis scenarios it is also becoming increasingly important in structuring the communication and conveyance of insights and stories in a compelling format. Visual data-driven stories have proliferated in many different forms, from talks [8], to animated infographics and videos [1, 9, 7], to interactive online visualizations.

Data-driven storytelling is also compelling for a wide range of applications. In enterprise scenarios, the output of data analysis (often reports and slide-based presentations) has to be conveyed to decision makers. In scientific research, interactive visualizations are increasingly used to convey data-driven discoveries to peers or used to communicate complex findings to a broader audience. In education scenarios, interactive visualizations are used by teachers to explain mathematical concepts or to illustrate biological or physical mechanisms. Many questions arise as interactive visualizations are used beyond data exploration by experts, for communication purposes to a broader audience. Research on understanding of static images in cognitive psychology and perception must be extended to encompass more advanced techniques (videos and interactive applications). Visualization literacy, defined as the ability to extract, interpret, and make meaning from information presented in the form of an (interactive) data visualization is also a crucial component for data-driven storytelling research. Assessing the visualization literacy of an audience and developing techniques to better teach how to decode interactive visualizations has started to attract the attention of our research community [4] However a plethora of research remains to be done. For example, research on how visualizations can lie [11] or at least how they may introduce bias in the reader's mind has focused on static visual representations but has not yet been extended to other medium. Similarly it is crucial for advancing researches in visualization to assess the role data-driven storytelling can play in easing the comprehension of a messages or in increasing their memorability.

The visualization research community needs to reflect on data-driven storytelling and to develop a research agenda to investigate how advanced data-driven stories are understood by the audience, identify factors that makes them compelling as well as factors that can introduce bias in their perception. By learning from master storytellers from other fields (journalism, design, art and education) strategies to craft successful stories, our community will be able to reflect on these questions and eventually build novel consuming tools that

Figure 1 Converging on topical groups from hundreds of individual ideas.

engage a broad audience while minimizing perception bias, as well as build novel authoring tools to craft high quality data-driven stories.

One domain where there has been extensive and practical progress on the question of data-driven storytelling is data journalism. News sites like FiveThirtyEight or the New York Times' The Upshot have seen a recent surge of attention and interest as a means of communicating data-driven news to the public. By carefully structuring the information and integrating explanation to guide the consumer, journalists help lead users toward a valid interpretation of the underlying data. Because of the rapid and practical progress of data-driven storytelling in the domain of journalism, our seminar sought to put some of the top practitioners from that field together with computer science researchers to discuss the challenges and opportunities of data-driven communication.

The Dagstuhl seminar was structured to leverage the interdisciplinarity of the attendees by first tapping into a divergent design thinking process meant to enumerate the range of issues that are relevant to data-driven stories. Hundreds of index cards and sticky notes were sacrificed as participants generated ideas (see Figure 1).

We then clustered these ideas to arrived at a set of key themes, including:

- Techniques and Design Choices for Storytelling
- Exploration and Explanation
- From Analysis to Communication
- Audience
- Evaluation
- Devices and Gadgets
- Ethics

Groups of participants formed around common interests and each of these major themes were then the focus of discussion. Each work group was geared towards developing an outline and plan to produce a written chapter for a forthcoming edited book on the topic of data-driven storytelling. Some groups met for a day or two and then reformed around other topics, whereas other groups spent the entire week going deep in exploring a single

topic. And as if the daytime activities weren't enough, additional evening breakout groups formed around additional topics of interest like Education in Data Visualization, Urban Visualization, and the Technology Stack for data-driven stories.

In-between the intense, small group sessions the entire group came together daily for five-minute lightning talks on a wide array of relevant topics. These stimulating talks primed the group for approaching data-driven storytelling from different perspectives and were an entertaining and informative way to share creative ideas or results in small and easily digestible nuggets. Among the more than 25 lightning talks, topics ranged from storytelling with timelines, to mobile visualization, the use of data comics, visual literacy, affect and color, data-story design workflows, and even the visualization of data through cuisine.

Outcomes

Our initial goal of the seminar was to have groups work intensively on their chosen topic(s) so that an outline and workplan could be developed to write a contributing chapter to a book on data-driven storytelling. The book is underway and will have contributions on each of the main themes outlined above, as well as an introductory chapter by the editors / organizers of the Dagstuhl seminar. Moreover, our creative contributors at the seminar produced other outputs as well: curated lists of example data driven stories, as well as of storytelling techniques were created and will be published online, and a blog has pulled together some of the formative impressions of participants (https://medium.com/data-driven-storytelling).

Below we briefly summarize the expected contents of each of the chapters that will form the book.

Techniques and Design Choices for Storytelling

This chapter will discuss techniques and design choices for visual storytelling grounded in a survey of over 60 examples collected from various online news sources and from award-winning visualization and infographic design work. These design choices represent a middle ground between low-level visualization and interaction techniques and high-level narrative devices or structures. The chapter will define several classes of design choices: embellishment, explanation, exploration, navigation, story presentation, emphasis, focus, and annotation. Examples from the survey for each class of design choices will be provided. Finally, several case studies of examples from the survey that make use of multiple design choices will be developed.

Exploration and Explanation in Data-Driven Stories

This chapter will explore the differences between and integration of exploration and explanation in visual data-driven storytelling. Exploratory visualizations allow for a lot of freedom which can include changing the visual representation, the focus of what is being shown and the sequence in which the data is viewed. They allow readers to find their own stories in the data. Explanatory stories include a focused message which is usually more narrow and guides the reader often in a linear way. Advantages and disadvantages of exploration and explanation as well as dimensions that help to describe and classify data-driven stories will be developed. The space is described by identifying freedom, guidance regarding representation, focus and sequence as well as interpretation as important dimensions of data-driven storytelling and existing systems are characterized along these dimensions. Recommendations will be

developed for how to integrate both aspects of exploration and explanation in data-driven stories.

From Analysis to Communication: Supporting the Lifecycle of a Story

This chapter will explore how tools can better support the authoring of rich and custom data stories with natural / seamless workflows. The aim is to understand the roles and limitations of analysis / authoring tools within current workflow practices and use these insights to suggest opportunities for future research and design. First, the chapter will report a summary of interviews with practitioners at the Dagstuhl seminar; these interviews aim to understand current workflow practices for analysis and authoring, the tools used to support those practices, and pain points in those processes. Then the chapter will reflect on design implications that may improve tool support for the authoring process as well as research opportunities related to such tool support. A strong theme is the interplay between analytical and communicative phases during both creation and consumption of data-driven stories.

The Audience for Data-Driven Stories

Creators of data-driven visual stories want to be as effective as possible in communicating their message. By carefully considering the needs of their audience, content creators can help their readers better understand their content. This chapter will describe four separate characteristics of audience that creators should consider: expertise and familiarity with the topic, the medium, data, and data visualization; expectations about how and what the story will deliver; how the reader uses the interface such as reading, scrolling, or other interactivity; and demographic characteristics of the audience such as age, gender, education, and location. This chapter will discuss how these audience goals match the goals of the creator, be it to inform, persuade, educate, or entertain. Then it will discuss certain risks creators should recognize, such as confusing or offending the reader, or using unfamiliar jargon or technological interfaces. Case studies from a variety of fields including research, media, and government organizations will be presented.

Evaluating Data-Driven Storytelling

The study of data-driven storytelling requires specific guidelines, metrics, and methodologies reflecting their different complex aspects. Evaluation is not only essential for researchers to learn about the quality of data-driven storytelling but also for editorial rooms in media and enterprises to justify the required resources the gathering, analyzing and presentation of data. A framework will be presented that takes the different perspectives of author, audience and publisher and their correspondent criteria into account. Furthermore it connects them with the methods and metrics to provide a roadmap for what and how to measure if these resulting data-driven stories met the goals. In addition, the chapter will explore and define the constraints which might limit the metrics and methods available making it difficult to reach the goals.

Devices and Gadgets for Data Storytelling

This chapter will discuss the role of different hardware devices and media in visual data driven storytelling. The different form factors offer different affordances for data storytelling affecting their suitability to the different data storytelling settings. For example, wall displays

are well suited to synchronous co-located presentation, while watches and virtual reality headsets work better for personal consumption of pre-authored data stories.

Ethics in Data-Driven Visual Storytelling

Is the sample representative, have we thought of the bias of whoever collected or aggregated the data, can we extract a certain conclusion from the dataset, is it implying something the data doesn't cover, does the visual device, or the interaction, or the animation affect the interpretation that the audience can have of the story? Those are questions that anyone that has produced or edited a data-driven visual story has, or at least should have, been confronted with. After introducing the space, and the reasons and implications of ethics in this space, this chapter will look at the risks, caveats, and considerations at every step of the process, from the collection/acquisition of the data, to the analysis, presentation, and publication. Each point will be supported by an example of a successful or flawed ethical consideration.

Conclusion

The main objective of this Dagstuhl seminar was to develop an interdisciplinary research agenda around data-driven storytelling as we seek to develop generalizable findings and tools to support the use of visualization in communicating information. Productive group work converged to delineate several research opportunities moving forward:

- The need for interfaces that enable the fluid movement between exploratory and communicative visualization so that storytelling workflow is seamless and powerful.
- The need to develop typologies of visual storytelling techniques and structures used in practice so that opportunities for supporting these techniques can be sought through computing approaches.
- The need to develop evaluation frameworks that can assess storytelling techniques and tools both scientifically and critically.
- The need for design frameworks that can guide the structure of visual information for experiences across different output devices, both existing and future.
- The need to understand the audience and their role in co-constructing meaning with the author of a data-driven story.
- The need for ethical frameworks that should guide tool development for visual data-driven communication.

These opportunities were productively enumerated at the Dagstuhl seminar and are in the process of being written up as chapters in our book on data-driven storytelling.

References

1 All the medalists: Men's 100-meter freestyle: Racing against history. Web, 2012. http://www.nytimes.com/interactive/2012/08/01/sports/olympics/racing-against-history. html?_r=1&.

2 Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. Understanding data videos: Looking at narrative visualization through the cinematography lens. In *Proceedings of CHI: SIGCHI Conference on Human Factors in Computing System*, pages 1459–1468, New York, NY, USA, 2015. ACM.

3 Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, December 2011.

4 Jeremy Boy, Sara Johansson Fernstad, Martin Turner, Simon Walton, David Ebert, Jean-Daniel Fekete, Andy Kirk, and Mario Romeo. Eurovis 2014 workshop: Towards visualization literacy. visualization literacy workshop. EuroVis 2014, 2014. https://www.kth.se/profile/178785/page/eurovis-2014-workshop-towards-visualiza/.

5 Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think.* Morgan Kaufmann Publishers Inc. San Francisco, CA, USA., 1999.

6 William C. Cleveland and Marylyn E. McGill. *Dynamic Graphics for Statistics (1st ed.).* CRC Press, Inc., Boca Raton, FL, USA., 1999.

7 Instance of data storytelling to explain the subprime mortgage crisis in the usa. Web, 2014. http://www.bloomberg.com/dataview/2014-02-25/bubble-to-bust-to-recovery.html.

8 Hans Rosling. The best stats you've ever seen. TED Talks, 2006. https://youtu.be/hVimVzgtD6w.

9 The wealth report 2013: Examining high net worth individuals from around the world. Web, 2013. http://www.elitehavenssales.com/news/47/The-Wealth-Report-2013-Examining-High-Net-Worth-Individuals-from-around-the-world.

10 J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics.* IEEE CS Press, USA., 2005.

11 Edward Tufte. *The Visual Display of Quantitative Information, Second Edition.* Graphics Press, USA., 1991.

2 Table of Contents

Working groups

3 Overview of Talks

3.1 Data Clips

Fereshteh Amini (University of Manitoba – Winnipeg, CA)

Data videos and animated infographics are becoming an increasingly popular medium for conveying digital stories with data. However, data video creation is challenging as it requires one to hone a unique and combined set of skills in areas such as storyboard design, data visualization, and motion graphics. In this project, we introduce DataClips, an authoring tool designed to help craft animated infographics. We designed DataClips by first undergoing a qualitative examination of existing data videos, authored by reputable sources. This provided us with a rich set of the most important and common visual properties that define a data-driven clip, or what can be considered as being as an atomic unit of a data video.

3.2 Telling Stories about Dynamic Networks with Graph Comics

Benjamin Bach (Microsoft Research – Inria Joint Centre, FR)

Comics are a well known medium to communicate stories by unfolding temporal events in visual space. We explore the design space of comics to present insights into changes in networks.and other visual languages, such as weather charts and assembly instruction, to present insights into changes in networks. However, while comics evolved around stories in "the real world", depicting changes in data (here, dynamic networks) poses several challenges: how to identify data elements across panels? how to switch between overview and detail, how to convey importance of data elements, and so forth. To understand the potential of comics as a storytelling medium, we first created a variety of comics during a 3 month structured design process, involving domain experts from public education and neuroscience. This process led to the definition of 8 design factors for creating graph comics and propose design solutions for each. Results from a qualitative study suggest that a general audience is quickly able understand complex temporal changes through graph comics, provided with minimal textual annotations and no training.

3.3 Affect in vis: Colour and animation

Lyn Bartram (Simon Fraser University – Surrey, CA)

The creation of affect – feelings, experiences, or impressions – has a central role in creating immersive and engaging stories. But affect is also important in the context of how an evoked experience or feeling amplifies, augments or moderates the cognitive interpretation of a visualization: it can enhance interpretation, promote engagement and appropriately "frame"

the information. We are investigating how different properties of color and animation may contribute to affective interpretation of visualizations with the objective of increasing the expressive capacity of visual representation.

3.4   The Works: All the way from data collection to visualizations in Subspotting

Dominikus Baur (Andechs, DE)

Subspotting mapped the cell phone reception on the New York City subway. While officially there is none available, stray signals from aboveground let people access the network at certain places on certain lines. In this project, Daniel Goddemeyer and Dominikus Baur collected and visualized the reception quality on the full subway and turned it into an iPhone app, two posters and an interactive website.

3.5   Thoughts on Visualization literacy

Jeremy Boy (NYU Polytechnic School of Engineering, US)

In this lightning talk I present my thoughts on visualization literacy. I introduce the different building blocks of a common visualization to illustrate what needs to be understood in order to comprehend the visualization. I provide my definition of visualization literacy, and propose a series of perspectives for ongoing and future research. I discuss other takes on literacy, specifically that it is more than simply reading or understanding: it also about being able to write or author visualizations. Finally, I suggest that there are different levels to the concept of visualization literacy, spanning from the low level immediate comprehension of graphics to the higher level comprehension of expressive forms like rhetoric, persuasion, deception or poetry.

3.6   Storytelling with timeline data

Matthew Brehmer (University of British Columbia – Vancouver, CA)

This is a talk about storytelling with timeline or event sequence data. The dominant form of timeline presentation is a linear, chronological encoding pioneered by Joseph Priestley in 1765. But this is certainly not the only form of timeline encoding. Based on a survey of over 250 timeline infographics and timeline visualization tools, we proposed a design space for timelines with three dimensions: representation, scale, and layout. This design space allows a designer or storyteller to present a different narrative points related to different aspects of timeline data, as some combinations of dimensions from our design space are better suited to

communicate different narrative points. We developed an environment for creating timelines that combine these dimensions. But what if you wanted to tell a story that involves more than one type of narrative point? For this reason, we also explored the possibilities for animating timelines, to allow authors to create stories that accommodate different narrative points without loss of context. I present an example story produced using our environment.

3.7 Demo of SandDance using storytelling with unit visualizations

Steven M. Drucker (Microsoft Research – Redmond, US)

SandDance is a visualization system that blends between exploration and presentation. It uses the notion of representing every data element on the screen in a series of visualizations with animated transitions that help both the presenter and the audience maintain their context as different representations highlight different aspects in the data.

3.8 Network Visualisation: a "higher-order" visual analytics for data exploration and communication

Tim Dwyer (Monash University – Caulfield, AU)

Visualisation of quantitative data using basic chart graphics is well understood in its applicability to the analysis of low-dimensional data and time-series. It is clear that people are able to obtain insights (such as patterns, trends, clustering, outliers and so on) through visual inspection of a very simple mapping of such data to a picture. However, machine learning techniques are also very good at identifying such features in simple data. If the ultimate aim of such data analysis is to make connections and build knowledge, then perhaps we should be trying harder to leapfrog the simple mappings of data to chart graphics and visualise the knowledge directly to support a higher-order form of reasoning about data. In this discussion we posit that network visualisation can be used as such a "higher-order" analysis tool. We examine some use-cases where this is already being attempted and opportunities for the visualisation community to further explore this space.

3.9 Towards cultural sensitivity/sensibility: Visualizing texture and structure of cultural collections

Marian Dörk (Fachhochschule Potsdam, DE)

The research project "Visualizing Cultural Collections" (VIKUS – Visualisierung kultureller Sammlungen) investigates graphical user interfaces and the potentials inherent to visual exploration of digitized cultural collections. Within the scope of the project, researchers

from various fields such as information visualization, interface design, and cultural sciences, will collaborate to develop and evaluate graphical interfaces for interactive exploration of cultural objects. Recently, there has been a particular research interest in representations that show it all, while still revealing the individual pieces. Examples include SandDance (Steven Drucker et al.) and Image plots (Lev Manovich et al.). I have demonstrated the prototype visualization "PAST VISIONS penned by Frederick William IV", which shows the collection of historical drawings by Frederick William IV alongside a thematic and temporal arrangement. His drawings are part of the collection of prints and drawings of the Prussian Palaces and Gardens Foundation Berlin-Brandenburg (SPSG) and have previously been published in an online inventory catalogue. Cultural collections poses some unique questions for narrative in visualization: How to pursue the risk of getting lost as an opportunity for open engagement with our cultural heritage? Can we create the opportunity for various readings? In which ways can we treat visualization as both method and object of interpretation?

3.10 10x10 stories

Kennedy Elliott (The Washington Post, US)

My unique perspective on data-driven storytelling comes from my own field of news graphics and news data journalism. The key to our work is simply telling stories that connect with readers – intellectually, emotionally and/or visually. For every project, we consider whether the reader would likely understand the key editorial components of the visual story, if the technical design leads the reader through the story effectively and finally if the technical development is intuitive, bug-free and suits the reader on any particular device. Our data-driven stories involve data types that range from conventional to more abstract: from continuous values to pictures, videos and satellite imagery.

3.11 Thoughts on breaking up the linearity and passivity of video storytelling

Kennedy Elliott (The Washington Post, US)

News organizations and content producers have yet to come up with effective ways to integrate video into textual, data and visual stories. Video is passive and poses and interruption to an otherwise active activity of reading or interacting with a story online. Readers have no control over the speed of progression of the video and surrender to the pace the video storyteller intended. Audio can also be a deterrent in choosing to watch a video. I will present a number of possible solutions to these long-standing problems of video and how they integrate into stories.

3.12 An area of conflicts: The antagonists and promoting factors of dataviz projects in newsrooms

Christina Elmer (Spiegel Online GmbH – Hamburg, DE)

Newsrooms are special ecosystems with their own sometimes strong mechanisms and values system. We as dataviz practitioners should be aware of the forces arising from that. We should use promoting factors like beauty, virality, uniqueness, transparency, personalization and impression of depth – as arguments to gather and to justify the resources and manpower we need for those projects. On the other hand, there are some antagonists that can make it difficult to produce such pieces within newsrooms. Many news topics have to be finished within deadlines, sometimes hours after the investigation starts. The cost pressure in many news outlets limits the resources for dataviz projects and there also is still a quite remarkable gap between technology and journalism to be overcome. The rising mobile usage of news websites influences the possible size and granularity of the visual contents. Thus, both the usual priorities and workflows have to be considered and adjusted to data-driven projects – to enable success within the system and a sustainable change towards more dataviz in the news.

3.13 Practicalities of visual storytelling on the smaller, connected, portable canvas

Xaquin Gonzalez Veira (The Guardian Visuals – London, GB)

In the last year (2015) the Guardian has effectively transition to a mobile first workflow for its interactive storytelling and data visualizations. In that time, we have notice a few patterns and collected a few considerations. There are obvious limitations of the size the canvas itself and landscape v. portrait orientations, but also optimizations on load, limits for autoplaying videos, different in-app browsers, different time commitments, advantages of gestures on touch screens (like in the 'Build your own coalition' interactive) . . . We have also learned that some projects may require a completely different expression of the content. In the 'Mekong River' interactive documentary, the desktop experience is simply a long scrolling page and uses autoplay videos on scroll; the mobile expression, which was developed and conceptualized first, is a set of cards that give you a curated story when swiping vertically, and at certain points, the options of finding out more about that section when swiping horizontally.

3.14   Visualizing the news at The Guardian

Xaquin Gonzalez Veira (The Guardian Visuals – London, GB)

In 2014, The Guardian merged the graphics and interactive teams, motion graphics specialists, interactive documentarians and the picture desk – with a renewed emphasis on digital – under the umbrella of Visuals. The idea is to push for a holistic story editing: each element graphics, text, video, pictures . . . works depending on how it plays in the context of the whole story. It also means a renewed effort on interactive data visualizations, and the size of the department gives it the opportunity to be proactive in finding newsworthy stories. The interactives 'How China's economic slowdown could weigh on the rest of the world' and 'Unaffordable country: Where can you afford to buy a house?', both in collaboration with the also newly created Data team, are an example of our data-driven visual storytelling philosophy, and among other tenets: an emphasis on new angles for stories and a search for affect in the connection between the visuals and the text.

3.15   Using Data-Driven Storytelling to Make Data and Science Relatable

Jessica Hullman (University of Washington – Seattle, US)

As data and scientific results are increasingly prevalent in public spheres, we need tools to help demystify difficult aspects of data and science for interested non-domain experts. I'll describe tools and mechanisms I'm developing that formalize strategies that the best human narrators rely on explain complex concepts. Our work in scientific text simplification learns mappings between complex scientific terms and simpler equivalents from both scientific journals and science news articles. A user of our interactive authoring and reading applications can access simplifications on demand as they read or write about science. Our tools for supporting measurement comprehension take an input measurement that is unfamiliar to a user and output analogies that relate that measurement to familiar objects and landmarks (300 acres is about the area of a well-known park in your town). Our work in visualizing uncertainty shows that presenting uncertainty as an animated set of draws provides a more directly experienceable, countable, and ultimately more understandable experience of probability than static plots of distributions.

3.16   Visualization assembly process as a story?

Samuel Huron (ENST – Paris, FR)

In the last few years I saw more and more data-driven video on YouTube where people are assembling visualization physically or visually. In these videos, people are using movable

independent visual elements map to data unit (I named tokens). These elements could be pictograms, pixels (like in SandDance), stones (like Hans Rosling shortest ted talk), lego bricks, bubble gums and others. In most of these videos the authors support a data-driven narrative, by assembling seamlessly these elements over time to construct more complex representations. During my lightning talk, I explored the following question: How and why the temporal assembly process of such constructed visualization can be a visual story process? To discuss this question, I compared commonalities and differences of the aforementioned examples with some unpublished observations conducted during a previous study. I conclude the lighting talk with a series of open research questions illustrated by other examples of people doing similar videos in various contexts.

3.17 Presentation-oriented vis techniques and how they differ from the usual analytical view

Robert Kosara (Tableau Software – Seattle, US)

Information Visualization is focused on the analysis of data, which shapes the techniques that get published and the criteria used to evaluate them. The goals for presentation are different, however. In particular, generality of techniques is not nearly as important, since the author of a presentation or story can try out different ideas more easily than during analysis. Using three examples, the connected scatterplot, ISOTYPE, and the StreamGraph, I argue that we need to understand techniques better that are useful for presentation and develop the tools to evaluate them.

3.18 Data Stories: Interactivity vs. linear storytelling – which way to go?

Ulrike Köppen (Bayerischer Rundfunk, DE)

Working in data journalism means to find answers to the questions: Do you want your users to lean back or to get involved in the story? Or both? And when exactly? In one of the last BR Data projects we designed an interactive player that combined explanatory and exploratory elements. The idea was to tell a linear story but to allow the user at certain story points to dig into the data (find the project here http://schnee-von-morgen.br.de/stage-1). Some of our takeaways: Don't give the user choices right at the beginning Don't present the whole data set within a story … … but exactly those data sets that fit to the narration at that point

3.19   Classification of common data driving forces

Giuseppe Santucci (Sapienza University of Rome, IT)

Data driven stories use visualizations as building blocks for the story authoring. This talk presents a first classification of the most common data driving forces, i.e., how data characteristics are presented and used in the stories, like functions, proportions, and correlations. Moreover, it discusses how to formally express the mapping between the data and the abstract characteristics of the visualization. One possible outcome of this activity is the definition of a basic set of visualizations that can be used to quickly insert a 'visualization bricks' in the story during the authoring phase.

3.20   Responsible/ethical data visualization

Jonathan Schwabish (Urban Institute – Washington, US)

The changing value and availability of data has democratized people's ability to use data to conduct analysis, show results, and tell visual stories. But that same democratization opens the door for people to misunderstand and misuse data for their purposes. Using data responsibly and ethically is something all data creators and users must take seriously. More generally, presenting data responsibly may encompass a variety of different areas including risk and uncertainty, method and source transparency, data and visualization literacy, cultural, societal, and ethnic diversity of both users and creators.

3.21   Data-Driven Storytelling – An InfoVis/Computer Science Perspective

John T. Stasko (Georgia Institute of Technology – Atlanta, US)

In this talk, I present a perspective on the information visualization (InfoVis) research community by discussing its two main applications: analysis (exploration) and presentation (explanation). Until recently, research for analytic purposes has dominated the community. In particular, InfoVis thrives in situations of exploratory data analysis by employing clever techniques, multiple views, and interaction. Analysis benefits from visualizations that show all the variables, illustrate both overview and detail, and facilitate comparison. Such visualizations may thus require training and practice. Furthermore, visualizations designed for analysis emphasize completeness, accuracy/truth, flexibility and comprehensibility. This seminar focuses on the other application of InfoVis, its use for presentation purposes including storytelling. For this objective, people frequently design with goals of clarifying, focusing, highlighting, simplifying, and potentially persuading, and thus may not show all cases or attributes of the data. Additionally, somewhat in contrast to analysis, visualizations designed for presentation emphasize learnability, comprehension, usability, accuracy/truth, interest, and appeal.

3.22 Data Cuisine

Moritz Stefaner (Truth and Beauty – Lilienthal, DE)

Data Cuisine investigates the language of food to express information. It is a series of participatory workshops organized by Susanne Jaschko and Moritz Stefaner. In the workshops, we team up with a local chef and help participants to invent *data dishes* that represent local data in edible form. The dishes can vary in ingredients, taste, shape, color, temperature, texture, size, presentation, . . . to express information. The workshops end with a tasting and discussion of the developed dishes. We see "foodification" as form of data expression that might be imprecise, but on the other hand sensually intriguing, very association rich, and fosters debate around the data.

3.23 Visual Mementos

Alice Thudt (University of Calgary, CA)

In my talk I discuss the creation of visual mementos as a new application area for visualization. Visual mementos are visualizations of personally relevant data for the purpose of reminiscing, and sharing of life experiences. Today more people collect digital information about their life than ever before. The shift from physical to digital archives poses new challenges and opportunities for self-reflection and self-representation. Drawing on research on autobiographical memory and on the role of artifacts in reminiscing, we identified design challenges for visual mementos: mapping data to evoke familiarity, expressing subjectivity, and obscuring sensitive details for sharing. Visual mementos can make use of the known strengths of visualization in revealing patterns to show the familiar instead of the unexpected, and extend representational mappings beyond the objective to include the more subjective. To understand whether people's subjective views on their past can be reflected in a visual representation, we developed, deployed and studied a technology probe that exemplifies our concept of visual mementos. Our results show how reminiscing has been supported and reveal promising new directions for self-reflection and sharing through visual mementos of personal experiences.

3.24 Cognitive principles for visual explanations and narratives

Barbara Tversky (Stanford University, US)

Telling stories with data brings together many disparate strands of inquiry that have been studied in a variety of disciplines: ancient visual story-telling; perception and comprehension

of the events from which stories are crafted, the production and creation of external representations of various kinds of information, the structures of stories, the understanding of stories, perception, comprehension, memory for various visual displays their use for understanding, inference, discovery, and action. Documenting what and how people create stories and represent data can provide cognitive principles for their design. Crafting visual stories of data depends on all that as well as the constraints of the media, notably print and digital, and the newsworthiness of the stories.

3.25   Exploring Data Sketching

Jagoda Walny (University of Calgary, CA)

Hand-drawn sketching is a common, accessible, and expressive method for generating visual data representations. We studied sketches of a single small dataset drawn by people with varying degrees of visualization experience, along with their reports of what they learned about the data. The sketches revealed a rich representation continuum from numeracy to abstraction; the data reports revealed a data report spectrum from individual data items to speculative data hypothesis. Relating the two results in strong potential directions for data-driven storytelling research.

3.26   Software Engineering in the newsroom

Stefan Wehrmeyer (Correctiv – Berlin, DE)

Software Engineering is present in all areas of media organisations: from the systems administration and IT-support (e.g. email system) to publishing content and tracking usage. Bigger media organisations that are well staffed with technologists can take matters further: they can better engage their audience, conduct investigations based on data and publish custom data visualisations and interactives. However, many older small to medium size media organisations have difficulties deploying these extra resources to take their publishing efforts further. Their constraints are financial, but they also suffer from accumulated technical debt or overreliance on outsourced services. Specifically their Content Management System acts as a gateway and restricts what they can publish. Additionally, journalism processes are focused on producing one-off articles and the editorial part of media organisation is often not used to the iterative nature of software development and the necessity for continued maintenance. Re-use of story-specific software between articles and extraction of their utility into libraries is difficult to fit into the publishing cycle. Many Free and Open Source software products have adopted a main-sponsor / multi-stakeholder approach. Smaller organisations can benefit from production-ready and maintained software without needing the same in-house development capacity as bigger organisations. A shared ownership approach for newsroom software and tools can support small and medium size media organisations that might otherwise struggle to survive.

3.27   Practical Process: An empirical study about visualization design workflows from around the world

Benjamin Wiederkehr (Interactive Things – Zürich, CH)

Visualization has proven to be a strong tool to explore and evaluate complex information. Let's discover how we can use data visualization as a way to tell engaging and compelling stories that are meaningful to your audience. Benjamin Wiederkehr, Managing Director of Interactive Things, shares insights how his team designs and develops visual and interactive stories based on quantitative and qualitative data. Based on a case study from their daily work, the long-form piece "Keine Zeit für Wut" for the Neue Zürcher Zeitung, Benjamin talks about the potentials and pitfalls of a structured process for telling data-driven stories.

3.28   Telling a data-driven story over a year, one day at a time: Cusum charts, cycling world records and community engagement

Jo Wood (City University – London, GB)

Since the early 1900s, cyclists have been telling data stories through their accounts and charts of miles covered by bicycle. Magazines invited readers to chart their weekly mileage and send in their charts for publication. This soon developed into a competition for the furthest ridden in a year. In 1939 Tommy Godwin broke that record with 75065 miles (120805 km). His progress in comparison to previous record holders can be shown as a cumulative distance chart. By rotating and rescaling such a chart, a stronger visual story can be told, in the form of a Cusum chart with the horizontal axis representing time, and the vertical representing distance above or behind a baseline. That baseline represents a goal in the form of a cumulative distance of another rider's progress for comparison. This has the advantage of showing local changes rapidly as they happen, while the smoothing effect of cumulative distance shows the bigger picture. This design was used in a community of cyclists to document the daily progress of five riders who challenged the 75 year old record in 2015/16. This form of chart is amenable to annotation, further increasing its story telling value. It was illustrated by showing the current record holder Kurt Searvogel's progress over the year.

4 Working groups

4.1 Evaluating Data-Driven Stories

Fereshteh Amini (University of Manitoba – Winnipeg, CA), Gordon Bolduan (Universität des Saarlandes, DE), Matthew Brehmer (University of British Columbia – Vancouver, CA), Christina Elmer (Spiegel Online GmbH – Hamburg, DE), and Benjamin Wiederkehr (Interactive Things – Zürich, CH)

The study of data-driven storytelling requires specific guidelines, metrics, and methodologies reflecting their different complex aspects. Evaluation is not only essential for researchers to learn about the quality of data-driven storytelling but also for editorial rooms in media and enterprises to justify the required resources the gathering, analyzing and presentation of data. The following framework takes the different perspectives of author, audience and publisher and their correspondent criteria into account. Furthermore it connects them with the methods and metrics to provide a roadmap for what and how to measure if these resulting data-driven stories met the goals. In addition, we explore and define the constraints which might limit the metrics and methods available making it difficult to reach the goals. We conclude with a graphic that gives an overview how these entities are linked to each other.

4.2 Devices and Gadgets for Data Storytelling: watches to augmented reality

Dominikus Baur (Andechs, DE), Tim Dwyer (Monash University – Caulfield, AU), Xaquin Gonzalez Veira (The Guardian Visuals – London, GB), and Bongshin Lee (Microsoft Research – Redmond, US)

This chapter will discuss the role of different hardware devices and media in visual data driven storytelling. The different form factors offer different affordances for data storytelling affecting their suitability to the different data storytelling settings, as identified by Lee et al. For example, wall displays are well suited to synchronous co-located presentation, while watches and virtual reality headsets work better for personal consumption of pre-authored data stories.

4.3   Ethics in data-driven visual storytelling

Gordon Bolduan (Universität des Saarlandes, DE), Nicholas Diakopoulos (University of Maryland – College Park, US), Marian Dörk (Fachhochschule Potsdam, DE), Kennedy Elliott (The Washington Post, US), and Xaquin Gonzalez Veira (The Guardian Visuals – London, GB)

Is the sample representative, have we thought of the bias of whoever collected or aggregated the data, can we extract that certain conclusion from the dataset, is it implying something the data doesn't cover, does the visual device, or the interaction, or the animation affect the interpretation that the audience can have of the story? Those are questions that anyone that has produced or edited a data-driven visual story has, or at least should have, been confronted with.

After introducing the space, and the reasons and implications of ethics in that space, the chapter will look at the risks, caveats, and considerations at every step of the process, from the collection/acquisition of the data, to the analysis, presentation, and publication, every point supported by an example of a successful or flawed ethical consideration.

4.4   Narrative Design Patterns for Data-Driven Storytelling

Jeremy Boy (NYU Polytechnic School of Engineering, US), Benjamin Bach (Microsoft Research – Inria Joint Centre, FR), Lyn Bartram (Simon Fraser University – Surrey, CA), Paolo Ciuccarelli (Polytechnic University of Milan, IT), Steven M. Drucker (Microsoft Research – Redmond, US), Yuri Engelhardt (University of Twente, NL), Ulrike Köppen (Bayerischer Rundfunk, DE), Moritz Stefaner (Truth and Beauty – Lilienthal, DE), Barbara Tversky (Stanford University, US), and Jo Wood (City University – London, GB)

In this chapter, we introduce a set of narrative design patterns for data-driven stories. These patterns aim to facilitate the creation of rich stories told with data visualization. There are many different ways to tell a same story. Emphasis can be put on certain events rather than others, characters can play more or less important roles, the implication (or voice) of the narrator can be more or less present. These variations usually have different effects on readers' emotions and/or understanding of the sequence of events told in the story. We identify a series of structural narrative patterns that can be used on their own or in combination to tell stories in an almost infinite number of ways. We present fifty of them here, and illustrate how they may help storytellers rethink the stories they want to tell. We assume that these storytellers already know what story they want to tell; who they want to tell it to; and why they want to tell it, i.e., what kind of effects and outcomes they want their stories to have. To frame our data-driven story design patterns, we also propose a threefold defining framework for data-drive stories, based on flow, data perspective, and framing (or context).

4.5   Audience

*Kennedy Elliott (The Washington Post, US), Steven M. Drucker (Microsoft Research –
Redmond, US), Samuel Huron (ENST – Paris, FR), Robert Kosara (Tableau Software –
Seattle, US), and Jonathan Schwabish (Urban Institute – Washington, US)*

Creators of data-driven visual stories want to be as effective as possible in communicating
their message. By carefully considering the needs of their audience, content creators can help
their readers better understand their content. We describe four separate characteristics of
audience that creators should consider: expertise and familiarity with the topic, the medium,
data, and data visualization; expectations about how and what the story will deliver; how the
reader uses the interface such as reading, scrolling, or other interactivity; and demographic
characteristics of the audience such as age, gender, education, and location. We then discuss
how these audience goals match the goals of the creator, be it to inform, persuade, educate,
or entertain. We also discuss certain risks creators should recognize, such as confusing or
offending the reader, or using unfamiliar jargon or technological interfaces. To help support
the chapter, we provide evidence from X case studies from a variety of fields including
research, media, and government organizations.

4.6   Organizational Management

*Jonathan Schwabish (Urban Institute – Washington, US), Christina Elmer (Spiegel Online
GmbH – Hamburg, DE), and Benjamin Wiederkehr (Interactive Things – Zürich, CH)*

Creating and producing data-driven stories requires a variety of personnel, skillsets, and
tools. Be it in news organizations, technology or other private-sector firms, public or non-
governmental organizations, and in the design and education fields, organizational structure
will impact how a data-driven visual story is created, told, and published. In this chapter,
we ask the important question of how organizations set up and run structures for visual
storytelling? Are they hierarchical organizations with multiple levels of management, or are
they flatter with multiple managers? Are content producers broken into separate groups
or pulled together into broader teams? Are there common practices across these different
sectors and what appear to be best practices and management structures?

To write this chapter, each author will speak to a handful of experts in the field of
data-driven storytelling to see how they group and organize their production process. These
interviews, as well as the existing literature on management organizations, will inform the
basis of the chapter.

4.7   Exploration and Explanation in Data Driven Stories

Alice Thudt (University of Calgary, CA), Jason Dykes (City University – London, GB), Theresia Gschwandtner (TU Wien, AT), John T. Stasko (Georgia Institute of Technology – Atlanta, US), and Jagoda Walny (University of Calgary, CA)

In our group we discussed the difference between and integration of exploration and explanation in visual data-driven storytelling. Exploratory visualizations allow for a lot of freedom which can include changing the visual representation, the focus of what is being shown, and the sequence in which the data is viewed. They allow readers to find their own stories in the data. Explanatory stories include a focussed message which is usually more narrow and guides the reader usually in a linear way. We identified some advantages and some disadvantages of exploration and explanation as well as dimensions that help to describe and classify data-driven stories. We begin to describe the space by identifying freedom, guidance regarding representation, focus and sequence as well as interpretation as important dimensions of data-driven storytelling and characterize existing systems along these dimensions. As an outcome we aim to describe the space of exploration and explanation in data-driven stories in light of recent vis research and exploratory visualization and provide practical recommendations on ways in which we might integrate both aspects.

4.8   Techniques and Design Choices for Storytelling

Robert Kosara (Tableau Software – Seattle, US), Fereshteh Amini (University of Manitoba – Winnipeg, CA), Benjamin Bach (Microsoft Research – Inria Joint Centre, FR), Matthew Brehmer (University of British Columbia – Vancouver, CA), and Stefan Wehrmeyer (Correctiv – Berlin, DE)

We discuss techniques and design choices for visual storytelling grounded in a survey of over 60 examples collected from various online news sources and from KIIB award-winning visualization and infographic design work. These design choices represent a middle ground between low-level visualization and interaction techniques and high-level narrative devices or structures. We define several classes of design choices: embellishment, explanation, exploration, navigation, story presentation, emphasis, focus, and annotation. We provide examples from our survey for each class of design choices. Finally, we offer several case studies of examples from our survey that make use of multiple design choices.

4.9 From analysis to communication: supporting the lifecycle of a story

Melanie Tory (Tableau Software – Palo Alto, US), Fanny Chevalier (INRIA, FR), Marian Dörk (Fachhochschule Potsdam, DE), Jessica Hullman (University of Washington – Seattle, US), Bongshin Lee (Microsoft Research – Redmond, US), Giuseppe Santucci (Sapienza University of Rome, IT), Jarke J. van Wijk (TU Eindhoven, NL), and Stefan Wehrmeyer (Correctiv – Berlin, DE)

We will explore how tools can better support the authoring of rich and custom data stories, with natural / seamless workflows. We aim to understand the roles and limitations of analysis / authoring tools within current workflow practices and use these insights to suggest opportunities for future research and design.

We will first report a summary of interviews with practitioners; these interviews aim to understand current workflow practices for analysis and authoring, the tools used to support those practices, and pain points in those processes. We will then reflect on design implications that may improve tool support for the authoring process as well as research opportunities related to such tool support. A strong theme is the interplay between analytical and communicative phases during both creation and consumption of data driven stories.

Participants

- Fereshteh Amini
University of Manitoba –
Winnipeg, CA
- Benjamin Bach
Microsoft Research – Inria Joint
Centre, FR
- Lyn Bartram
Simon Fraser University –
Surrey, CA
- Dominikus Baur
Andechs, DE
- Gordon Bolduan
Universität des Saarlandes, DE
- Jeremy Boy
NYU Polytechnic School of
Engineering, US
- Matthew Brehmer
University of British Columbia –
Vancouver, CA
- Sheelagh Carpendale
University of Calgary, CA
- Fanny Chevalier
INRIA, FR
- Paolo Ciuccarelli
Polytechnic University
of Milan, IT
- Nicholas Diakopoulos
University of Maryland –
College Park, US
- Marian Dörk
Fachhochschule Potsdam, DE
- Steven M. Drucker
Microsoft Research –
Redmond, US
- Tim Dwyer
Monash University –
Caulfield, AU
- Jason Dykes
City University – London, GB
- Kennedy Elliott
The Washington Post, US
- Christina Elmer
Spiegel Online GmbH –
Hamburg, DE
- Yuri Engelhardt
University of Twente, NL
- Xaquin Gonzalez Veira
The Guardian Visuals –
London, GB
- Theresia Gschwandtner
TU Wien, AT
- Nathalie Henry Riche
Microsoft Research –
Redmond, US
- Jessica Hullman
University of Washington –
Seattle, US
- Samuel Huron
ENST – Paris, FR
- Christophe Hurter
ENAC – Toulouse, FR
- Ulrike Köppen
Bayerischer Rundfunk, DE
- Robert Kosara
Tableau Software – Seattle, US
- Bongshin Lee
Microsoft Research –
Redmond, US
- Giuseppe Santucci
Sapienza University of Rome, IT
- Jonathan Schwabish
Urban Institute –
Washington, US
- John T. Stasko
Georgia Institute of Technology –
Atlanta, US
- Moritz Stefaner
Truth & Beauty – Lilienthal, DE
- Alice Thudt
University of Calgary, CA
- Melanie Tory
Tableau Software – Palo Alto, US
- Barbara Tversky
Stanford University, US
- Jarke J. van Wijk
TU Eindhoven, NL
- Jagoda Walny
University of Calgary, CA
- Stefan Wehrmeyer
Correctiv – Berlin, DE
- Benjamin Wiederkehr
Interactive Things – Zürich, CH
- Jo Wood
City University – London, GB