Visualization Viewpoints

Editor:
Theresa-Marie Rhyne

More Than Telling a Story: Transforming Data into Visually Shared Stories

Bongshin Lee and Nathalie Henry Riche
Microsoft Research

Petra Isenberg
Inria

Sheelagh Carpendale
University of Calgary

Visualizations are increasingly used "in the wild" to tell compelling stories supported by data. Examples include Hans Rosling's talks on human development trends using Gapminder (gapminder.org) and the use of visualizations on the Web (see Figure 1 for an example) including news media such as the *New York Times* and the *Guardian*. Given the practical power visualizations can have in communicating facts and opinions, the visualization research community has begun to pay more attention to the need and use of visualization as a storytelling medium to tell compelling data stories.

In this article, we take a closer look at how the visualization community has discussed visual storytelling. We argue that so far the community has been using the term "storytelling" in a broad way, without a clear consensus or discussion on what a visual data story encompasses. Yet, despite this wide view on what constitutes a data story, visualization research on storytelling has mainly centered on how data visualization components contribute to communication. Instead, we argue for first narrowing the scope of what visual data stories encompass and then expanding our research focus to the whole process of transforming data into visually shared stories, including formative processes such as crafting narrative structures.

Narrowing the scope of what is termed a *data story*, for instance, by distinguishing between a visual data story and a data visualization, helps us open the door for a more detailed examination, covering the aspects of the visual data storytelling process that have so far received less research attention. We discuss the entire process of transforming data into visually shared stories, incorporating steps involved in finding insights (explore data), turning these insights into a narrative (make a story), and communicating this narrative to an audience (tell a story). We conclude with research opportunities in visualization as a storytelling medium in the light of this broader process.

Storytelling Scope

The word "storytelling" has a long history during which it has been used in many ways in different domains. As a result, it invokes meaning and nuance without having one single agreed-upon definition. Yet, most descriptions of storytelling require some sort of controlled delivery or presentation of information.[1] This delivery can be with or without a predefined temporal or narrative structure, but it always contains components that form the "story" (structures, elements, and concepts) and those that influence the "telling" part of storytelling (people, tools, and channels).

Similarly, no single crisp definition of a "visual data story" has emerged. We argue against using the broadest sense, which would consider any shared visualization to be a data story, given that visualization is by definition focused on making data visible—and thus on presenting data to a viewer. A visualization can be thought of as having both design elements that form a story and

 Published by the IEEE Computer Society

presentation methods that have been used to tell it.[2] If we take this broadest view on visual data stories, however, any images containing even simple charts with little explanations or reading aids would be called visual data stories. To facilitate a more practical discussion on the structure of visual data storytelling in the visualization community, we need to consider narrowing the scope of visual data stories.

Several research papers have considered this problem of defining what storytelling in the domain of visualization encompasses. Edward Segel and Jeffrey Heer introduced the term *narrative visualization*, emphasizing visual stories that include a narrative, or a series of causally related events.[3] They also showed that storytelling in the context of visualization cannot be separated from talking about visual narratives, but it still needs to be more concretely defined—in particular because the differences and commonalities to traditional storytelling mechanisms and strategies are not well understood yet. They attempted to develop a better understanding of visual data stories by formulating the design space of narrative visualization. Jessica Hullman and Nicholas Diakopoulos began with Segel and Heer's term narrative visualization and defined it as a genre that combines interaction techniques for exploratory control over insights gained and communicative, rhetorical, and persuasive techniques for conveying an intended "story."[4] Similar to Segel and Heer, the authors did not clearly define what constitutes a visual data story but aim to better describe the visualization techniques involved in communicating an intended message. Although these two approaches are a great first step, as they are valuable for understanding communication mechanisms for visualization in general, it is still unclear exactly what is or is not a visual data story.

In discussing the process of transforming data into visually shared stories, we propose that visual data stories can be thought of as follows:

- A visual data story includes a set of story pieces—that is, specific facts backed up by data (such as how energy consumption has changed over the years).
- Most of the story pieces are visualized to support one or more intended messages. The visualization includes annotations (labels, pointers, text, and such) or narration to clearly highlight and emphasize this message and to avoid ambiguity (especially for asynchronous storytelling).
- Story pieces are presented in a meaningful order or with a connection between them to support the author's high-level communication goal, which can range from educating or entertaining the viewers with illustration of facts to convincing or persuading them with thought-provoking opinions.

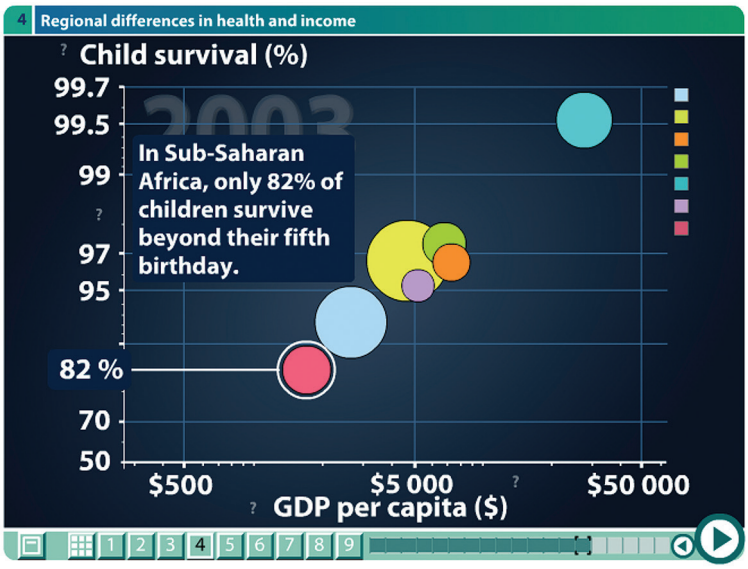

Figure 1. Example visual data story. A full Flash presentation, "Human Development Trends, 2005," is available at www.gapminder.org/downloads/human-development-trends-2005.

Taking this narrower view of a visual data story, we can exclude several types of stories not related to visualization, allowing us to uncover research opportunities relevant to the visualization community. For example, typical stories told in books and movies are not pertinent to our discussion unless they are focused around data visualizations. In addition, we exclude Web-based interactive visualizations that support completely free exploration without any guidance. We also do not consider visual data stories to include, for example, charts posted on the Web unless they are enriched with written explanations or annotations that help the viewer capture the intended message. For example, many visualizations that include only a title above the chart usually leave the interpretation of the content entirely to the viewer.

It is important to note that these last examples—simple charts created from data—can be turned into visual data stories according to our narrower scope if the author added appropriate annotations for a simple story or combined several to form a longer, more complex story. However, this process of turning data into a visual story has so far received little attention. We argue that it is just as important to consider research on the process of creating a visual data story as it is to understand which features best communicate it. We describe such a possible process and open areas for research next.

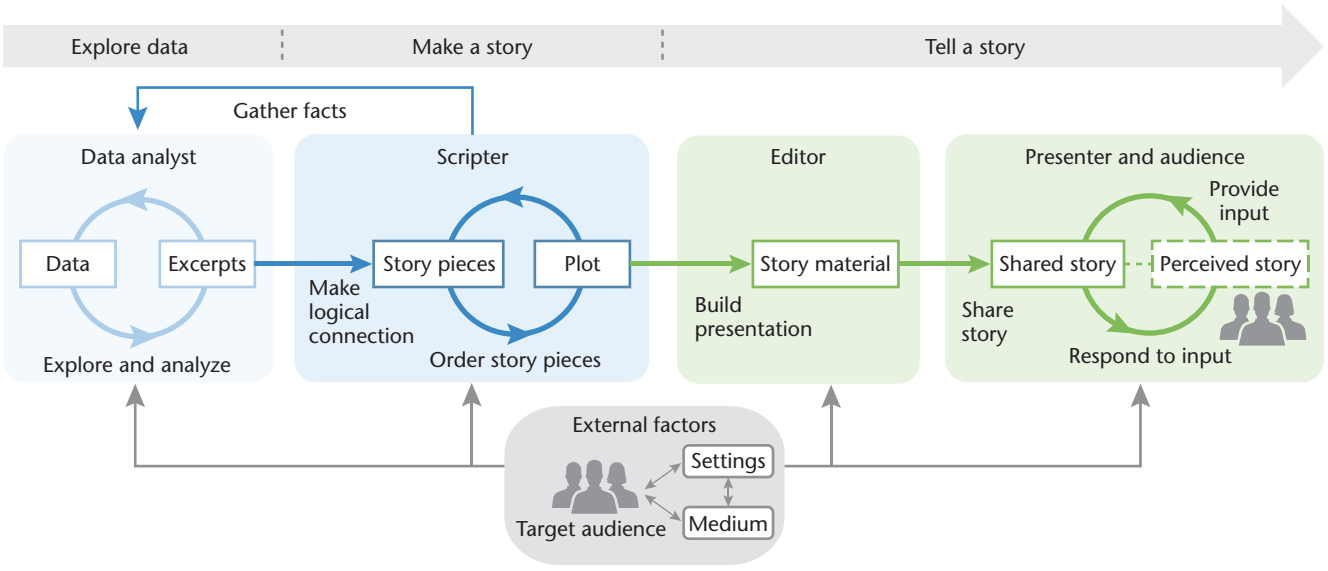

Figure 2. Storytelling process. The visual data storytelling process involves transforming data into visually shared stories via three main components (exploring data, making a story, and telling a story), although the progression need not be linear.

Visual Data Storytelling Process

In the past, few research papers have considered a wider visual storytelling process. An exception is the article by Robert Kosara and Jock Mackinlay,[5] which aimed to provide a high-level overview of the state of storytelling research. They briefly mentioned a working model for how stories are constructed based on how journalists work and argued that visualization tools for analysis are rarely usable for presentation purposes.

Here, we contribute a much more detailed description of the storytelling process in visualization with regard to activities, artifacts, and roles involved to develop a more encompassing look at the visual storytelling process and to uncover open areas for research. We propose the visual data storytelling process (VDSP) (see Figure 2) as a working model, which is derived from other models in the data journalism literature.[6] The VDSP summarizes the main roles and activities that visualization storytellers engage in as they turn raw data into a visually shared story, along with the types of artifacts that result from these activities.

While the storytelling process is linear in Figure 2 for illustration purposes, it does not have to begin on the left and move to the right; it can contain many loops and multiple occurrences of each component. For the sake of simplicity, the three main components—exploring data, making a story, and telling a story—are introduced with their respective artifacts in a linear order.

Exploring Data

Exploring data involves the set of activities centered around exploring and analyzing data. Data is the raw material that constitutes the source of the visual data story content. Pertinent data excerpts are collected through exploratory analysis. These may be simple, such as recorded data facts or steps from the analysis process. They may be more complex, such as derived data insights, interesting sets or sequences within the data, and/or process details and variations. They may include the first quick externalizations of the data, such as charts from spreadsheets or hand sketches made during the analysis. At this point in the process, this collection of excerpts may or may not be tied to any specific visual representation. The result of data exploration when making a visual data story is a collection of the chosen data excerpts.

Making a Story

To make a story, the data excerpts gathered in step one need to be assembled into a storyline that is interesting, illuminating, and compelling. The sequence plays a critical role in a story; the same set of excerpts can have impact or can fall flat. A significant part of making a story is the process of constructing the storyline or plot. The activities involved are ordering, establishing logical connections, developing flow, formulating a message, and creating the denouement. These activities that are often intertwined may be achieved sequentially, simultaneously, or through multiple iterations. Furthermore, while developing the storyline, it may be necessary to go back to the explore data stage to gather more excerpts (insights or evidence). The final outcome of this step in the process is the plot of a story that describes how the story pieces are related (in time, cause and effect, patterns, and so on) and what they mean in an overall context.

Table 1. Four common storytelling settings.

Example scenario	Time	Place	Audience participation level
Live presentations	Synchronous	Colocated	Low
Dynamic discussions	Synchronous	Colocated	High
Recorded videos, static infographics	Asynchronous	Distributed	Low
Guided tours, interactive infographics	Asynchronous	Distributed	High

Telling a Story

Telling a story is the general process of materializing the abstract plot and delivering the story. It consists of the following activities: building a presentation (creating story material with the chosen medium), sharing the story using the story material, and finally receiving and handling the feedback from the audience.

In the building phase of telling a story, a plot and story pieces are taken and turned into story material. Story material is the materialization of each piece of this abstract content through the development of visual representations, interactions, animations, annotations, or narration. For example, story material could be one or more visualizations assembled in a slide deck, a video with narration, an infographic presented on a poster, or a demo planned with an interactive system for the live presentation.

This story material turns into a shared story once it is delivered to at least one person. Ultimately, the perceived story is what the audience understands through the storytelling experience. We acknowledge that, in the current practice, making a story is often merged with the presentation building phase of the telling a story component. For example, a storyteller may sequence specific visual representations to build a plot while making a presentation. However, by making the distinction explicit in the process we can capture many other ways these two components can interact. Also, making a story and building story material require different skill sets, involving the two different roles we describe next.

Roles in the Visual Storytelling Process

A number of different roles are involved in turning data into a visually shared story:

- The data analyst engages in the process of exploring and analyzing the data that is the foundation of the story.
- A scripter builds the plot using the chosen excerpts.
- The editor prepares the story material.
- The presenter is responsible for delivering the story.
- The audience experiences the story and provides feedback.

Note that it is possible (and often likely) that one person plays multiple roles. For example, one person can analyze the data, build a plot, create the story material, and then present it. Yet, in many professional settings, it is also possible that the roles are taken on by different people. For example, professional analysts or statisticians may be hired to dig through the data and provide interesting excerpts, a journalist may take this input and build a plot, a graphic designer may prepare infographics using the plot with the prepared story pieces, and a professional typesetter will include the infographics in the news medium.

External Factors: Audience, Setting, and Medium

A set of external factors may impact the storytelling process. In every step (when searching for data, creating a plot with story pieces, building story material, or delivering a visual story), considerations regarding the target audience at the receiving end of the shared story are typically taken into account. A next factor concerns the setting (context), which determines the way a visual story is presented. On a higher level, one can characterize the setting by time and place where the story is conveyed. The setting can be further characterized by the level of possible audience participation. Table 1 shows how different combinations of the setting can lead to commonly encountered storytelling scenarios with data.

The medium is another factor that influences how the story material will be created and presented as well as consumed. For example, a story consumption experience with a static image significantly differs from that with an interactive infographics, supporting a simple, controlled interaction. The choice of media (for example, video, images, text, narration, and interactive system) comes with presentation parameters that are more or less appropriate (such as color palettes or fonts).

Ethics in Visual Data Storytelling

When visualizations are leveraged to make intended messages more comprehensible and persuasive, visual representations can be misused unintentionally, offer controversial emphasis, or even be abused intentionally. Alexander B. Howard argues that, given the rapidly increasing amount of

digital media, data journalists must be more conscious of ethics than ever before.[7] It seems sensible that the ethics maintained in journalism should be upheld in visual data storytelling as well. To increase the transparency and credibility of visual data stories throughout the visual data storytelling process, we need to make it easier and more desirable for authors to share underlying data and their analysis process.

Research Opportunities

The storytelling process we have described here helps us identify and articulate several research opportunities. We first discuss missing efforts for the two components of the visual data storytelling process—making a story and telling a story. We then reflect on whether it is desirable to build an encompassing tool that supports the entire visual data storytelling process. Finally, we underline the need to develop novel methodologies to assess the quality of both authoring and experiencing visual stories.

We need to make it easier and more desirable for authors to share underlying data and their analysis process.

Help People Make a Compelling Story

While the visualization community has a plethora of techniques supporting the exploration and analysis of data, not much attention has been put on helping a scripter make a story—that is, identifying a sequence of compelling story pieces. In practice, people often make a story (even if just in their minds) while they are either exploring the data or building a presentation. Yet, visualization systems do not usually incorporate help for the general making a story phase, and they particularly fall short in supporting people in the steps to collect and organize excerpts from data exploration that are potentially interesting for the final visual data story. Furthermore, visualization systems do not usually provide support for building a plot from such a collection. As a result, people have to keep track of a set of interesting visualizations, often resorting to capturing them on an auxiliary storage such as creating a document with screen captures and notes. IBM's i2 Analyst's Notebook is a rare exception in that it is aimed at streamlining this process (see www-03.ibm.com/software/products/en/analysts-notebook). However, there is no explicit support for selecting and ordering these elements to form a coherent plot with a beginning and an end.

Because the structure and sequence in stories (the plot) can influence the reception of a story, Hullman and her colleagues recently investigated the forms and reactions to sequencing in linear, slideshow-style presentations using visualizations.[8] There are research opportunities in further exploring ways to provide guidelines for making a good story, suggesting different plots, and letting people experiment with several plots and select the most compelling one. For example, as work on Show Me[9] suggests an appropriate visualization for given data properties such as data type and dimension with other possibilities, a system for story building can automatically suggest possible data stories for given story properties (such as story point, intended message, and goal). This could benefit novices as well as people with extensive training in communication (such as data journalists) or with inherent storytelling skills (the ability to create a compelling story).

Make It Easier to Tell a Story

Even authors inherently skilled at creating compelling story structures face the challenge of producing the story material. This step may prove extremely difficult and time consuming for people with low programming skills. Several online tools (such as Many Eyes) are accessible to nonexperts, offering a set of visualizations that are based on predefined templates. They usually support the creation of visualizations through three main steps: import data, select the visualization type, and configure the visual attributes (such as color palettes). Although there is some support for simple interactions, these are usually limited to things such as mouse over and click, making these tools largely only suitable for rather simple stories.

To support more sophisticated interactions and animations, people have resorted to programming using libraries such as D3 (www.d3.org). These libraries have been adopted by many designers and data analysts, but they require programming skills. This motivated recent research on helping people with little or no programming skills create custom visualizations. A recent example is Ellipsis, a graphical interface for story authoring with preexisting visualizations and the Ellipsis domain-specific language (DSL).[10] It is an interesting research space to support the easy creation of custom annotations through direct manipulation and reuse of existing story elements.

Most of the existing tools are meant to help peo-

ple create material that can be shared on the Web asynchronously. However, a presentation based on an interactive visualization could allow the presenter to be more responsive to the audience by interacting with the data and showing the action in real time. For example, Tableau introduced a feature called Story Points[11] that allows people to capture a set of visualizations with specific states as a dashboard (a story point) within Tableau. These can be recorded with captions and reordered with simple drag-and-drop interaction. Designing an easy-to-use, yet powerful experience for authoring the story material for interactive visualization system is an interesting future research topic.

Delivering a story live also presents a unique opportunity to perceive the audience's reaction as the story is told, potentially allowing the presenter to adjust the presentation style in real time to provide a better experience. There are opportunities to help the presenter monitor audience feedback and assess their engagement as well as to facilitate or suggest alteration to the ongoing presentation. In addition, other recent research has focused on techniques to better engage the audience. For example, SketchStory helps presenters perform an engaging storytelling by attracting attention and creating anticipation.[12] However, it only offers limited annotation and interaction with the data during the presentation. More research on supporting advanced features such as emphasizing different components via annotation, highlighting, and zooming during a live presentation is a promising avenue for research.

One Tool for the Whole Process?

One research opportunity is an attempt to combine data analysis, scripting, editing, and presenting functionality all into one tool or a suite of tools. However, we need to reflect on whether this is a desirable outcome.

On the one hand, a holistic tool may facilitate every requirement of each of the storytelling components and aid people in taking on the different roles in the process. For example, the scripter's work may be significantly lessened as new or updated data is processed by the analyst. Thus, the design of an encompassing tool or a suite of tools that allows a fluid switch between different storytelling components and activities, keeps track of the data, and updates story material as updated data appears seems like a promising endeavor. On the other hand, such an encompassing tool may either be extremely feature-rich to support various data types, visualization styles, and presentation techniques, making it difficult to learn and use, or limited in functionality and, thus, only useful for certain types of stories with certain types of data.

Beyond Evaluating a Story

Kosara and Mackinlay discussed evaluation as a research challenge with a focus on understanding the effectiveness of a story in conveying its message.[5] Although we agree that this question is important, evaluation challenges exist beyond studying the story itself. Understanding where, when, and why the visual storytelling process is successful is another fundamental challenge to tackle.

Understanding why a story is or is not effective may require an answer that goes beyond looking at story delivery and the story's reception by the audience. Several factors in the visual data storytelling process can influence a story's success and can lead to poorly delivered stories still being successful or brilliantly delivered stories failing. Designing evaluation methodologies across the process we described earlier is, thus, a major research question. It is important to derive study protocols, new methods, and analysis practices for each phase of the visual data storytelling process and attempt to evaluate the process as a whole.

Because storytelling involves either people taking on multiple roles (data analyst, scripter, editor, and presenter) or several people working together, the story creation and the delivery process can become a collaborative activity. As such, a wealth of research opportunities open up. For example, collaborative evaluation metrics could be applied or we could study communication successes and failures between different people involved in the process.

Emerging Scenarios

Our visual data storytelling process supports several scenarios involving storytelling with data by taking different paths through the components and different manifestations of each of the components. We envision that other storytelling opportunities exist that include visualization support, and it is interesting to consider whether our visual data storytelling process could be expanded, by adding steps or activities, or modified to more clearly capture emerging scenarios. For example, our storytelling process could support the audience actually taking on the role of data analyst, scripter, and/or editor. Involving the audience in the story creation process by opening up the editing and scripting roles is an exciting emerging scenario. For example, Sprint's Global Water Experiment (www.visualizing.org/sprint/global-water-experiment), where the audience could edit an initial visualization and

data, illustrates the potential of collaborative story creation. Such emerging scenarios can empower the audience and make the raw data and the story creation process more transparent.

We aim to facilitate better structured discussions around compelling techniques for storytelling with data visualization by drawing a line between a visual data story and a general data visualization and narrowing the scope of what a visual data story is. Considering the entire process of transforming data into visually shared stories along with this more focused definition, we believe it is possible to widen the scope of research around visual data stories. By pursuing these new avenues of research in visual data storytelling process, visualizations can enable more effective storytelling with data.

References

1. B. Alexander, *The New Digital Storytelling: Creating Narratives with New Media*, Praeger, 2011.
2. T. Strothotte and C. Strothotte, *Seeing between the Pixels*, Springer-Verlag, 2003.
3. E. Segel and J. Heer, "Narrative Visualization: Telling Stories with Data," *IEEE Trans. Visualization and Computer Graphics*, vol. 16, no. 6, 2010, pp. 1139–1148.
4. J. Hullman and N. Diakopoulos, "Visualization Rhetoric: Framing Effects in Narrative Visualization," *IEEE Trans. Visualization and Computer Graphics*, vol. 17, no. 12, 2011, pp. 2231–2240.
5. R. Kosara and J. Mackinlay, "Storytelling: The Next Step for Visualization," *Computer*, vol. 46, no. 5, 2013, pp. 44–50.
6. J. Gray, L. Chambers, and L. Bounegru, *The Data Journalism Handbook*, O'Reilly Media, 2012.
7. A.B. Howard, *The Art and Science of Data-Driven Journalism*, Tow Center for Digital Journalism; http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf.
8. J. Hullman et al., "A Deeper Understanding of Sequence in Narrative Visualization," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2406–2415.
9. J. Mackinlay, P. Hanrahan, and C. Stolte, "Show Me: Automatic Presentation for Visual Analysis," *IEEE Trans. Visualization and Computer Graphics*, vol. 13, no. 6, 2007, pp. 1137–1144.
10. A. Satyanarayan and J. Heer, "Authoring Narrative Visualizations with Ellipsis," *Computer Graphics Forum (Proc. EuroVis)*, vol. 33, no. 3, 2014, pp. 361–370.
11. E. Fields, "8.2 Preview: Story Points," Tableau Software, 20 May 2014; www.tableausoftware.com/about/blog/2014/5/82-preview-tell-story-your-data-story-points-30761.
12. B. Lee, R.H. Kazi, and G. Smith, "SketchStory: Telling More Engaging Stories with Data through Freeform Sketching," *IEEE Trans. Visualization and Computer Graphics*, vol. 19, no. 12, 2013, pp. 2416–2425.

Bongshin Lee *is a researcher in Microsoft Research's neXus Group. Contact her at bongshin@microsoft.com.*

Nathalie Henry Riche *is a researcher in Microsoft Research's neXus Group. Contact her at nath@microsoft.com.*

Petra Isenberg *is a research scientist in Inria's Aviz project. Contact her at petra.isenberg@inria.fr.*

Sheelagh Carpendale *is a professor at the University of Calgary. Contact her at sheelagh@ucalgary.ca.*

Contact department editor Theresa-Marie Rhyne at theresamarierhyne@gmail.com.