

Telling Stories with Data Visualization

María Teresa Rodríguez
INESC TEC,
Faculdade de Engenharia,
Universidade do Porto
mrodriguez@fe.up.pt

Sérgio Nunes
INESC TEC, DEI,
Faculdade de Engenharia,
Universidade do Porto
sergio.nunes@fe.up.pt

Tiago Devezas
Labs. SAPO/U.Porto,
Faculdade de Engenharia,
Universidade do Porto
tdevezas@fe.up.pt

Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

ABSTRACT

In this article we survey the historical background and development of information and data visualization, and an overview of the intersection of data visualization with storytelling applied to the field of data journalism, where it finds its most widespread use in narrative visualizations. We start by explaining why the mere act of visualization can be highly useful to readers, helping them discover patterns and comprehend information. Backed by historical references, we will describe how some of the first data visualizations were used to explain facts, understand certain events, and determine courses of action. We will then outline how storytelling and narrative techniques are being currently used with data visualization to leverage the power of visual expression. Our goal is to characterize storytelling with data as a vibrant and interesting field that current journalism practices employ to help readers understand and form opinions on complex facts. By presenting concepts like storytelling with data and data stories, we aim to spark interest in further research in the applications of data visualization and narrative.

Keywords

data journalism, data visualization, storytelling

1. INTRODUCTION

In recent years there have been significant advances in displaying massive amounts of data to make the information contained in the data easily accessible and understandable. Major contributors to this exponential growth include the open data movement [1], the collection of information by internet companies, and the surge of social media. Other developments witnessed since the inception of the internet have to do with how the information and media is produced and consumed, creating new models of media consumption. One of these new models of consumption can be experienced in “narrative visualization”, a combination of storytelling techniques and interactive data visualizations. Examples of this type of work are regularly used as narrative artifacts

in the data journalism of news sites like The New York Times [2] and The Guardian [3]. Every day, applications and continuous use of visualization techniques have established that the main purpose of using visualizations is to provide “insight” on a particular topic, make sense of it, and then help the reader form an opinion on it. On the other hand, storytelling and visual expression are ancient forms of expression, rooted as part of human culture and perform an indispensable role in our inherent ability to make sense of the world around us. In Section 2 we describe how visualization leverages human cognition to help viewers process and comprehend information efficiently. Section 3 discusses the intersection of storytelling and data visualization in the field of journalism, and the approaches proposed by different authors. Finally, in Section 4, we discuss some of the benefits and challenges of incorporating storytelling strategies in data visualization projects, as well as potential future directions of research.

2. THE WHY BEHIND VISUALIZATION

The visualization of information is important due to its ability to help the human eye “see” things that are difficult, or impossible, to understand in large datasets and abstract textual based information. From the field of human-computer interaction, Card explains the capacity of the human brain to comprehend visual stimuli and proposes the notion that our limited mental capacity is assisted by a method of “externalizing” cognition [4]. In this sense, he offered an example of external cognition by describing how much easier it is to learn a task like multiplication by doing it on a piece of paper, instead of only doing it mentally.

This process of external cognition, using visual means as a basis, informs why “the fields of Perception, Graphic Design, Information Visualization, Statistics, Data Mining, and Cartography”, can be brought together as a “unified approach for the effective handling of complex data” [5].

2.1 Perception and insight

Ware explains the important role visualization plays in cognitive systems [6]. He argues that “visual displays present the highest bandwidth channel from the computer to the human”, making around 20 billion neurons in the brain capable of processing and analyzing visual information to provide a mechanism able to find important patterns. He also argues that the term *visualization* has come to mean something more than a mere image of something in the mind, but “an external artifact that supports decision making”.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
NHT'15, September 1, 2015, Guzelyurt, Northern Cyprus.
Copyright 2015 ACM ISBN 978-1-4503-3797-7/15/09 \$15.00.
<http://dx.doi.org/10.1145/2804565.2804567>.

“Providing insight” is considered to be the main objective of using information visualization and the by-product of an ongoing exploration with the visualization [7], and is important to understand what this process consists of. It includes four distinct phases:

1. Provide Overview: it allows a person to grasp the big picture concept of a dataset, allowing to understand what they know and don’t know.
2. Adjust: the exploration of a dataset by adjusting the level of abstraction, enabling the person to change perspective on the dataset making sense of different aspects.
3. Detect Pattern: the person usually finds what he was looking for and gains new knowledge by finding trends, frequencies, and structures in the dataset.
4. Match Mental Model: metaphors enable an effective mapping of data by allowing people to match the visual representation of the dataset with a mental model.

Dove and Jones [8] view insight as a two-part process: first, the insight experience of gaining new understanding or knowledge; and secondly, the consequence or product gained from the exploration of a visualization. This second part is how the InfoVis community generally understands insight. For Dove and Jones this communication of insights is the best opportunity narrative visualization can offer to not only consider the product of the visualization but also to consider the experience. An excellent example of considering experience as part of this process is the project “The Fallen of World War II” by Neil Halloran [9], where a data visualization of the number of dead civilians and soldiers during World War II is presented in an interactive and cinematic form, and the experience of the ravages of war is clearly palpable in the presentation of the data of the loss.

2.2 Data visualization

It is usual to think about data visualization as a modern application in statistical graphs and charts, where it finds its wider and most common application. But in reality, “the modern representation of quantitative information has deep roots”, and we can trace them all the way to the earliest applications in cartography and statistical graphics [10].

Although data visualization techniques have been studied and developed during the last 20 years, like the recent example in the melding of storytelling and visualization, the *Gap-minder* [11] project, information visualization examples have been created to show and explain certain events throughout history. Figure 1 displays an early example of information visualization in Florence Nightingale’s diagrams [12] to analyze the causes of mortality of the British army in India during the Crimean War in the 1850’s, to present this information to the Queen and Members of the Parliament.

Tufte [13, 14] identifies how storytelling is present in data visualization throughout history in the dot map designed by the English physician John Snow crafted to explain the outbreak and dispersion of cholera cases that occurred in London in 1854, and by Charles Minard’s description of Napoleon’s campaign to Moscow in 1812 using a map, seen in Figure 2, to represent in two dimensions six types of data [15], and functioned primarily as a means to tell the story of the soldier’s struggle leading to a disastrous and unfruitful

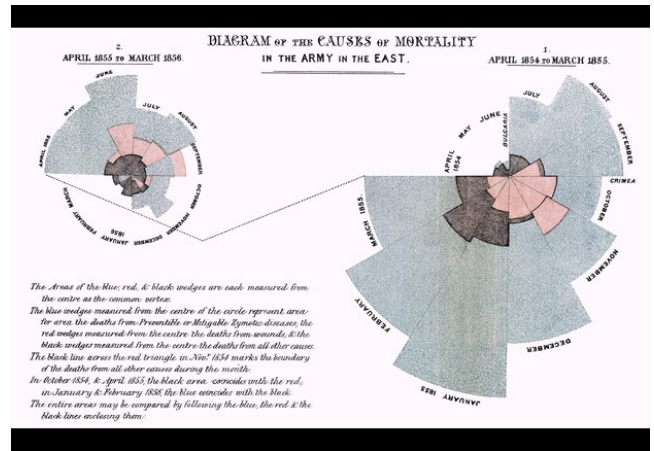


Figure 1: Mortality of the British Army by Florence Nightingale [12].

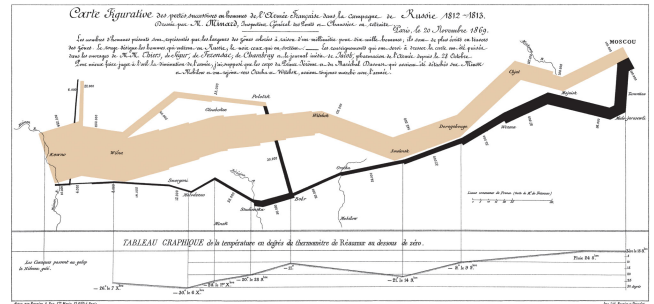


Figure 2: Charles Minard’s map of Napoleon’s Russian campaign of 1812 [15].

campaign. Tufte encourages the use and sharing of data-rich illustrations that represent all the available data, or as he calls them: “data rich displays”, as a means to always represent the truth the data is trying to convey.

2.3 Information visualization and infographics

Working professionals of the field have maintained a distinction between the two disciplines, affirming that infographics present information using statistical charts, maps, and diagrams, while information visualization offers visual “tools” that an audience can explore to analyze datasets. Where infographics tell stories designed by communicators, information visualization helps readers discover stories by themselves. In contrast, Cairo proposes that “Infographics and data visualization should exist in a continuum” [16], meaning that both disciplines should be approached in a similar way.

3. DATA VISUALIZATION AND STORYTELLING

The act of storytelling goes back in time as far as the first forms of human communication. It was well established long before humans learned how to write. It is a process that allows us to pass experiences from human to human through many media: oral tradition, printed word, visual images, radio, cinema, television.

DATA-DRIVEN JOURNALISM = A PROCESS

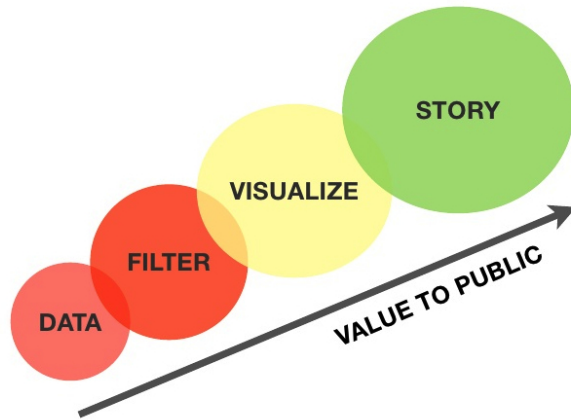


Figure 3: Data Journalism process [20].

The use of visualization to aid the storytelling process is widespread, typically by leveraging the use of diagrams and charts included in a larger body of text. In this structure, the image is used to illustrate the details explained in the story.

3.1 Narrative Structure

Storytelling strategies are very much dependent of the medium and the genre of the story. In this sense, stories presented through film and the moving image employ different mechanisms than stories told primarily through text. In a greater sense, a story is usually told in a sequential form: a beginning, a middle, and an end [17].

Grey et al., in the *Data Journalism Handbook*, pose the question: “What makes data journalism different to the rest of journalism? Perhaps it is the new possibilities that open up when you combine the traditional *nose for news* and ability to tell a compelling story, with the sheer scale and range of digital information now available” [18]. Data journalism can also help journalists tell complex stories through infographics. This explanatory journalism has helped audiences predict and anticipate answers to certain questions of social, economic, and political relevance, based on the stories that are being published by news media. These visualizations have traditionally been about basic and straightforward narratives based on the artifacts of storytelling.

Lorenz [19] developed a three-part workflow of data journalism, that includes data visualization as a critical step before the crafting of a news story is produced:

- Localization, retrieving, and integration of the data.
- Visualization of the data, where the output of the data can be observed and patterns are conveyed from it.
- Story composition from the analysis derived after the visualization.

Figure 3 identifies the stages of data journalism as a process that enhances the value the public receives from the news content.

3.2 Storytelling with Data

One of the first things to understand, is the distinction between “Data” and “Story”. Data, by itself, is an intrinsic form of information. This information has to be extracted by following often technical and rigorous procedures of cleaning, structuring, and sorting, so the data becomes understandable before visualization techniques can be applied to it, paving the way to the crafting of a coherent story. “Story is an interactive form of communication, where information is brought into a context that people can understand, remember, discuss and tell others about” [20]. In this sense, the two main reasons to use visualization as a form of presentation in news stories are: *exploratory*, to understand what is present in the data; *communicative*, to display and discuss visualizations with the public by telling stories based on the data and the facts.

In one of the earliest treatments of storytelling in information design and data visualization, Gershon and Page [21] describe a cinematic scene represented through visualization. This was an effort to identify that technology can provide new media and new genres to present information in a story-like construct, although it does not offer clear techniques or a strategy to understand the interactions between the two.

Drawing from news media data visualization to identify distinct genres of narrative visualization, the seminal work by Segel and Heer [22] offers a clear and concise study of narrative visualization and storytelling with data. They were able to develop a framework that suggests design strategies for narrative visualization in journalistic storytelling, in order to prevent certain difficulties presented when the user is given too much freedom and control over the narrative.

Segel and Heer placed narrative visualizations in a spectrum of *author-driven* and *reader-driven* approaches. In an author-driven approach, the structure of the visualization is strictly linear and includes no interactivity; its main advantage is efficient communication. The purely reader-driven approach is highly interactive and it doesn’t offer a clear path of the story; its best used when discovery tasks are to be performed to formulate hypothesis. Figure 4 shows the three structures or narrative strategies introduced by Segel and Heer. The defining distinction between the three has to do with how much author driven involvement is in the crafting of the path of the story:

1. *Martini Glass Structure*: it starts with an author-driven approach with the default view of the visualization, and once the intended narrative is complete the user is capable of exploring the data freely.
2. *Interactive Slideshow*: the visualization is presented as a regular slideshow where the user is able to interact with certain points of the presentation before allowing the story to advance to the next stage.
3. *Drill-Down Story*: it is based on a strict reader-driven approach where a general theme is presented and the user can interact with particular points of the visualization to reveal additional details and background information relevant to the main theme. However, it still requires a high level of authoring to determine where the interactions are going to take place.

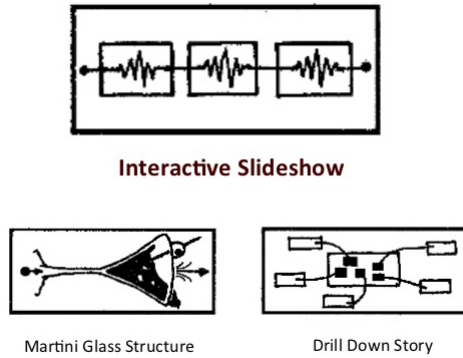


Figure 4: Segel and Heer’s narrative visualization structures [22].

3.3 User Interpretations

Hullman and Diakopoulos [23] take the research a step further by taking into account the rhetorical techniques used to influence user interpretation, an aspect often excluded from InfoVis research, functioning as a way to understand what is called “framing effects” [24] relevant to InfoVis, where small changes in visual findings can produce changes in opinion.

Their visualization rhetoric framework introduces the term *editorial layers* that include the data, visual representation, textual annotations, and interactivity, where different rhetorical techniques can be applied to the construction of narrative visualizations and convey meaning. The decision making process of a designer entails a series of conscious choices to add information (in the form of suggestions), or to omit information, thus affecting the end-user interpretations.

4. DISCUSSION AND FUTURE WORK

Without storytelling techniques “visualizations are not able to provide explanations about the subject and depend too much on the audience’s ability to interpret the data correctly” [25]. When a balance is achieved in the amount of interaction the designer provides to the user to discover the story by himself, some of the challenges of interpreting the visualization can be lessened. Considering this balance as a starting point to convey meaning in InfoVis strategies may be the greatest contribution made by Segel and Heer [22] in their framework for narrative visualizations.

An interesting consideration in the analysis of narrative visualization is the level of persuasion and influence in the end-user’s opinion of the overall story, when persuasive and rhetorical strategies are introduced in the visualization. The decision to employ these strategies, permeate the overall success or failure of the visualization and harness great power when used as a narrative technique in data journalism, when they leverage the efficacy of visual perception to communicate efficient ideas. Consequently, news media plays an important role in how topics of different natures are perceived by public opinion since it has increased in adoption by blogs [26] and mainstream news media alike.

However, it is important to take into account a few concerns and obstacles that might surface when too many nar-

rative elements are introduced in the visualization [27] specifically, whether or not they will distract the user from the data and the original message the author wishes to convey. Figueiras states that “although having a direction will help users that are less familiar with the subject, an undirected exploration can help proficient users find new interpretations of the data and even discover meanings that were not foreseen by the creators of the visualization. It’s important to understand every building block of the visualization in order to create a narrative that doesn’t overpower the data” [25].

We feel that in order to categorize visualization strategies as successful or unsuccessful in its intention to convey a message, an empirical evaluation should be conducted to determine if the user was “invested” in the data of the story, or “distracted” by interactive features or visual constructs that will consequently hinder the core message. There is a large amount of literature on narrative structuring techniques for visualization, but there are no clear guidelines that creators can use to find the best narrative strategies for different types of visualizations. Prior work has shown the benefits of having flexible narratives with landmarks and spaces for the user to freely explore [27], but achieving balance between user interaction and efficient communication always seems to be an issue to consider.

We have identified new directions of research pointed out by researchers in the area that can shed some light on how this topic should be addressed. Kosara and Mackinlay [28] propose that the focus should be given to the presentation of the elements of storytelling in the visualization, because of a lack of systematic studies of visualization techniques and their effectiveness as storytelling devices. Segel and Heer [22] propose that new research should be conducted by focusing exclusively on the reader’s experience and behavior when interacting with narrative visualizations. Similarly, the need to understand the mechanism of rhetorical devices and ways of communicating meaning was identified by Hullman and Diakopoulos [23]. A meaningful set of research questions proposed by Figueiras [29] articulate very succinctly new avenues of research in narrative visualization:

- what makes it work (level of knowledge communication);
- which of the narrative visualizations that are being produced in news media, advertising, research, education, etc. are having the desired effect on users;
- how and where should narrative elements be placed;
- how should the story be structured;
- what is the impact of these stories on the users.

What became evident in this preliminary research of the storytelling space in the field of information visualization is that it’s “starting to take shape” [25], and there are prevalent discussions that are taking place to articulate well grounded techniques and conventions to produce the desired communication with the audience.

5. ACKNOWLEDGMENTS

This work was partially supported by SIBILA Project - NORTE-07-0124-FEDER-000059, financed by the North Portugal Regional Operational Programme (ON.2 – O Novo

Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency, Fundação para a Ciência e a Tecnologia (FCT) and Labs. SAPO/U.Porto, financed by PT Comunicações.

6. REFERENCES

- [1] Tim O'reilly. What is web 2.0: Design patterns and business models for the next generation of software. *Communications & strategies*, 1(1):17, 2007.
- [2] 2014: The year in interactive storytelling, graphics and multimedia - nytimes.com.
<http://www.nytimes.com/interactive/2014/12/29/us/year-in-interactive-storytelling.html#data-driven-stories>.
- [3] Data | the guardian.
<http://www.theguardian.com/data>.
- [4] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999.
- [5] Benjamin Jotham Fry. Computational Information Design. *Sciences New York*, PhD(May 1997), 2004.
- [6] Colin Ware. *Information Visualization: Perception for Design*. Elsevier, 2013.
- [7] Ji Soo Yi, Youn-ah Kang, John T Stasko, and Julie a Jacko. Understanding and characterizing insights: How Do People Gain Insights Using Information Visualization? *Proceedings of the 2008 conference on BEyond time and errors novel evaluation methods for Information Visualization - BELIV '08*, page 1, 2008.
- [8] Graham Dove and Sara Jones. Narrative visualization: sharing insights into complex data. *Interfaces and Human Computer Interaction (IHCI 2012)*, pages 21–23, 2012.
- [9] Neil Halloran. The fallen of world war ii - data-driven documentary about war and peace.
<http://www.fallen.io/ww2/>.
- [10] Antony Chen, Chun-houh; Härdle, Wolfgang; Unwin. *Handbook of Data Visualization*. Springer Handbooks Comp.Statistics. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [11] Hans Rosling. Gapminder, 2009.
- [12] I. Bernard Cohen. Florence Nightingale. *Scientific American*, 250(3):128–137, 1984.
- [13] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 2nd edition, 1983.
- [14] Edward R Tufte and David Robins. *Visual explanations*. Graphics Cheshire, CT, 1997.
- [15] John Corbett. Charles josephminard: Mapping napoleon's march, 1861, 2012.
- [16] Alberto Cairo. *The Functional Art*. New Riders, San Francisco, CA, 1st edition, 2012.
- [17] Norma J Livo and Sandra A Rietz. *Storytelling: Process and practice*. Littleton, Colo.: Libraries Unlimited, 1986.
- [18] Jonathan Gray, Lucy Chambers, and Liliana Bounegry. *The Data Journalism Handbook*. O'Reilly Media, 1st edition, 2012.
- [19] Mirko Lorenz. Data-driven journalism: What is there to learn? In *IJ7-Innovation Journalism Conference*, Stanford, CA, 2010.
- [20] Data-driven journalism : What is there to learn? *European Journalism Centre*, 2010.
- [21] Nahum Gershon and Ward Page. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31–37, August 2001.
- [22] Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(March):1139–1148, 2010.
- [23] Jessica Hullman and Nick Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17, 2011.
- [24] James N. Druckman. On the limits of framing effects: Who can frame? *The Journal of Politics*, 63(04), 2008.
- [25] Ana Figueiras. How to tell stories using visualization. In *Information Visualisation (IV), 2014 18th International Conference on*, pages 18–18. IEEE, 2014.
- [26] Bruce Etling, Hal Roberts, and Robert Faris. Blogs as an alternative public sphere: The role of blogs, mainstream media, and tv in russia's media ecology. *SSRN Electronic Journal*, 2014.
- [27] Nicholas Diakopoulos. Game-y information graphics. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 3595–3600. ACM, 2010.
- [28] Robert Kosara and Jock Mackinlay. Storytelling: The next step for visualization. *Computer*, pages 44–50, 2013.
- [29] Ana Figueiras. Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations. In *Information Visualisation (IV), 2014 18th International Conference on*, pages 46–52. IEEE, 2014.
- [30] Seref Sagiroglu and Duygu Sinanc. Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pages 42–47. IEEE, May 2013.