Promoting insight: A Case Study of How to Incorporate Interaction in Existing Data Visualizations

Ilo Alexandre
Faculdade de Ciências Sociais e Humanas
Universidade Nova de Lisboa
Lisbon, Portugal
mail@iloaguiar.com

***Abstract*—Visualizations became essential to understand large datasets and abstract information. Data Visualizations are increasingly used to tell compelling stories, particularly in journalism. In order to engage the users with complex data, interaction techniques are incorporated into storytelling, emerging narrative visualization. Narrative visualization attempts not only to present information with context, but also to allow users to explore the dataset, find patterns or structures in the data, providing the users with control over the insights she/he gains from interaction. Interaction techniques play a key role in narrative visualization improving understanding of the data and stimulating insights through discovery. In this paper, we explore the benefits of adding interaction techniques to data visualization. Drawing on case studies from Portuguese news media, we identified strategies to enhance storytelling in visualizations and to provide better insights through different types of interactivity.**

Data visualization; interaction; insight; storytelling; case study

I. Introduction

Over the last decade the interest in data visualization increased in different areas, such as marketing, business management and education. Using visualization techniques to understand large datasets and abstract information became essential nowadays. The reason is that visualization of information amplifies cognition and makes data that is not naturally accessible to the bare eye [6, 7] visible and understandable.

The large amount of data that is publicly available (e.g., Open Government Data [12], WikiLeaks [4]) combined with sophisticated, easy-to-use tools and applications made the phenomenon of data visualization a trend in journalism, leading to the emergence of a new term: data journalism [14]. The New York Times[1] and The Guardian[2] were among the pioneers [25] and they remain a reference in the subject. Although there is still no consensus on its definition [11, 20], the data journalism process can be described in a few steps: collect the data, check and clean it, contextualize and combine it with others datasets, choose a narrative visualization and add storytelling and interactive techniques [5, 19].

Data visualizations are critical in journalism not only to tell complex stories, but also strengthen the role of accountability of journalism [2, 13]. The practices to create data visualizations are being researched all over the world [3, 11, 16, 20, 22] but evaluation of these visualizations is still scarce [18]. Assessments are important to measure the efficiency of visualizations.

Data visualizations should not only present information, but also allow users to explore that information [6]. Therefore, interactive visualizations can help users to gain insight [9] and make decisions [22]. *The Fallen of World War II*[3] and *Is It Better To Rent or Buy?*[4], respectively, illustrates these capability. The first is an interactive documentary that examines the World War II casualties and allows users to pause during key moments to interact with the charts and dig deeper into the numbers. The second is a tool that accompanies an article about housing prices. The tool allows users to input their own information and calculate at the price of the house, mortgage rates, and down payment to help she/he determine if it is better to rent or buy.

In this paper we examine narrative strategies [26], processes to gain insight [28] and interaction techniques [10] used by online news media to tell stories and engage the audience. We collected examples from previous interviews with Portuguese newspapers editors and used them as case studies. Taking an empirical approach, we analyze four visualizations, their narrative strategies, interaction techniques, procedures leading to the insight and how they could be improved with another interactivity techniques. Finally, we discuss some of the benefits and challenges of introducing interaction techniques in data visualizations, as well as research opportunities in this field.

II. Background

According to Card, Mackinlay and Shneiderman, "The purpose of visualization is insight, not pictures" [7]. Building on work by North [21], Plaisant, Fekete and Grinstein they define insight as "a nontrivial discovery about the data or as a complex, deep, qualitative, unexpected, and relevant assertion" [23]. Based on the model of sensemaking

[1] http://www.nytimes.com/interactive/2015/us/year-in-interactive-storytelling.html
[2] http://www.theguardian.com/data
[3] http://www.fallen.io/ww2/
[4] http://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html

2375-0138/16 $31.00 © 2016 IEEE
DOI 10.1109/IV.2016.15

proposed by Pirolli and Card [24], Yi et al. [28] introduce four processes through which people gain insight while using an Information Visualization (InfoVis) system: provide overview, adjust, detect pattern, and match mental model.

In the *Overview*, the user looks at the "big picture" and examines the data available for future explorations. "Even though observing an overview may not directly help a person gain insight, it appears to play an important role by helping people make sense of and find which areas they need to investigate more, thereby promoting further exploration of the dataset" [28]. During the *Adjust* step, filtering helps to explore the data at different levels of abstraction and enables the selection of specific information. The *Detect Pattern* step enables the user to find what she/he has been looking for and discover new knowledge that they did not expect to find by accessing specific distributions, trends or structures in the data.

Finally, the *Match Mental Model* transforms information into a visualization, a physical space for effective exploration, making sense of the data. "A visual representation of data can decrease the gap between the data and user's mental model of it, thereby reducing cognitive load in understanding, amplifying human recognition of familiar presences, and linking the presented visual information with real-world knowledge" [28].

Introducing storytelling in visualizations is another way to help the user to gain insights, introducing context and meaning in the visualization. Segel and Heer [26] called narrative visualization this combination of storytelling techniques and interactive data visualizations. Analyzing the elements used by news media to tell visual data stories Segel and Heer identified some patterns and introduced three structures or narrative strategies: Martini Glass Structure, Interactive Slideshow, and Drill-Down Story. The first structure starts with an author-driven approach with the default view of the visualization and only once the author's intended narrative is complete, the visualization opens up to a reader-driven stage where the user is free to explore the data interactively. In the Interactive Slideshow the visualization is presented as a typical slideshow with some interactivity in each slide. Drill-Down Story is based on a strict reader-driven approach, where the user can choose any order possible to interact with the story.

Building on Segel and Heer's design space analysis, Hullman and Diakopoulos define narrative information visualizations as "a style of visualization that often explores the interplay between aspects of both explorative and communicative visualization" [15]. According to Hullman and Diakopoulos, narrative visualizations depend on a combination of persuasive, rhetorical techniques to transmit an intended story to users as well as exploratory, "dialectic strategies aimed at providing the user with control over the insights she gains from interaction" [15].

Interactivity techniques play a key role in narrative visualization stimulating insights through discovery. Interactive data visualization offers an opportunity for the audience to experience a similar moment of insight to that which the author had experienced earlier [10]. One of the first and arguably the most well-known interaction techniques taxonomy is *The Visual Information Seeking Mantra by Shneiderman* [27]. Since then there have been several efforts to produce visualization taxonomies [1, 8, 17, 29]. Taxonomies help to achieve a better understanding of the design space of interaction, but as appointed by Yi et al. [29] defining a comprehensive taxonomy is challenging because InfoVis is still a growing field and it is highly possible that taxonomies get antiquated. Therefore, we apply in our case studies the interaction techniques categories proposed by Figueiras [10]: filtering, selecting, abstract/elaborate, overview and explore, connect/relate, history, extraction of features, reconfigure, encode, participation/collaboration, and gamification.

Filtering is one of the most common techniques to reduce complexity, allowing the user to see how the data representation is affected when parts that are uninteresting to him/her are eliminated or deemphasized. The original data can be recovered whenever the users wishes by resetting the criteria.

Selecting techniques provide users with the ability to mark items of interest to keep track of it. These also evidentiate the data of interest and enables the users to see it in contrast with other items.

Abstract/Elaborate interactivity techniques provide the users with the capability to adjust the level of abstraction of a data representation. Zooming, details-on-demand and linking are some abstract/elaborate techniques. Zooming allows users to see an overview of the visualization (zoom-out) or to see a closer, more detailed view (zoom-in) without altering the representation. Details-on-demand gives the users extra information upon the selection of an item, providing "backstories that not only help in the level of engagement of the user but also provide relevant details" [10]. Drill-down options, tool-tips and pop-ups by mouse-hover or click are some examples of details-on-demand interactions. Linking techniques give access to external information, as do hyperlinks to the reference data or to give access to a different visualization method.

Overview gives a general context of the dataset while exploring interactions enable users to examine a different subset of data. Both techniques help the user to better understand the data, allowing patterns to be more easily identified and insights gained.

Connect/relate interactions enables the users to see the relationships between the data items by highlighting links between the items that are already represented in the visualization or even by showing items that are relevant to an item that the user has interest in and that were previously hidden [10].

Reconfigure interactive techniques allow the users to change the way data items are arranged. The new spatial arrangement of the representation can reveal hidden characteristics of data and provide different perspectives on the dataset. Encode techniques enable the users to change the fundamental visual representation of the data (e.g., color, size, and shape) facilitating the discovery of new insights about the data.

History provides ways for the users to undo and replay his/her actions. This technique allows users to recover from mistakes in the data exploration and to progressively refine the exploration.

Extraction of features enables to extract data in which the users are interested. Exploring the data often becomes a lengthy and complex task, therefore allowing the users to extract the data so it can be shared, dissected, or even seen in other visual representations, can reduce that complexity and result in better insights [10].

Participation and collaboration techniques allow the users to contribute to the data visualization interpretation and understanding, sharing their insights. Finally, gamification interaction techniques show the data in a more playful way to the users. Gamified visualizations can conduct the user's attention to a particular subset of the data. This type of interaction is the least common because its production is time consuming and very complex. Although gamification can include most of the other interaction techniques, what makes it different is the inclusion of video game elements, such as ranks, reputations, time constraints, levels, goals, etc.

III. Case Study

We asked four Portuguese newspaper editors to choose a data journalism example in their respective newspapers. The case studies are from the online news media Jornal de Notícias[5], Público[6], Diário de Notícias[7] and Correio da Manhã[8]. We explore the processes through which people gain insight [28], the narrative strategies [26] and interactivity techniques [10] used in each example.

A. *Forest fire in Portugal*[9]

The visualization *Forest fire in Portugal* (*Os fogos e as vítimas deste ano em Portugal*), shown in Fig. 1, displays an interactive map showing the numbers of forest fires and its victims in Portugal. The main map shows the numbers of forest fires and burnt area in each district of the country. Complementary graphics with the month comparison and the year comparison of the forest fires and the burnt area are provided. It has also a timeline with information about the victims and what happened. Since the user can choose if he/she wants to start exploring the map or the graphics, this visualization is considered to have a Drill-Down Story structure, according to Segel and Heer's categories for structure [26].

In terms of interactive elements, *Forest fire in Portugal* has at least two techniques categories: Overview and Explore and Abstract/Elaborate. The map of Portugal with the fire flame pictograms of different sizes representing the burnt area provides an overview that gives the user a general context to understand the dataset [10]. It is possible to get details of each region by mouse-hover the pictograms. This is an example of Details-on-demand, a type of Abstract/elaborate interaction. It is also possible to obtain more information about the victims by details-on-demand. This visualization presents three procedures leading to the insight: Provide Overview, Adjust and Match Mental Model.

Figure 1. Forest fire in Portugal.

Forest fire in Portugal would probably be enhanced if it were possible to relate two or more regions at once because to compare the forest fires and the burnt area of two different regions the user have to memorize the data from one and click on the other. The user should also be allowed to reconfigure the data and compare the forest fires or the burnt area. The timeline with the victims should display graphics to show trends of date of death, age and location and enable the user to filter by these criteria. Finally, the visualization should allow users to collaborate reporting fires in their regions. Those interactivity techniques may give more insights about the data to the user.

B. *Evolution of marital status in Portugal*[10]

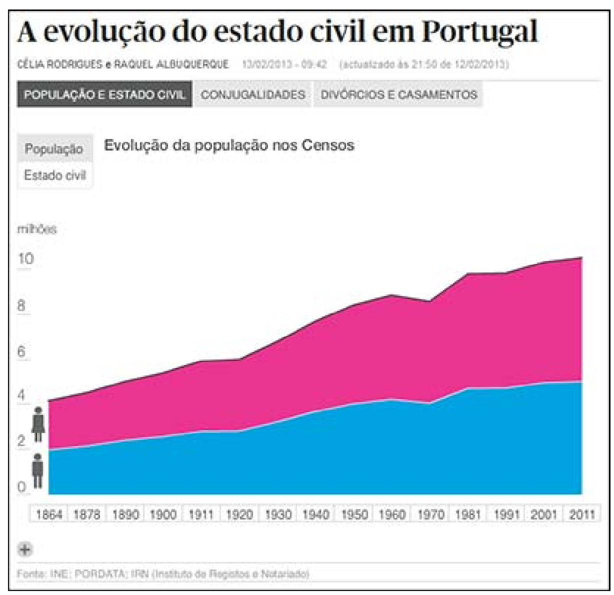

Figure 2. Evolution of marital status in Portugal.

[5] http://www.jn.pt/paginainicial/
[6] http://www.publico.pt/
[7] http://www.dn.pt/
[8] http://www.cmjornal.xl.pt/
[9] http://www.jn.pt/multimedia/infografia.aspx?content_id=3402610
[10] https://www.publico.pt/multimedia/infografia/a-evolucao-do-estado-civil-em-portugal-46

The visualization *Evolution of marital status in Portugal* (*A evolução do estado civil em Portugal*), shown in Fig. 2, collected the numbers of marital status of all censuses conducted in Portugal since 1864 and organized the statistics in three different tabs. The first one has two graphics: one with the population growth over the years and the other with the evolution of six types of marital status throughout the years. In the second tab, there are three static slides with graphics about types of union, number of single and divorced. The third tab has comparative maps with numbers of divorces in each municipalities of Portugal in 1984 and 2011. This visualization has an order imposed by the author and is classified as Slideshow in Segel and Heer's categories [26].

Evolution of marital status in Portugal allows the user to interact and further explore particular points of the visualization. The first slide enables the user to filter the population data by decade and by gender. It is also possible to reconfigure the data of the evolution of the marital status and get extra details of evolution of divorcees and widowers by details-on-demand interaction. This interactivity technique is also used in the last slide to provide the user information about the municipalities with more marriages and more divorces. This visualization displays elements from all four processes to gain insight: Provide Overview, Adjust, Detect Pattern Match Mental Model.

Evolution of marital status in Portugal's last slide could be improved with techniques such as zooming to see details of the maps, since it is difficult to visualize some municipalities and encoding to change the colors of the representation. Another way to enhance the insights in the visualization is creating filters for the user to be able to choose a marital status to see in the map and then select specific municipalities to compare the data (Relate technique). Lastly, since the author mentions the data sources, it should have hyperlinks allowing the user to further explore the data.

C. *For every less younger person there is one more elderly*[11]

Figure 3. Evolution of marital status in Portugal.

Accompanying the article *For every less younger person there is one more elderly* (*Por cada jovem a menos há um idoso a mais*), the Diário de Notícias published an interactive visualization about the aging process of the Portuguese population in the beginning of 21st century, shown in Fig. 3. *For every less younger person there is one more elderly* is a typical example of Slideshow visualization, with a more author-driven approach than reader-driven approach.

Slideshows work well with complex dataset and narratives [26], however *For every less younger person there is one more elderly* is far from complex. From the five slides of the presentation two incorporates interactions and only Details-on-demand technique is employed in the visualization. There is also only one process leading to the insight: Match Mental Model.

For every less younger person there is one more elderly contains some static graphics that could be interactive. The ten most populous cities, for example, could have a filter technique to select only the data that the user is interested. Enabling the user to relate these cities with other cities around the world would also be interesting. The same could happen to other graphics, since there is a lack of context data in the visualization. Another improvement would be to show the cities according to number of older and/or younger people allowing the user to filter and select by city and age. Finally, the sentences of politicians should have hyperlinks to the full interview.

D. *Map of homicides in Portugal*[12]

Map of homicides in Portugal (*Mapa dos homicídios em Portugal*) compiled all homicides in Portugal. The visualization, shown in Fig. 4, uses Google Maps[13] to locate the crimes and on the bottom there is the circumstances in which the victims lost their lives. The visualization does not impose a path to follow, leading the user to explore the map and the details about the victims freely. Therefore, *Map of homicides in Portugal* has a Drill-Down Story structure and a reader-driven approach, once it "presents a general theme and then allows the user to choose among particular instances of that theme to reveal additional details and backstories" [26].

The overview map shows in which regions crimes are concentrated. Each homicide is represented by a pictogram of a cross. Extra information about the victims pops up by mouse-hover, a details-on-demand interactivity technique. It is also possible to zoom-in and see exactly where the murders happened. Details about the homicides are provided along the bottom: year of the crime, location, victim's name, victim's age, weapon used and relationship between the victim and the perpetrator. It is possible to reconfigure and filter the data on each criteria, highlighting what is interesting to the user. This visualization displays elements from all four processes to gain insight: Provides Overview, Adjust, Detect Pattern and Match Mental Model.

[11] http://www.dn.pt/portugal/interior/por-cada-jovem-a-menos-ha-um-idoso-a-mais-3261494.html

[12] http://www.cmjornal.xl.pt/mortes_violentas.html

[13] http://maps.google.com/

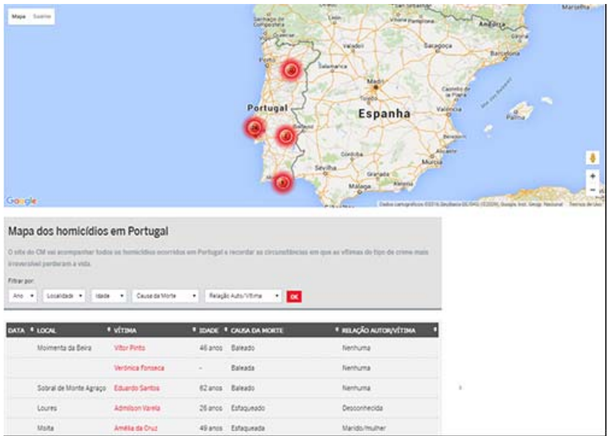

Figure 4. Map of homicides in Portugal.

Map of homicides in Portugal would probably be enhanced if a hyperlink to a story about the crime was added to each murder. To enable change from the used map to a heat map may improve insights about the data because "providing encoding techniques that allow the user to fundamentally change the visual representation can facilitate the discovery of new insights about the data." [10]. Finally, two less common interactive techniques: collaboration and extraction. The first could be achieved enabling the user to share a specific state of the visualization and the second allowing the user to extract the data so it can be dissected and reused in other visualizations.

IV. CONCLUSIONS AND FUTURE WORK

Storytelling and interactive visualizations are a trend in journalism, but there is a need to consider how those elements can engage the audience and help them gain more insights. In this paper, we conducted an analysis of narrative strategies [26], processes to gain insight [28] and interactivity techniques [10] on four online news articles. The research has two main outcomes. First, it shows that some Portuguese newspapers are doing data visualizations, since all case studies have a narrative structure and they adopted at least one interaction technique. Second, it shows that all the case studies could be improved and how that could be done. We explore the benefits of adding interaction techniques to visualizations to enhance the storytelling and to provide better insights.

In InfoVis there still is an absence of studies of visualization techniques and their effectiveness as storytelling devices [18]. This paper aims to be a contribution to help filling the gap, evaluating and understanding how Portuguese newspapers utilize narrative elements and which interactive elements they use to reach the audience. However, more evaluation studies focusing in data journalism is needed to identify patterns in the narrative structure and interactivity techniques.

It is also important to conduct research on readers' experience. Our analysis reveal that Details-on-demand technique was the only one used in all case studies but how is the impact of this interaction on the audience? Which are their favorite interaction technique? Do they prefer author-driven approach or reader-driven approach?

A short answer for those questions does not seem to exist since users have different levels of proficiency and, while some readers will prefer a more complex narrative, others will feel intimidated by sophisticated graphics. Nevertheless we can try to achieve broad guidelines to help to create more comprehensible, interesting and memorable data visualizations to the general public.

REFERENCES

[1] R. Amar, J. Eagan and J. Stasko, "Low-level components of analytic activity in information visualization," In Information Visualization 2005. INFOVIS, 2005. IEEE Symposium on, pp. 111–117. October 2005.

[2] C. W. Anderson, E. Bell and C. Shirky, "Post-industrial journalism: adapting to the present," Tow Center for Digital Journalism, Columbia University. 2012.

[3] E. Appelgren and G. Nygren, "Data journalism in Sweden," In Digital Journalism 2 (3) 2014, pp. 394-405. November 2014.

[4] S. Baack, "A new style of news reporting: Wikileaks and data-driven journalism," In Cyborg Subjects: Discourses on Digital Culture, pp. 113–122. 2011.

[5] P. Bradshaw, "The inverted pyramid of data journalism," Online Journalism Blog. 2011. Retrieved from http://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/.

[6] A. Cairo, The Functional Art: An introduction to information graphics and visualization. 1st ed., Berkeley, CA: New Riders. 2012.

[7] S. K. Card, J. D. Mackinlay and B. Shneiderman, Readings in information visualization: using vision to think. San Francisco, CA: Morgan Kaufmann. 1999.

[8] A. Dix and G. Ellis, "Starting simple: adding value to static visualisation through simple interaction," In Proceedings of the Working Conference on Advanced Visual Interfaces. AVI '98, pp. 124–134. 1998.

[9] G. Dove and S. Jones, "Narrative visualization: sharing insights into complex data," In Interfaces and Human Computer Interaction 2012. IHCI. July 2012.

[10] A. Figueiras, "Towards the understanding of interaction in information visualization," In Information Visualisation (iV) 2015. 19th International Conference on, pp. 140-147. July 2015.

[11] K. Fink and C. W. Anderson, "Data journalism in the United States: beyond the 'usual suspects'," In Journalism Studies 16 (4) 2015, pp. 467-481. August 2014.

[12] J. Gray, "Towards a genealogy of open data," In General Conference of the European Consortium for Political Research in Glasgow. September 2014.

[13] J. T. Hamilton and F. Turner, "Accountability through algorithm: developing the field of computational journalism," Report from the Center for Advanced Study in the Behavioral Sciences. July 2009.

[14] A. Howard, "The art and science of data-driven journalism," Tow Center for Digital Journalism, Columbia University. 2014.

[15] J. Hullman and N. Diakopoulos, "Visualization rhetoric: framing effects in narrative visualization," In Visualization and Computer Graphics, IEEE Transactions on, 17 (12), pp. 2231-2240. December 2011.

[16] J. Karlsen and E. Stavelin, "Computational journalism in Norwegian newsrooms," In Journalism Practice 8 (1) 2014, pp. 34–48. July 2013.

[17] D. A. Keim, "Information visualization and visual data mining," In Visualization and Computer Graphics, IEEE Transactions on, 8 (1), pp. 1-8. Jan-Mar 2002.

[18] R. Kosara and J. Mackinlay, "Storytelling: the next step for visualization," In Computer 46 (5), pp. 44-50. May 2013.

[19] M. Lorenz, "Data-driven journalism: what is there to learn?" In IJ7-Innovation Journalism Conference, 7-9 June 2010.

[20] J. De Maeyer, M. Libert, D. Domingo, F. Heinderyckx and F. Le Cam, "Waiting for data journalism: a qualitative assessment of the anecdotal take-up of data journalism in french-speaking Belgium," In Digital Journalism 3 (3) 2015, pp. 432-446. November 2014.

[21] C. North, "Toward measuring visualization insight," In Computer Graphics and Applications 2006, IEEE 26 (3), 6-9. May-June 2006.

[22] S. Parasie and E. Dagiral, "Data-driven journalism and the public good: 'computer-assisted-reporters' and 'programmer-journalists' in Chicago," In New Media Society 15 (6) 2012, pp. 853-871. September 2013.

[23] C. Plaisant, J. D. Fekete and G. Grinstein, "Promoting insight-based evaluation of visualizations: from contest to benchmark repository," In Visualization and Computer Graphics 2008. IEEE Transactions on, 14 (1), pp. 120-134. Jan-Feb 2008.

[24] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," In Proceedings of International Conference on Intelligence. 2-4 May 2005.

[25] S. Rogers, Facts are Sacred. Main edition. London: Guardian Faber Publishing. 2013.

[26] E. Segel and J. Heer, "Narrative visualization: telling stories with data," In Visualization and Computer Graphics 2010. IEEE Transactions on, 16 (6), pp. 1139–1148, Nov-Dec 2010.

[27] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," In Visual Languages 1996. Proceedings, IEEE Symposium on, pp, 336-343. September 1996.

[28] J. S. Yi, Y. Kang, J. T. Stasko and J. A. Jacko, "Understanding and characterizing insights: how do people gain insights using information visualization?" In Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization (BELIV '08). 2008.

[29] J. S. Yi, Y. Kang, J. T. Stasko and J. A. Jacko, "Toward a deeper understanding of the role of interaction," in Information Visualization. In Visualization and Computer Graphics. IEEE Transactions on, 13 (6), pp. 1224-1231. Nov-Dec 2007.