Qualitative Studies on Human-data Interaction and Data Analysis Collaboration in Latin America

Caroline Queiroz Santos
UFVJM
Diamantina - MG, Brazil
caroline.queiroz@ufvjm.edu.br

Milene Selbach Silveira
PUCRS
Porto Alegre - RS, Brazil
milene.silveira@pucrs.br

ABSTRACT

Professionals from different areas deal with data in various formats and do not always know how to extract information and knowledge from those datasets. Our studies focused on getting the user (data analyst) to reflect and project interactive narrative visualizations based on data. We used a model to help users to build knowledge about computing resources and techniques they may appropriate and employ in a data analysis context. Our results indicate pertinent ways of representing relevant narrative aspects in data visualizations. The purpose of this work is to investigate people's needs in relation to data analysis, and define strategies and solutions to bridge the gap between data-analysis professionals and computational solutions. Additionally, given the importance of culture in HDI research, we believe we need to direct our efforts to develop human-data interaction in LatAm context.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Visualization design and evaluation methods**.

KEYWORDS

data visualization, data analysis, human-data interaction, design

INTRODUCTION

Advances in information and communication technologies, as well as in large data volumes access, have often required knowledge and skills to collect, process, present, and analyze these data. Professionals from different areas deal with large sets of data in several formats and do not always know how to extract information and knowledge from them. We conducted studies with students and professionals who deal with this problem, collecting their opinions related to data visualizations' aspects [5, 6].

Later, we conducted studies with savvy professionals, who work or study large volumes of data and are acquainted to data collection and visualization tools. Thereafter, we verified how professionals who claim to be ignorant of these types of tools deal with the data they collect and analyze.

These studies focused on getting the user (data analyst) to reflect and project their data interactive narrative visualization [7]. We used a model that introduced these professionals to a new way of planning and presenting their data, as well as helped them to build knowledge about computing techniques that they may appropriate and employ in a data analysis context.

The obtained results indicate pertinent ways of representing relevant narrative aspects in data visualizations. We have also verified that the model is satisfactorily descriptive to assist in the design of visualizations, whether narrative or not.

In this scenario, we propose to investigate people's needs concerning to data analysis in order to define strategies and seek solutions to bridge the gap between data-analysis professionals and computational solutions designed for them. In other words, for this LatAm workshop, we intend to collaboratively investigate and apply concepts from the emerging field of Human-Data Interaction (HDI) [1–3, 8], doing this by placing humans in the center of data flow and providing mechanisms for citizens to gain insights and awareness about HDI context, coming to explicitly interact with the data.

CONTEXT AND GOALS

Our main study in this field concerns about how to help data analysts in the efficient presentation of large volumes of data, providing techniques that support the design processes, analysis and understanding of these data.

We created a customizable interface design model to support the construction of interactive narrative visualizations of data extracted from social media [4]. The model is descriptive, having

fundamentals on the Semiotic Engineering theory and its view on End-user Development, as well as on narrative visualization concepts. It considers that the data analyst is interested in organizing their data to be presented in such a way to tell a story. Thus, the model offer a set of components to structure the design space, coming to assist designers and data analysts (co-designers) to reflect on different aspects that influence on the construction of customizable interactive narrative visualizations.

We carried out four studies and, based on the results triangulation, we obtained the bases for refining and consolidating the model idea. After conceiving the model, we conducted a study focusing on its use by social-media data analysts, whose profiles are presented in Table 1.

Subsequently, in the context of tourism data in Brazil (whose profiles are presented in the Table 2.), from a specific scenario, we noticed that those professionals don't use yet techniques and tools for processing and presenting data, which may be compromising the extraction of relevant information and knowledge from data they collect. The model introduced these professionals to a new way of planning and presenting their data, as well as helped them to build knowledge about computing names and techniques which they may appropriate and employ in a data analysis context.

Awareness is an important element in this area and there is a need to make users aware of the data embedded in IHD solutions [8]. We verified the potential epistemic value of our model, considering that its use can instigate important discussions and reflections during the process of designing customizable interactive narrative visualizations, assisting both designers and co-designers in its epistemic character.

What is the key LatAm research focus/problem domain you aim to address?

We intend to investigate and identify ways for the LatAm end-user to be aware of what these data are, as well as how they are collected and stored (by social media, government, business, and people) so that they can make decisions about what to do with that data, how to process them and how to analyze them in the best way, given cultural aspects and contexts within Latin America.

How does your research currently addresses this problem domain?

Our research tries to comprehend this problem by investigating the user needs and also how much knowledge and perception they have about everything that involves the availability, collection, treatment, and presentation of the data with which they deal.

How does pursuing a regional research agenda can help advance the state of the art?

IHD is an interdisciplinary research topic involving elements of computer science and areas such as communication, law, sociology, psychology, human behavior, economics, and others. The research in this field is recent and have the purpose of analyzing the impacts of data collection, filtering, processing and analysis (from several sources) in the actual society. In this sense, a regional research

Table 1: Profile of participants.

ID	Age	Education	Do you use visualization tools?
I1	58	History	No
I2	42	History	No
I3	28	Journalism	No
I4	37	Journalism	Yes
I5	32	Publicity	Yes
I6	25	Systems Analysis	Yes
I7	28	Business Admin.	Yes
I8	29	Journalism	Yes

Table 2: Profile of participants.

ID	Age	Genre	Education	Role
P1	30	M	Turismologist	Decision maker
P2	33	F	Turismologist	Decision maker
P3	41	F	Turismologist	Data modeler
P4	37	F	Economist	Data modeler

agenda could help researchers to understand the user's needs within their cultural context and propose ways to meet those needs in a better way.

What actionable mechanisms can be designed to realize the proposed agenda?

Narrative visualization is an emerging data visualization field and has been used for the purpose of communicating information efficiently and intuitively as well as providing greater engagement in data exploration. We understand that to discuss fundamental concepts for interactive narrative visualizations design and explore these concepts – in theory and practice – can aim to enable reflection on decisions to be taken in narrative visualization's design and, consequently, in the HDI.

PROPOSAL FOR COLLABORATIVE RESEARCH METHODS

Collaborative research methods that could apply in this field, considering researchers in different sites are: ethnography, interviews, triangulation of data, identifying the needs of end-users, creation of models from data, reflections about tools design, and tools evaluations.

REFERENCES

[1] Afra Mashhadi, Fahim Kawsar, and Utku Günay Acer. 2014. Human Data Interaction in IoT: The ownership aspect. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*. 159–162.

[2] Pedro E.L.R. Meireles and Heiko Hornung. 2016. Human Values in Human-Data Interaction. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems (IHC '16)*. ACM, New York, NY, USA, Article 51, 4 pages.

[3] Richard Mortier, Hamed Haddadi, Tristan Henderson, Derek McAuley, and Jon Crowcroft. 2013. Challenges & opportunities in human-data interaction. *University of Cambridge, Computer Laboratory* (2013).

[4] Caroline Queiroz Santos. 2018. *CIViS: Modelo de design de interface customizável para apoiar a construção de visualizações narrativas interativas de dados extraídos de mídias sociais.* Tese de Doutorado. Escola Politécnica, PUCRS, Porto Alegre, Brasil.

[5] Caroline. Q. Santos, Holisson Cunha, Carlos Teixeira, Daniele Ramos de Souza, Roberto Tietzmann, Isabel Manssour, Milene Silveira, Marcelo Träsel, Duncan Dubugras Alcoba Ruiz, and Rodrigo Barros. 2017. Media professionals' opinions about interactive visualizations of political polarization during Brazilian presidential campaigns on Twitter. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*. 1939–1948.

[6] Caroline. Q. Santos, Roberto Tietzmann, Marcelo Träsel, Sílvia M. W. Moraes, Isabel H. Manssour, and Milene S. Silveira. 2016. Can visualization techniques help journalists to deepen analysis of Twitter data? Exploring the "Germany 7 x 1 Brazil" case. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*. 1939–1948.

[7] Edward. Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148.

[8] Mateus Rambo Strey, Roberto Pereira, and Luciana C. de Castro Salgado. 2018. Human Data-Interaction: A Systematic Mapping. In *Proceedings of the 17th Brazilian Symposium on Human Factors in Computing Systems (IHC 2018)*. ACM, New York, NY, USA, Article 27, 12 pages. https://doi.org/10.1145/3274192.3274219