User Experience for Non-Expert Audiences in Data Exploration

Tylosky Natasha, Knutas Antti, Qureshi Majad, Wolff Annika

DOI: 10.1145/3593743.359375

User Experience for Non-Expert Audiences in Data Exploration

Natasha Tylosky
Finland LUT University

Antti Knutas
Finland LUT University

Majad Qureshi
Finland LUT University

Annika Wolff
Finland LUT University

ABSTRACT

Data Exploration remains an untapped area of research, and as of yet few guidelines exist for how to design a data exploration tool. In this paper we conduct a qualitative study of a data exploration tool that has been created using a series of principles for designing data exploration systems for non-expert audiences. In the course of this study, we both evaluate the tool, dubbed the Data Explorer, and refine the design principles that it was based on.

KEYWORDS

Data Exploration, Design Principles, Interactive Data Visualization, Sensemaking, User Experience

1 INTRODUCTION

Data exploration can be described as the facilitation of knowledge extraction and information perception [2]. Visual data exploration, the subset of exploration that this paper will focus on, provides users with a visual means to interact with data and provide their own input. Data exploration allows users to examine different subsets or categories of data to gain understanding and insight [25]. The goal of data exploration for non-expert users is to engender sensemaking in said users, by providing a more intuitive means of interacting with data, as traditional data visualization techniques may have a high barrier of entry to understand [8].

The user experience of data exploration systems, specifically ones aimed at a non-expert audience (i.e. audiences who have no domain knowledge of the topic of a particular data set or data sets), remains an under-researched area in the field of Human Computer Interaction, as few such systems or tools have been built and evaluated. This study will evaluate the user experience of a data exploration software tool, called The Data Explorer, with the aim of both determining what improvements can be made to this tool, and refining the design principles the tool was built on.

The system that we are evaluating was built based upon the design principles first laid out in the paper [19], these principles are meant to guide designers in creating data exploration systems for non-expert users. The principles in the aforementioned paper were defined via a systematic mapping study but remain unevaluated using real-world participants. Through evaluating the user experience of The Data Explorer, we will refine the aforementioned principles with the aim of making them generalizable to other data exploration tools and systems for non-expert audiences. We will discuss these principles in details in the background section.

2 BACKGROUND

2.1 Related Work

Interest in Human Data Interaction (HDI) has been increasing over the past decade or so [22], as has the interest in data exploration as a subsection of HDI and HCI research [3, 7, 26]. And while in recent years some guidelines exist for designing data visualization systems in HDI [20, 21], there remains a gap when designing for data exploration specifically for non-experts. Data exploration systems differ from interactive data visualization systems in that they allow the user to both manipulate data [17] and facilitate sensemaking of data[8]. Sensemaking can be described as the process of understanding data within a certain framework and adapting the framework as new information is gathered [11]. Whereas a more typical interactive data visualization system will allow a user to manipulate data, they do not encourage users to engage in their own sensemaking about said data.

2.2 Software Artifact

The Data Explorer is a tool which allows local people to give input on data that directly relates to them in some way via the creation of data stories. The Data Explorer features a variety of data sets that relate to a specific topic such as water quality or income within a certain region. A topic in this context is simply any subject matter that researchers, educators, or community leaders may want to disseminate to a wider community. The long-term goal of the Data Explorer is to serve as a potential tool for citizen science and to provide a way for researchers to include non-experts in the research process. Data inclusivity is especially important when researchers are outsiders working with local communities [24].

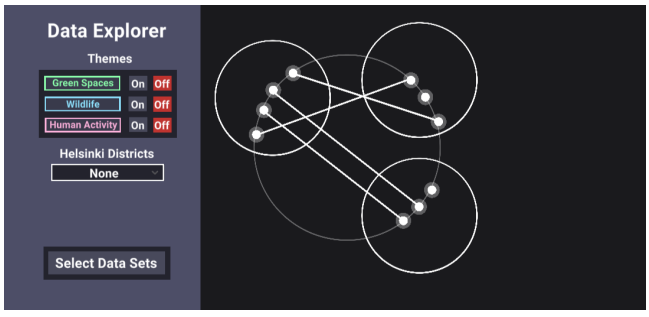

Figure 1: The Data Explorer's opening screen.

The Data Explorer allows users to view example data stories made from expert curated data sets, and then create their own data story using these curated data sets. Each data story contains relevant data sets that are visualized within a comic and includes

links to where the data set was taken from. The data comics provide the user with a point of entry into a data set as research has shown that a visual form or representation of data facilitates understanding, [10], users can then choose to look at the raw data set itself should they so wish.

Figure 2: A data comic open with The Data Explorer, with the user's data set collection shown on the side.

Each data set is represented by a white point or "star", each data story is represented by a circle or "constellation", data sets that are in multiple data stories are connected via lines. Users may also sort the data sets by theme or location as a means of looking for a particular type of data. When a theme is selected such as "green spaces" all data sets that relate to that theme will change color accordingly, and when a location is selected all data sets related to that location will change to a diamond shape. The current prototype of the Data Explorer is themed around the topic of biodiversity in Helsinki and therefore all the data sets, data comics, data stories, themes and locations relate to this topic.

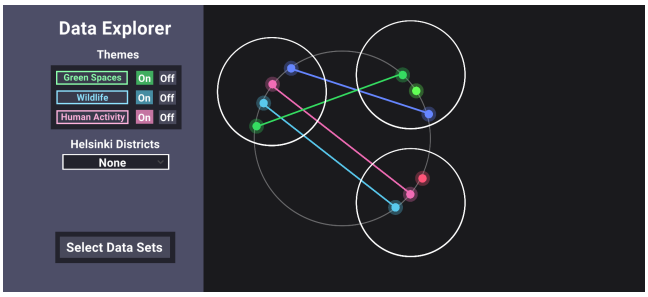

Figure 3: The themes buttons applied, and a location selected in the Data Explorer.

2.3 Original Design Principles

This tool was built using the design principles for data exploration systems aimed at non-expert users laid out in a previous research paper and therefore serves as a means to evaluate the efficacy of said principles. The original design principles were as follows:

(1) Allow users to find their own patterns in the data.
(2) Visualize relationships between data sets.
(3) Use artistic expression to engage non-expert users.

These design principles are meant to guide designers and software developers towards building tools which will facilitate the data exploration process, in particular, they are meant for tools which assist members of the general public (non-experts) in interacting with scientific data. In this study we will analyze the efficacy of these principles using the thematic analysis method.

3 RESEARCH METHODS

3.1 Data Gathering Methods

In this study our participants consisted of fifteen student volunteers who were unfamiliar with the Data Explorer tool, the volunteers ages ranged from 18 to 36, there were 8 female volunteers and 7 male volunteers. Our main data gathering method was the Think-Aloud Protocol [12], this method was chosen for this software tool as it can be performed easily by laypeople, this data gathering method involves asking participants to speak aloud about everything they are thinking of while interacting with the software. For this study we chose the Concurrent Think-Aloud Protocol, wherein users speak about the actions they are doing while they are doing them. Our secondary data gathering methods were eye-tracking [14] and mouse-tracking [9].

For guidance on interacting with the tool the users were given a series of 6 tasks to perform, these tasks asked the user to achieve a certain goal but did not tell them directly how to achieve it, thereby allowing them to explore the tool freely. The Think-Aloud data was collected via live audio recordings, these audio recordings were then made into transcript using speech to text software, and the eye-tracking and mouse-click tracking data was recorded via an eye-tracking software tool. A Parallel Scan Paths (PSP) visualization of the tracking eye data was used for the analysis. This eye tracking visualization method was chosen because it combines, a gaze duration sequence diagram (showing how long users looked at a point), a fixation point diagram (showing the path users eyes took) and a gaze duration distribution diagram (where users focused on average) [15]. The tasks were as follow:

(1) Find and open bat data set window.
(2) Add bat data set window to 'my data sets' collection.
(3) Find and open a data story that relates to bats.
(4) Find and open noise data set window.
(5) Add noise data set to 'my data set' collection.
(6) Create and upload placeholder data story.

3.2 Analysis of Data

For our data analysis process, we had three HCI researchers analyze the footage and transcripts of the users using qualitative data analysis methods.

3.2.1 Transcript Analysis. To analyze the Think-Aloud sessions we compared user's transcripts to the aforementioned design principles via a thematic synthesis [4]. We did this by searching for codes that relate to the stated purpose of the design principles, then examining the user's perspective in the case of the transcripts, and then categorizes those codes into themes that either provided a perspective on one of the design principles or related to a potential new principle. We also searched for the most common usability problems that users were facing.

3.2.2 Eye tracking, Mouse Tracking, and Screen Capture Analysis. For the analysis of the eye tracking, mouse tracking and screen capture data we chose to use a dyadic analysis technique [16] wherein a group of researchers examine the video footage of each participant and make qualitative assessments of the eye tracking, mouse tracking and screen capture data shown in the footage. For this study we chose to include a series of coding categories to assist the researchers in their assessment of the footage. This dyadic analysis allowed us to thoroughly examine principle 1 and 2 as via coding the footage we could examine if users are finding patterns in the data and understanding the relationships between data. Additionally, the eye-tracking data allows us to more deeply evaluate principle 3, as we could directly see if participants were more likely to focus on, I.e., be engaged by artistic representations on the screen [23].

3.3 Coding Process

In order to code the data, we first came up with a series of initial categories of codes that we would be looking for, as well as the types of codes that may relate to these categories. Each category had transcript codes (codes found in the transcript), passive action codes (codes that relate to a user's gaze) and active action codes (any codes that involve users making an intentional action). Each category related to one of the three design principles, below you can see a diagram of the initial codes categories.

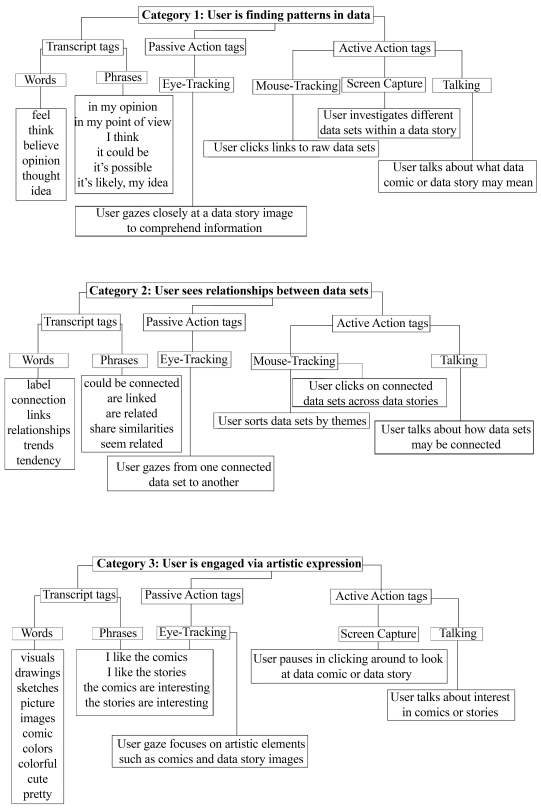

Figure 4: Diagram of initial codes categories.

3.4 Thematic Synthesis

A thematic synthesis was then conducted wherein we synthesized shared themes from the codes. A few key themes emerged, namely a. themes that related to visualizing relationships, b. themes that related to users finding patterns in data, c. themes that related to engaging users through artistic elements, d. themes that related to clearly labeling elements in the visualization, and e. themes that related to providing narrative context to users. These themes became the basis for our refined principles for designing data exploration systems for non-experts. Below you can see a diagram of the process, and our codebook is linked in the footnotes.
1

Figure 5: Diagram of the thematic synthesis process.

4 FINDINGS

4.1 Refined Principles for Designing for Data Exploration for Non-Experts

From the thematic synthesis we derived a refined version of the previous design principles which we elaborate upon below.

4.1.1 Allow users to find their own patterns in the data. In our previous paper we proposed that a data exploration system or tool should allow users to find their own patterns in data and form their own hypothesis about what may be happening in the data. This principle was seen in action when users were able to find their own patterns in the data as indicated by their actions and comments. Enabling pattern finding in data remains a key element of data exploration, as without giving users a means to find patterns data

[1] https://doi.org/10.5281/zenodo.7686010

exploration would not be possible [19]. In our qualitative analysis we found that a hindrance to this pattern finding process was a lack of understanding of the narrative context of a data set or data story. As such we have added another principle for designing for data exploration for non-experts that relates to narrative context.

4.1.2 Provide narrative context. In this paper we are defining narrative context as the set and setting of where the data was taken from [24], and the central story of that data exploration system [6]. A lack of narrative context can leave a user with a lack of direction within a data exploration system. As data exploration systems are by their design, intentionally open-ended terms of interaction, failing to provide sufficient narrative context can lead users to not understanding the purpose or goal of the exploration system. This was reflected when users indicated that they felt lost, confused or uncertain in the purpose of the tool or what their ultimate goal as a user was.

4.1.3 Visualize relationships between data sets. One of the other original principles was that a data exploration system should visualize relationships between data sets. This principle proved useful and necessary as without visualizing the relationships between data sets users are unable to engage in sensemaking [13]. However, a common problem that emerged in this area was that users were often confused by the lack of labeling in The Data Explorer. Therefore, we propose to add another principle that relates to proper labeling of elements in a data exploration system.

4.1.4 Clearly label all elements. Our analysis shows that clearly labeling all elements in a data exploration system or tool will drastically reduce user confusion. Many of the most common pain points encountered in The Data Explorer related to confusion over improper labeling, such instances where users were unable to find the same data story twice or were confused about which window represented a data comic and which a data story.

4.1.5 Use artistic expression to engage non-expert users. Finally, our last design principle remains unchanged from the previous paper. It has been previously established that engaging users through visual and artistic elements is effective [1, 5, 18] and this is reflected in our findings. We counted over 40 instances across all users of a user either actively or passively looking at an element artistic element on the screen. Likewise, users frequently verbally expressed positive interest and engagement with the artistic elements of the tool.

4.2 Recommendations for changes to Data Explorer Tool

We also compiled a list of recommendations for feature changes that can be made to the Data Explorer tool based on common pain points and the extended design principles. This list consists of 10 recommendations based on a total of 19 pain points that were found across all users. These feature change recommendations can be seen in the graphic below.

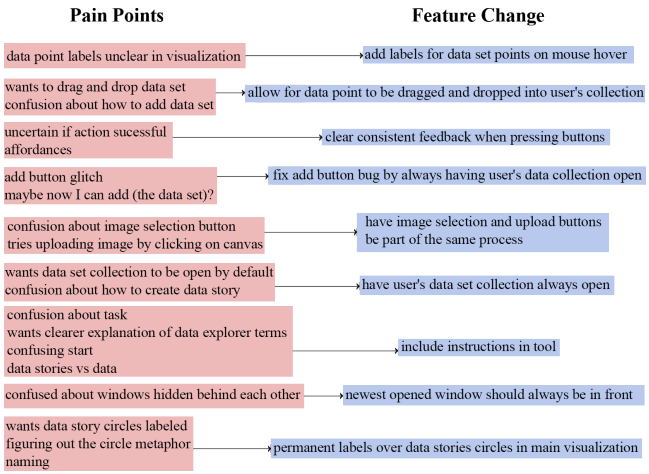

Figure 6: Feature Changes based on Pain Points.

5 LIMITATIONS

One limitation that we faced in this study is that we were limited to three researchers for the coding and thematic analysis process. Ideally, we would have preferred to have had more researchers participate in the coding process, as codes are not objective measurements and there is always the potential for biases to arise during code selection. Additionally, in an attempt to allow the users to explore the tool freely we chose tasks that asked users to achieve specific goals but did not provide users a wider context of how to achieve those goals. This led to some users being initially confused as to what they were being asked to do and in part motivated the addition of the principle of "providing narrative context". Preferably we would have liked to have a workshop with a small group of preliminary users to test out the task list, however due to both time constraints and resource constraints this was not possible. Eventually we would like to conduct an additional study of on the Data Explorer using a similar list of task that have been improved based on the user feedback from this study.

6 CONCLUSION

The aim of this qualitative study was to improve upon a previously established series of design principles for designing data exploration tools for non-expert audiences. We did this by assessing our data exploration tool, which was built upon said principles, with volunteer non-expert users. From this assessment, we expounded those initial design principles by adding two more. The final principles were as follows. 1. Allow users to find their own patterns in the data. 2. Provide narrative context. 3. Visualize relationships between data sets. 4. Clearly label all elements. 5. Use artistic expression to engage non-expert users. In the future we hope to iterate upon the Data Explorer based on these refined principles and test it in collaboration with a real world organization where data dissemination, and exploration and inclusion tools for the general public are needed, such as a non-profit, government or community organization. Likewise these principles can be used for designing industry software tools wherein the general public must interact with big data.

7 ACKNOWLEDGEMENT

Thank you to the participatory communication of science (ParCos) project, for funding this research. The ParCos project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 872500.

REFERENCES

[1] Benjamin Bach, Nathalie Henry Riche, Sheelagh Carpendale, and Hanspeter Pfister. 2017. The Emerging Genre of Data Comics. *IEEE Computer Graphics and Applications* 38 (05 2017), 6–13. https://doi.org/10.1109/MCG.2017.33

[2] Nikos Bikakis, George Papastefanatos, and Olga Papaemmanouil. 2019. Big Data Exploration, Visualization and Analytics. *Big Data Research* 18 (12 2019), 100123. https://doi.org/10.1016/j.bdr.2019.100123

[3] Nikos Bikakis and Timos Sellis. 2016. Exploration and Visualization in the Web of Big Linked Data: A Survey of the State of the Art.

[4] Daniela Soares Cruzes and Tore Dybå. 2011. Recommended Steps for Thematic Synthesis in Software Engineering. *2011 International Symposium on Empirical Software Engineering and Measurement* (2011), 275–284.

[5] Vanessa Echeverria, Roberto Martinez-Maldonado, Roger Granda, Katherine Chiluiza, Cristina Conati, and Simon Buckingham Shum. 2018. Driving data storytelling from learning design. 131–140. https://doi.org/10.1145/3170358.3170380

[6] Nahum Gershon and Ward Page. 2001. What Storytelling Can Do for Information Visualization. *Commun. ACM* (2001). https://doi.org/10.1145/381641.381653

[7] Stratos Idreos. 2013. *Big Data Exploration.*

[8] Stratos Idreos, Olga Papaemmanouil, and Surajit Chaudhuri. 2015. Overview of Data Exploration Techniques. 277–281. https://doi.org/10.1145/2723372.2731084

[9] Pascal Kieslich, Felix Henninger, Dirk Wulff, Jonas Haslbeck, and Michael Schulte-Mecklenbeck. 2019. *Mouse-tracking: A practical guide to implementation and analysis.* 111–130.

[10] Andy Kirk. 2016. Data Visualisation: A Handbook for Data Driven Design.

[11] Gary Klein, J.K. Phillips, E.L. Rall, and Deborah Peluso. 2007. A data-frame theory of sensemaking. *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (01 2007), 113–155.

[12] C. Lewis and J. Rieman. 1993. Task Centered User Interface Design: A Practical Introduction. (01 1993).

[13] Riccardo Mazza. 2009. *Introduction to Information Visualization.* https://doi.org/10.1007/978-1-84800-219-7

[14] Ana Pellicer-Sánchez and Kathy Conklin. 2020. *Eye-tracking as a data collection method.* 370–382.

[15] Michael Raschke, Tanja Blascheck, and Michael Burch. 2013. *Visual Analysis of Eye Tracking Data.* https://doi.org/10.1007/978-1-4614-7485-2_15

[16] J. Saldaña. 2009. *The coding manual for qualitative researchers.*

[17] Martin Schroeder, Dan Cornford, and Ian Nabney. 2009. Data Visualisation and Exploration with Prior Knowledge. *Communications in Computer and Information Science* 43, 131–142. https://doi.org/10.1007/978-3-642-03969-0_13

[18] Edward Tufte. 2001. The Visual Display of Quantitative Information / E.R. Tufte. *American Journal of Physics* 31 (01 2001). https://doi.org/10.1109/MPER.1988.587534

[19] N Tylosky, A Knutas, and A Wolff. 2023. Principles for Designing for Data Exploration for a Non-Expert Audience. *Manuscript accepted for publication* (2023).

[20] Eliane Victorelli and Julio dos Reis. 2020. Human-data interaction design guidelines for visualization systems. 1–10. https://doi.org/10.1145/3424953.3426511

[21] Eliane Victorelli, Julio dos Reis, Antonio Santos, and Denis Schiozer. 2020. *A Design Process Integrating Human-Data Interaction Guidelines and Semio-Participatory Design.* 334–360. https://doi.org/10.1007/978-3-030-40783-4_16

[22] Eliane Zambon Victorelli, Júlio Cesar dos Reis, Heiko Horst Hornung, and Alysson Bolognesi Prado. 2020. Understanding human-data interaction: Literature review and recommendations for design. *Int. J. Hum. Comput. Stud.* 134 (2020), 13–32.

[23] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing Effectiveness and Engagement of Data Comics and Infographics. https://doi.org/10.1145/3290605.3300483

[24] Alexander Loukissas Yanni. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society.*

[25] Ji Soo Yi, Youn ah Kang, John Stasko, and J.A. Jacko. 2007. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1224–1231. https://doi.org/10.1109/TVCG.2007.70515

[26] Elena Zudilova-Seinstra, Jean-bernard Martens, and Tony Adriaansen. 2010. Interactive data exploration and knowledge discovery. 421–422. https://doi.org/10.1145/1842993.1843099