

Exploratory Visualization for the Women's Print History Project

by

Parnian Taghipour

B.Sc., Amirkabir University of Technology, 2021

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
School of Computing Science
Faculty of Applied Sciences

© **Parnian Taghipour 2023**
SIMON FRASER UNIVERSITY
Summer 2023

Copyright in this work is held by the author. Please ensure that any reproduction or re-use is done in accordance with the relevant national copyright legislation.

Declaration of Committee

Name: Parnian Taghipour

Degree: Master of Science

Thesis title: Exploratory Visualization for the Women's Print History Project

Committee:

Chair: Xing-Dong Yang
Associate Professor, Computing Science

Sheelagh Carpendale
Co-supervisor
Professor, Computing Science

Thomas Shermer
Co-supervisor
Professor, Computing Science

Michelle Levy
Committee Member
Professor, English

Lawrence Kim
Examiner
Assistant Professor, Computing Science

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

- a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

- b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

- c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

I designed, developed, and studied an exploratory visualization for the Women’s Print History Project. The women’s print history project is a manually collected dataset that spans the years from 1700 to 1850, collecting publications in which women have been involved through any of the publication roles. My visualization research, done in consultation with a domain expert, is part of an ongoing trend to support humanities in visualization research. I designed and developed an exploratory visualization and then through a user study involving —data experts and general users—I investigated the tool’s effectiveness in meeting the experts’ research needs and its overall user experience. My contribution lies in understanding the domain needs and tasks, and in designing, and implementing a tool in response to these needs of the data experts.

Keywords: Visualization; Data Exploration; Design Study; Humanities datasets

Dedication

To my parents and lovely sister, for being the rock upon which I stand.

Acknowledgements

I am truly grateful for the unwavering support and invaluable mentorship provided by my supervisors, Dr. Sheelagh Carpendale and Dr. Thomas Shermer. I consider myself fortunate to have had the opportunity to work under Dr. Carpendale's guidance. My journey into the world of visualization and HCI began with her Visualization course, and she played a pivotal role in sparking of my interest in study-based research. Her mentorship style empowers you to stand on your own while offering huge support, especially during crucial deadlines. This project has brought me closer to embracing the mindset of a researcher, thanks to the knowledge I've gained under her guidance. The dual mentorship of both Sheelagh and Tom is a unique privilege. My journey to SFU commenced by Tom and has been consistently supportive of my decisions since. His encouragement led me to explore diverse areas and ultimately discover my passion for visualization and HCI.

I extend my heartfelt gratitude to my labmates at the IXlab for keeping a warm and supportive environment in the lab. I wish to acknowledge the invaluable advices of my colleagues, especially Maryam Rezaie, Foroozan Daneshzand, Narges Ashtari, and Rimika Chaudhury, who patiently answered my questions throughout the project. I must also mention my labmate Parsa Rajabi, with whom I collaborated on my first paper during his project.

Special thanks are due to Dr. Michelle Levy, whose commitment as a committee member and data expert significantly contributed to the project. Lastly, I extend my sincere appreciation to Dr. Lawrence Kim for accepting the role of my examiner and Dr. Xingdong Yang for chairing my thesis defense.

Table of Contents

Declaration of Committee	ii
Ethics Statement	iii
Abstract	iv
Dedication	v
Acknowledgements	vi
Table of Contents	vii
List of Figures	ix
1 Introduction	1
1.1 Motivation	2
1.2 Thesis Statement	4
1.3 Approach and Contribution	4
1.4 Outline	4
2 Literature Review	6
2.1 What Kind of Network we are Working with?	6
2.2 Visualization tools to help Humanities Studies	6
2.3 Information Seeking	9
2.4 Complex Network Visualization	11
3 Introduction to Data and Tasks	14
3.1 Data set	14
3.1.1 Data Challenges	15
3.2 Understanding the Domain and Task	15
3.2.1 Gathering and Verifying Data	15
3.2.2 Writing Spotlights and Making Podcasts	16
3.2.3 Data Changes	16
3.3 Design Goals	16

4	Design Process	18
4.1	Iterative Design Process	18
4.2	Design Methodology	19
4.3	Early Design Choices	20
4.4	Final Design Description	21
4.4.1	Timeline and Information Box	21
4.4.2	Filters	27
4.5	Implementation	28
5	Qualitative Study and Results	33
5.1	Study methodology	33
5.1.1	Tasks (1)	34
5.1.2	Tasks (2)	34
5.1.3	Semi-structured Interview	35
5.2	Data Analysis	35
5.3	Results	35
5.3.1	How did Participants encounter tasks?	36
5.3.2	Overview helps in different ways	37
5.3.3	Learning Curve	37
5.3.4	Data Experts See More in the Data	38
5.3.5	Data Discrepancies Prompt Exploration	38
5.3.6	Data Experts Encountering Things They Usually do not See	40
5.3.7	New Interesting Insights about Experts' Tasks	41
5.3.8	Non-Data Experts Suggesting Possible Use Cases	41
5.3.9	User Experience and Suggestions	42
5.4	Study Summary	43
6	Conclusions	44
6.1	Discussion on our Design Goals	44
6.2	Lessons Learned	45
6.2.1	Data Expert Collaboration Needs Effort from Both Sides	45
6.2.2	Using Familiar Designs is Not Always the Key	47
6.3	Limitation and Future Work	47
6.4	Conclusions	48
	Bibliography	49
	Appendix A Study Material	58
A.0.1	Pre-Study Questionnaire	58
A.0.2	Post-Study Questionnaire	61

List of Figures

Figure 2.1	Exploratory Search	9
Figure 4.1	In the initial designs, we attempted to stretch timelines for zooming purposes; however, we observed that numerous data points coincided at the same positions, and stretching the timeline did not alter this overlapping issue.	20
Figure 4.2	Overview of the prototype	22
Figure 4.3	Overview of the prototype	23
Figure 4.4	Hovering on a rectangle in different years reveals connections between two timelines in the same year. This figure is about persons and title timelines demonstrating the years in which individuals initiated their involvement in publications and the subsequent years of their contributions.	24
Figure 4.5	clicking on lists will show the items as a scrollable list, and will push all of the items down	25
Figure 4.6	Hovering on the names in the scroll-able will show the links to the person's publications	26
Figure 4.7	Clicking on the names in the scroll-able list, a timeline for one person will appear with circles on the timeline representing the publications	29
Figure 4.8	Clicking on a circle with a black border on the timelines, a spiral will appear with the publications inside that circle around it	30
Figure 4.9	It is possible to change the color of the timeline using a color picker. In this image, the red timeline opened before the blue one because of that it is closer to the title timeline	30
Figure 4.10	After a long click on the timeline name-tag the trashcan button will appear that will delete the timeline	31
Figure 4.11	If there is a circle representing a firm or person on the screen the name of the person in the scrollable list will be bold	31
Figure 4.12	Clicking in the titles, the triangles will appear to help the user find the related firm and person	32
Figure 4.13	Filters	32

Chapter 1

Introduction

In my research, which I describe in this thesis, I extend the exploration of the potential of visualization in the humanities. Humanities researchers have broad research areas, with multi-dimensional data and complex datasets, which make them challenging and interesting for visualization. The definition of Humanities research based on Berkeley University's humanities website [3] is

Research in the Arts and Humanities is about asking questions and discovering new ideas, often based on analysis or close inspection of artifacts (texts, artwork, films or other human productions).

Based on this definition humanities research is not uniquely based on the quantity of data; it consists of critical thinking and making interpretations of the context of the data, patterns, and theories. Based on the humanities' challenging research process, it is difficult for visualization tools to actually demonstrate their potential in the humanities research process [63]. Humanities research is broad and consists of different fields: for this project, I collaborated with domain experts in the English department, working on the Women's Print History Project (WPHP).

Visualization is a powerful tool for representing data in visual form using visual variables, such as color, text, size, shape, etc. Its ability to present complex information in a visually comprehensible way created a revolution in the way we understand and analyze data, and it helps people perform better in different tasks. From scientific research and business analytics to urban planning and healthcare, visualization has proven valuable in uncovering patterns, revealing insights, and facilitating decision-making processes [100][75].

However, visualization is not limited to quantitative data alone. Its application in the humanities has emerged as a promising area of research, offering new possibilities for exploring and understanding culture, literature, historical data, and more. In short, visualization in the humanities extends beyond the presentation of results; it can help people interact with their data and help in having new insights and emerging ideas [26].

The possibility of using visualization in different areas created great opportunities for interdisciplinary research. With collaboration between data experts and visualization experts [114], it gets possible to make visualizations that can effectively help other disciplines research and get blended into their research process.

In this chapter, I provide a brief introduction to this thesis. Section 1.1, provides the motivation. Section 1.2, explains the research statement; In Section 1.3, we define the scope of this research. Section 1.4 describes the method and contribution of this research. Section 1.4 provides a brief overview of this thesis.

1.1 Motivation

In today's data-driven world, the collection and analysis of historical data has become increasingly important in gaining valuable insights and understanding various aspects of our past. As data becomes a more dominant part of everyday life, people strive to gather more information by collecting and analyzing data from different sources. Historical data, in particular, offers a wealth of knowledge about our cultural heritage, societal developments, and the lives of individuals who have shaped our history. However, extracting meaningful information solely from looking through historical records can be challenging. Therefore, visualization methods can be helpful to bring hidden patterns to the surface. In this thesis, I aim to explore the challenges associated with visualizing historical data, focusing specifically on the Woman's Print History Project (WPHP) dataset [58]. The WPHP is a growing hand-assembled dataset that aims to provide comprehensive information about women in publications between 1700 and 1836. It is already a valuable resource for understanding the contributions of women to the print industry during this period. Women contributed to these titles with different roles, for example, a woman might have contributed to a book as an author, illustrator, editor, publisher, etc. The dataset has three main sections: people [57], titles [110], and firms [35] each of which is being manually collected. I am working with researchers from this project, to be able to make an explorable tool for their dataset, to bring hidden information and patterns to the surface.

Researchers from WPHP are gathering all of the possible information and are exploring its reliability and trustworthiness, verifying accuracy whenever possible using different sources, such as English Short Title Catalogue (ESTC) [2], Eighteenth Century Collections Online (ECCO)[1]. However, a difficult aspect of collecting the data for the researchers is that the information is often not available in digital format, forcing manual data collection. Because it was historically collected by different people it may not be totally coherent, and there are inevitable gaps in the dataset, which reflect gaps in the historical knowledge. Consistent formatting might be difficult because the original information is so varied. Besides, this dataset includes information from marginalized groups that have been overlooked in the past. Thus much useful information may have been overlooked or not considered im-

portant enough to record. Therefore, the completeness of the data may vary significantly depending on the sources from which it was collected. Also, we need to consider that authors and contributors at that time may not always share the same amount of information about themselves. For example, some authors may have used pseudonyms or female authors may have published under male names, their husband's name, or even simply as a 'by a lady'. Additionally, there may be missing information about dates, with only a possible range or even no year known. In short, since this data is modeling complicated human/object relationships over time and space, there are interesting data complexities, and there are inevitable gaps in the dataset. Although these shades of missing data make it hard to work with this data and visualize it, they are a rich part of data to shed light upon that era. By visualizing the data, the researchers aim to identify and analyze these variations and incompleteness to gain a comprehensive understanding of the historical context, and the challenges women in print faced during the 18th and 19th centuries. Furthermore, the variations and incompleteness in the dataset can provide valuable insights into the social, cultural, and economic factors that influenced women's contributions to print during this time. Also, knowing about these resources and the relationships between data and these resources can help data experts to collect better-quality data. The combination of these challenges and finding a place for visualization in dealing with these challenges was the motivation behind starting this project.

The challenges associated with visualizing historical data are at the confluence of several active research directions in the field of visualization. This project involves working with a domain expert, leveraging personally collected data, dealing with various types of missing values, and handling a growing dataset with over 20,000 entries. Although 20,000 entries may be considered small in the context of big data [66], it presents challenges in terms of screen space usage and interactions. Consequently, designing an effective visualization requires careful consideration of these limitations while providing meaningful insights to the users.

While, just above, we have mentioned some of the anomalies in the WPHP dataset. Here, we want to emphasize how the combination of data anomalies can be an interesting challenge when approached from the perspective of data visualization and analysis. In the realm of data visualization, handling and working with collected datasets that have inconsistencies is a common challenge faced by the visualization community. One common practice to handle this situation is to do data cleaning and data wrangling before starting to work with the dataset. Cleaning the data will simplify the programming process and make it easier to visualize the data. However, as the inconsistency in our dataset represents the historical context, it needs to be preserved. Therefore, it is one of the major motivations of this research to face this challenge without altering, imputing, or deleting the records with discrepancies.

In short, this thesis presents a collaboration between visualization experts and domain experts from the WPHP. The collaboration seeks to develop a visualization that can effectively cope with data inconsistencies and discrepancies while enabling data experts to explore and understand the dataset's complexities. The visualization aims to support data exploration, helping experts identify patterns and trends, and providing a deeper understanding of the role of women in print history.

1.2 Thesis Statement

I am working with humanities scholars who are manually gathering data about gender roles in publications from 1700 to 1836. In my research, I am investigating whether an exploratory visualization tool designed for these experts can support them in their ongoing research.

To do this we have worked closely with data experts [114] [90] to:

1. Understanding the domain problem and data

In order to reach an understanding of Data experts, their tasks, and needs we have had several meetings with them and extracted their needs based on our conversation, sketching session with the main data experts, and discussed their tasks again with them.

2. Design a tool based on the understanding developed in my previous step

We followed the user-centered design process in our design process; after defining design goals based on our findings from the first part, we have done more brainstorming and sketching; next, we gradually refined our design while implementing and progressively making design decisions based on domain experts' feedback, and adding new features.

3. Qualitative study of our visual exploratory tool.

1.3 Approach and Contribution

Our research is a collaboration between data experts and visualization experts. The first contribution of this research is extracting the domain needs and tasks. The second contribution is an effective design and implementation of a tool that we studied to learn more about how it answers the needs of data experts. We explain our design goals and justified our design choices based on the design goals. Third, we studied our design with two different groups, data experts and general people, to get feedback both about the effectiveness of this tool in the expert's research and general user experience. Finally, we discuss the lessons we learned in this design study.

1.4 Outline

The following chapters describe various parts of our research:

Chapter 2 – A Background in Visualization Design:

In this chapter, we describe the current literature on visualization in humanities and the different goals, challenges, and designs, which were required to understand the domain, and challenges of our collaboration, and design our tool.

Chapter3 – Introduction to Data and Tasks:

In this chapter, we explain domain experts' needs and the data that we are using and extract design goals and considerations.

Chapter4 – Design Process:

This chapter describes how we designed XVis based on the functional requirements elicited in Chapter3, to support collaboration between a domain expert and a visualization designer in creating and discussing visualizations together.

Chapter5 – Qualitative Study and Results:

This chapter explains the study conducted to investigate the usefulness of XVis in domain experts' research, and also the usability of XVis. It also provides details on the approach taken to conduct the study and its results.

Chapter6 – Conclusion and Future Work:

This chapter provides the conclusion and guidelines for future work in this area.

Chapter 2

Literature Review

In this chapter, I will review previous research relevant to my work. My focus is on three areas: research on visualization tools for humanities, information seeking, and complex network visualization.

2.1 What Kind of Network we are Working with?

Before delving into the literature, it's important to establish a background on the network we are presenting. The existing network within the data comprises various types of entities: People, Titles, and Firms. These entities serve as the vertices of the network. It's worth noting that this network does not exhibit connections between all entities. Instead, Titles are linked to both People and Firms separately, resulting in two distinct types of edges: title-people and title-firms. Additionally, all connections between People and Firms occur through the intermediary of Titles.

2.2 Visualization tools to help Humanities Studies

I will start by going over some papers that are about digital humanities. The main goal is to learn the state of collaboration between Humanities researchers and visualization researchers and the challenges and findings in this area.

As the application of information visualization expands across various fields and proves its usefulness, the intersection between humanities and visualization has also seen a rise in research [12]. There are many papers [107, 4, 85, 28] that have explored text analysis and visualization to assist humanities researchers with their text-related tasks. Moreover, methodologies have been introduced for working with large text-based data sets in humanities, emphasizing both close reading and macro analysis [47]. Besides, various visualization techniques have been employed to aid humanities researchers in text mining and exploring large text collections, including word clouds [108], maps [43, 72], word tree maps [72], scatter plots [82], and parallel word clouds [24]. Graphs and connected list views have also been utilized to display relations within large text collections, as seen in the example of

jigsaw [101]. As an instance, Hinrichs et al. [43] combined text mining and visualization techniques to develop a tool for monitoring the growth of global commodity trade in the nineteenth century. Their design primarily relied on maps and bar charts. Several other tools and research. There were also other tools that had gathered the network out of the text. For example, Bingenheimer et al. [13] extracted a network of historical Chinese Buddhist monks from texts and created a visualization of their social network, consisting of 5,900 different individuals. They employed a social network visualization approach, representing the monks as nodes spread across the screen. However, the networks were usually uni-dimensional (meaning all of the vertices in the network are from the same category; more explanation in section 2.3)

One popular approach to deal with complex humanities data is incorporating multiple visualization views that focus on different metadata. This means having different sections of screen dedicated to different visualizations that are showing different aspects of data. Various investigative tools in the humanities have been developed to assist researchers in different topics. For example, an overview tool aids journalists in reading diverse documents by using multiple views like topic trees, scatter plots, and cluster and document lists [17]. In most of these tools these different visualizations are synced with each other, which means interaction with one of the views will alter the others, called "cross-filtering". Different cross-filtering techniques have been studied [113, 91]. As some examples, Meinecke et al. [71] focused on creating a steam graph and a multi-view interface to facilitate an explorable visualization for finding similarities among rappers based on their lyrics. EMDialog [46] developed an explorable visualization tool for museums, using Emily Carr data as a sample, and incorporated two visualizations to represent the context and temporal data. These studies also have shown that having multiple views is effective for explorability of the visualization tool. Some of these tools also have the option of visualize the networks as a part of their tools. As an instance, one of them [91] included a network view as one of the views on the screen, but it was organized only after applying filters, and it did not serve as the main network for exploration. Wireframe [103] separated different aspects of letter collections, presenting them in different facets such as time and network. It is also worth mentioning that various visualizations use multiple views of the dataset to assist data experts in identifying anomalies and understanding their underlying reasons [6][120]. In this project, while the usefulness of multiple views in data visualization has been acknowledged in many cases, certain constraints have led us to prioritize simplicity and clarity on a laptop screen. Considering these limitations, opting for multiple views might not be the best choice. Moreover, the project's primary focus is on exploring relations and networks within the dataset, which helped us to decide not to emphasize multiple views at this stage.

Another crucial aspect we examined during the literature review was the relationship between visualization and humanities research. Heinrich et al. [44] focused on using multi-view visualization for a specific literary collection, and developed a tool that promotes

critical thinking and curiosity. Their visualization included keyword clouds, timelines, and hierarchy trees as cross-views. As part of their research process, they have framed humanities design space within four aspects: audience, use case (result or process), approach (qualitative and quantitative), and visual qualities. Considering the usecase some research claimed that though visualization is utilized to present findings in humanities research, its usage in the research process is not as common [39]. One of the concerns of humanities researchers is that literature is not merely data to be treated in the same manner. It's crucial to acknowledge that identical words may have distinct meanings [5]. In recent years, researchers have looked into why the collaboration between humanities and visualization is not more effective [52] and used more user-centered designs to involve humanities scholars and better understand their needs as end-users [104]. Moreover, Lamqaddam et al.'s paper in 2021 [53] aimed to understand what humanities researchers think about visualization and how to make collaboration more effective. Their main finding was the importance of reducing the semantic distance between visual representation and the actual data. To achieve this, they suggested guidelines such as transparency of data and using terminology familiar to the experts. We understand that showing data experts their actual data can help them in using a visualization tool (Section 3.3, Design Goal 5). There are also researches claiming that presenting data in its raw form, without cleaning or imputation, offers the advantage of assessing data quality and gaining insights into missing and incomplete data [119][69]. Keeping missing data and errors in the visualization can significantly impact user experience and the sense-making process [77]. In our collaboration, our goal is to create a visualization that supports our collaborators in their research process. Based on our goal, we find it important to focus on the collaboration with data experts, and we made the decision not to clean or impute the data, and to keep the missing data on the screen (Section 3.3, Design Goal 2). Besides, in a subsequent paper, they introduced the concept of "sandcastle" in which they more focused on the research design process and invited researchers to reflect on their design process [45]. As a result, in this thesis, we also report our observations to reflect on our design process.

Another interesting topic in literature was the use of explorable visualization in humanities (I discussed explorability in general and more in section 2.3). Whitelaw [116] argued that search engines are not the ideal solution for browsing collections, suggesting the need for more generous interfaces for cultural collections. In terms of explorability, various concepts have evolved over time. Diverse researchers are also moving towards explorability in the humanities; for example, the use of 3D environments to showcase historical information in a network has also been investigated to improve serendipity. [40]. Dörk et al. [34] proposed four designs for their humanities collection, finding that incorporating time as an aspect of visualizations makes it easier for users to filter and explore the data. They also integrated a separate network view that was interconnected with the timeline view. However, Windhager [117], in 2017, reviewed 70 papers related to digital humanities and observed

that exploration concepts are often designed for casual users. However, considering that humanities research does not revolve around reaching a single interpretation, exploratory visualizations can be beneficial for their research process(Section 3.3, Design Goal 1).

In this initial section, we have explored the relevance of visualization research in the context of humanities. However, to strengthen our understanding of explorability and network visualization, which are the primary focus of this study, further investigation is required. In the forthcoming sections, while maintaining a central focus on these topics in humanities, we will also delve into relevant papers and research from other domains to gain comprehensive insights and inspiration for our visualization project.

2.3 Information Seeking

In this section, I will delve into the literature review on information seeking and explore the concept of "explorability" in the context of information seeking. Helping users to find information is one of the main blocks of this research project. Information seeking traditionally happens by using search engines employing different approaches. Information-seeking behavior is similar to the process of berry picking, and it is based on the back-and-forth process of browsing and searching. Through this process, the user will learn [10]. Marchionini [65] expanded the idea of user browsing behavior and separated search activities into three categories (Figure 2.1). He grouped learning and investigating phases together as exploratory searches, suggesting a balanced consideration of querying and browsing to create an explorable tool. Kules et al. [50] defined some characteristics based on information-seeking behavior to generate exploratory tasks suitable for user studies. The main feature of exploratory tasks based on their research was asking questions that were not exactly clear to the user. White et al. [115] attempted to address the exploratory search problem by exploring different tasks supported by various exploratory systems, moving toward "richly associated search across heterogeneous spaces." There are also two types of looking through exploratory visualization; one is with considering that the end-user already now the dataset and want to explore more [11], and the other is through investigative exploration [26].

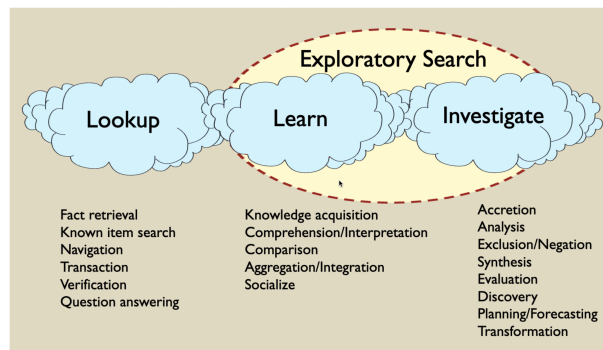


Figure 2.1: Exploratory Search

In terms of multi-view visualization, many visualizations for large collections of digital items use metadata to support searching, browsing, and filtering. There have been numerous visualizations and research utilizing data attributes as facets, which are different possible classifications of data that can be combined to filter resources. Facets are sets of categories that describe various aspects of the items in the collection. Visualizations using facets are useful in giving a high-level overview and making sense of the collection as a whole [56][19]. Early interfaces used lists or columns of facets to navigate resources and help users explore new choices [87] [118]. Faceted interfaces proved to be powerful for filtering large sets of resources. Later, researchers tried different approaches to create explorable visualizations from faceted information collections. For example, visual encodings that can show the distribution of items in different facet values [102], and techniques that can reveal cross-facet relationships in query results [31] have been used to guide users through their exploration. RomaAV [22] used multiple views to show the relationship between places, people, and history, enhancing the sense-making process. There is also research that uses network views along with other views to show the network of people working in art in the past and make connections between people explorable [86]. ConTour [79] used a faceted view to support drug discovery by facilitating exploration with different detailed views. In terms of multi-dimensional network data, there has been work toward using a cross-views faceted interface [48] [55]. In these works, they performed different queries for networks in different views; however, these views are in different small windows that are not linked to each other. There were also systems like Lifeline [81] which used one view to show faceted data in one interface by using different visual elements on a timeline; their design was good for showing missing data and exploration, but the exploration was limited to one item at a time and not the links between items. These tools are all about explorability, however, they focus on several views, which is not my focus in this project. Although these projects are different from mine, they gave me a better understanding of explorability. Although these papers were on different types of visualization, most of them tried to make their system explorable by using various approaches to navigate users through data. Besides, due to the large number of data points in collections, aggregation and grouping in the overview have become necessary to ease perceptual challenges [97]. However, interactions such as filtering and zooming in the overview can be challenging. User-centered methods like "details on demand" (degree-of-interest [20]), "overview+detail" [23], and "focus+context" [106] have been used in many explorable visualizations in humanities. However, the application of these methods is highly dependent on the type of visualizations. As a result, in the next section, I will concentrate on visualizing multi-dimensional networks.

New concepts emerged in visualization theory as a result of merging information visualization with concepts such as creativity, serendipitous discoveries, etc. The notion of information "flâneur" used by Dörk et al. [29], coming from the area of urban design, is one example. They define it as an information-seeking method. Search engines, as an old method

of information seeking, proved to be accurate and fast; however, information flâneur is about creativity, curiosity, and explorability. There are also other popular approaches focusing on explorable tools for investigating information. For example, the Bohemian bookshelf project was focused on having different views of books and considering books' physical features to support serendipity discoveries. Their method of having multiple entry points and coordinated views proved useful for casual users [37]. Smets [98] argues that there are four ways that designers can talk about serendipity. In our use case, we see serendipity as a feature. We need to show our experts unexpected parts of their data because they are interested in inconsistencies, and it has a positive effect on their research. This project will not have a search engine to increase the chance of unexpected findings as a result of free exploration of the tool (Section 3.3, Design Goal 1).

2.4 Complex Network Visualization

As one of the main focuses of this project is on understanding the intricate connections within the data, an overview of the complex networks literature becomes essential. The multi-model graph is defined as a graph $G(V, E)$ in which the vertices (V) can also be grouped into different categories [25]. The dataset we are working with and trying to model exhibits a connection structure comparable to a multi-model graph, where various entities such as people, firms, and books are interconnected. One of the key benefits of using multi-model visualization is its capability to reveal the reasons behind connections; for example, it can show that two people are connected because they have contributed to the same book. However, visualizing the multi-model graph can be challenging, because of the data complexity, and screen limitations [30].

Thinking about trying to have an indicator for each data point in the visualization (Unit visualization [78]), one way of dealing with the multi-model graph is to use projection to decrease the complexity. For example, only using the network between people, and using the titles data only to find the connection between people [49]. I am not focusing on this approach as it will also decrease the information. Another common practice is to represent it in a uni-modal manner, using different shapes and colors to distinguish between the different types of entities within the same graph[9]. However, even using the methods to decrease the visual complexity of these methods for large datasets[93] [94], these methods will either increase the complexity or decrease the information.

To reach a visually simple network visualization, we need to focus on aggregated visualization [36]. Aggregated visualization means merging several data points into a signal visual indicator; such as rectangles in bar charts. For example, Hieraxes [95] employs grid-based aggregation of units to facilitate the exploration and organization of large datasets. Another way to work with multi-model network data is to use multi-facet interface[8]s. Dattolo [27] used two different views one showing the citations related to each other on a timeline and

the other showing the similarities of articles. CiteSpace II [21] used aggregated citation patterns of papers to give an overview of the network. To increase the sense-making process in user exploration for research idea generation, ThoughtFlow [42] tried to show the relations between citations in networks in different views. PivotGraph [112] used a combination of grid view with networks to show the relations between nodes; their visualization was useful in generating a hypothesis. It is a useful visualization for having an organized view as long as the number of nodes is limited. To make it easier to follow the events in multi-view graph visualization, GraphTrail [30] showed the interaction history to facilitate exploration. To make it easier to follow the relationships, PivotPaths [33] make the user follow gradual changes, which means they are not drastically changing the interface by pivoting, but the user will move gradually between intersecting structures. This system separated the three categories into three different possible views that could be connected with lines; however, different nodes in different categories were always on the screen, making it difficult to use this visualization with larger datasets as the screen will be too crowded. The advantage of methods based on faceted views is that they allow users to see information about data overview and individual data points simultaneously. The disadvantage however is that it is hard to follow information on different views and connect them.

On the other hand, in some use cases, such as when the end-user is already familiar with the overview, it can be useful to first focus on the local and detail view, because that is how the user starts exploring[60]. Considering the network tasks that are not that related to the overview, Ham et al. [106] used the degree of interest in the graph to reduce the complexity of the overview and let the user explore the sub-graph of their interest easier; using one node as a starting point and iteratively explore the graph. EdgeMaps [32] used outgoing and incoming edges to visualize relations in the graph with more detail, and revealed the category of different nodes by positioning them based on similarity. Citeology [68] is an exciting visualization that shows the references and citations of a selected paper on a timeline. However, the only connection that was focused on in this visualization was citations, which means all of the lines had the same meaning. Polyarchy is also a visualization to demonstrate relations, especially when data has hierarchical relations [84]; they showed the parents and child of the selected item in a tree format, therefore it could have many repeated representations of the node on the screen, which increase the cognitive load.

Finally, they are methods that combined the aggregated network overview and vertices on the screen to facilitate different types of exploration, such as Graphicle that merged grid view and network[62]. However, as the system kept the structure of the network the screen was still complex. MMGraph [38], on the other hand, tried opposition the entities in the parallel node-link band [70], and show the connections separately to visualize multi-model social networks. the visualization was helpful in order to lessen the clutter and complexity of the interface, but it could not work with a large number of data. Going over these systems designed for similar datasets, we noticed the interesting complexity and the amount of

information displayed in different strategies utilized in these systems, Increasing data or attributes makes the visualization visually complex, while reducing complexity results in less data shown. Also, the need for hybrid systems between the local view and global view was clear to support the sense-making process [80]. Our main design concern is to find a balance to keep the main features of the connection in our dataset on the screen.

Chapter 3

Introduction to Data and Tasks

In this chapter, we delve into the data experts’ research process and the nature of their data. Our approach to understanding this information involved engaging in meetings with data experts and exploring their own websites.

During our initial meeting with the main domain expert, who acted as the principal investigator (PI) on the WPHP project, we had in-depth discussions about various aspects of their dataset. The domain expert conveyed the different types of questions they had concerning the dataset. However, at this early stage, formulating detailed questions posed a challenge. Nonetheless, through an iterative process and close collaboration with the domain expert, we refined the visualization design based on their valuable feedback and specific requirements. Furthermore, the domain expert introduced our visualization team to the terminology used in their dataset and provided valuable insights into its characteristics. This mutual understanding of the data and research will lead us later to Design goals.

3.1 Data set

In this project, our focus was on the Women’s Print History Project (WPHP) dataset, an ongoing collection of titles printed between 1700 and 1836—a significant period in both women’s history and print history—gathered from over 100 sources. This dataset comprises more than 25,000 titles, along with their associations with individuals and firms, and other relevant metadata.

Specifically, we concentrated on three key tables within the dataset: titles (containing 25,073 rows), Persons (with 12,178 rows), and firms (including 6,934 rows). It is essential to mention that, according to WPHP’s methodology, the primary dataset revolves around titles, while also recording contributions from firms and persons. Furthermore, the records document various roles held by individuals and firms in the books. For people contributors, these roles include authors, publishers, booksellers, printers, editors, translators, engravers, introductions, illustrators, compilers, and composers. As for firm contributors, their roles may be identified as publishers, printers, and booksellers.

3.1.1 Data Challenges

After reviewing the data, we observed a substantial amount of missing or partially incomplete data within the dataset. During our discussions with a data expert, we inquired if they could fill in the missing data based on their knowledge and the patterns observed in the dataset. In response, the data expert provided several scenarios, and here I will highlight a portion of their explanation to provide a clearer understanding of their response. *“Yeah, empty is better than inaccurate, especially because the database..people look at databases and they’re like that is fact it’s like somebody has like said that this is true”*. Based on the responses received, we can infer: 1. we lack historical information and so gaps are inevitable.

2. we do not want to pretend these gaps do not exist, as they demonstrate what can be recovered historically as well as what was considered valuable information at the time.

3. we want to preserve what information is in historical records, even if it is false or misleading because they was often a reason for providing the information given.

Based on these the first challenge(C1) of this dataset is dealing with inconsistent and missing data. As we do not want to perform data wrangling and cleaning, given that data inconsistencies are one of the intriguing aspects of the dataset, we must address programming challenges that arise from data discrepancies.

The second challenge pertains to the size of the dataset (C2). With numerous visualizations now catering to big data, handling such a large dataset with multiple attributes can significantly limit the available design options.

3.2 Understanding the Domain and Task

As per our methodology, the first step in our project was to engage with Domain Experts and understand their needs. After our discussions with the individuals involved in the WPHP project, we identified two main streams for their research group:

3.2.1 Gathering and Verifying Data

The primary focus of the WPHP is to collect data from various sources and compile it to showcase the significant contributions of women to the publishing industry. Three teams are responsible for gathering information related to people, firms, and titles. After entering the records, there is a verification process in which another person, sometimes the same individual, cross-checks the information using different resources to ensure accuracy. If all the information is successfully verified using two sources, the record is marked as verified. However, if complete information cannot be found, the record is marked as attempted verified. Ensuring data quality is a priority for the WPHP team. They strive to have the least missing data fields with accurate information but avoid entering uncertain or potentially incorrect data.

It's worth noting that the dataset initially focused on records from 1750 and onwards, so data from those years is more complete. In cases where information on a title's first page is insufficient, researchers use other reprinted versions or additional contributions by the same person to complete the details. To ensure accuracy, connections between different variables are used to verify the information. The WPHP team aims to improve the verification process by gathering more related information, particularly about contributors and contributions. They are also developing a new table to document relationships between firms. For example, they will record if two firms are related.

3.2.2 Writing Spotlights and Making Podcasts

The WPHP project also emphasizes communication of their findings through podcasts and spotlights. Spotlights are informative pieces in text format, ranging between blog posts and essays. Researchers use spotlights and podcasts to convey information that may not be captured in the dataset alone. These resources shed light on specific facts, historical patterns, or individual contributors. They usually decide their spotlight based on the patterns or information that inspire them while gathering the dataset. During the conversation with the PI, it was also clear that they are interested in the connections between entities such as who worked at which firm on what book. Which firms or people contributed to the same book?

3.2.3 Data Changes

The fact that the WPHP is an ongoing project suggests that it will continue to grow over time. As part of this expansion, new tables and fields will be added to their existing data. This dynamic nature of the project underscores the importance of designing a flexible and adaptable visualization tool that can accommodate future updates and changes in the dataset.

3.3 Design Goals

Considering the research needs of the data experts, it becomes evident that finding spotlight topics and verifying data are exploration tasks, where experts need to analyze different titles and individuals and explore their relationships to identify patterns and raise questions. Our main goal was to when possible, cater to the specific needs of the WPHP team. Based on that we have defined our design goals for this project:

DG1 To create an explorable visualization interface,

DG2 To, as much as possible, visualize all the data on the screen, examining contributions made over time,

DG3 To handle missing, partial, and anomalous data,

DG4 To have an adaptable visualization based on changes in database structure, and

DG5 To make full details of data accessible.

Therefore, our main focus for this project is to create an explorable visualization interface (DG1). Our intention is that as much as possible, this visualization interface will present all the data, including missing information, on the screen (DG2). Users should be able to follow connections between titles and contributors (both individuals and firms) and examine contributions made over time, comparing different roles played by individuals and firms.

To preserve the anomalies in the dataset that could provide valuable insights, we have chosen not to alter the records that lack consistency (DG3). The interface should be user-friendly, enabling researchers to navigate through the data effortlessly. Additionally, the visualization must be adaptable to accommodate potential variations in datasets, such as the appearance of new tables, and adding new records (DG4).

A key aspect of the design is the inclusion of missing information, providing a comprehensive view of the dataset. The visualization should not clean or impute the data, as this approach will allow researchers to assess data quality and identify potential gaps in the dataset.

To promote effective exploration and sense-making, the visualization should enable users to follow connections between different entities, such as titles, persons, and firms, across various time frames. This feature will support researchers in identifying trends, relationships, and patterns in the publishing industry.

Finally, allowing data experts to access the raw data is crucial (DG5) because they require the ability to check the records and texts directly in order to interpret the content accurately. By having access to the raw data, researchers can verify information, cross-reference details, and ensure the reliability of their findings derived from the visualization tool. This direct access empowers data experts to conduct in-depth analyses and contextualize the visual representations effectively, ultimately enhancing the overall utility and credibility of the visualization tool.

Chapter 4

Design Process

In this chapter, we will focus on our design process regarding the design goals and also explain the interface in detail.

4.1 Iterative Design Process

As I have used an iterative design process in this project, I am going to start by brief giving background about the iterative design process. The iterative design process is a method for solving problems that entails repeating a series of steps while improving a product or solution at each stage based on customer feedback and knowledge gained from earlier iterations. To develop efficient and user-centered solutions, this process is widely used in a variety of industries, including software design, visualization, and human-computer interaction (HCI). I will briefly explain the iterative design process in these three fields.

One of the first appearances of the iterative design process is in software engineering. Using iterative design in software engineering goes back to the mid-1950s. In 1970, the Waterfall model was one of the early models that formalized the iterative process, mentioning the results of the testing phase can update the first design. For more information on the progress of iterative design in software engineering before the agile model, you can check the Larmen and Basili paper in 2003 [54]. In 2001, the Agile Manifesto was formulated by a group of software practitioners, highlighting values like customer collaboration, responding to change, and iterative development. Different models appeared that are agile, among which Scrum is one of the famous ones because of its adaptability with different processes [67]. One of the benefits iterative design process in software engineering is its adaptability to requirement changes [61]. The main goal of using iterative design in software engineering is to get a better understanding of the requirements to have functional software.

Iterative design in Human-Computer Interaction (HCI) is a user-centered approach that involves creating and refining technology interfaces through a cycle of continuous feedback and improvement. It begins with thorough user research to understand their behaviors, needs, and preferences. Designers then create initial interface prototypes that align with user

requirements. These prototypes are tested with real users, who provide valuable feedback on usability and functionality. This feedback is used to refine the design, making adjustments to the layout, interactions, and user flows. Subsequent iterations involve further testing and refinement, ensuring that the interface becomes increasingly intuitive, efficient, and enjoyable for users. The iterative process in HCI aims to create technology interfaces that seamlessly integrate with users' cognitive processes and behaviors, resulting in a final design that optimally supports user tasks and experiences [7, 76].

In the visualization community, the goal of this iterative design process is to make the data representation more understandable and informative. Iterative design in Visualization begins with a deep understanding of the data's context and the objectives of the visualization. Designers create initial visual prototypes that aim to capture the essence of the data's patterns and relationships. Stakeholders and potential users review these prototypes, providing feedback on the clarity, accuracy, and visual effectiveness of the representation. Based on this feedback, designers refine visual elements, adjust color schemes, and optimize layouts. Iterations continue with multiple rounds of user testing and refinement to ensure that the visualization not only presents the data accurately but also facilitates users' understanding and decision-making processes [74].

In this project, we are working with data experts and we decided to use the iterative design process to have a better understanding of the data experts' tasks and requirements and also to have their feedback on the design at each stage. Based on the IEEE Visualization paper on design study methodology [90], working with domain experts is a highly iterative process. We continued using an iterative design process after fixing the requirements, to increase the functionality usability, and understandability of the visualization.

4.2 Design Methodology

As data experts and their unique problems are a major part of the design studies [90], we used different methods to involve the data experts' opinions in our design process the most. Considering the design goals extracted from the data experts' tasks that we discussed in the previous chapter, We decided to have a User-centered design process [88, 41]. Based on User-Centered Design (UCD) [64], we used a multidisciplinary design approach based on the active involvement of domain experts to improve our understanding of user and task requirements. We used an iterative design approach and subsequently ran a qualitative study to better understand our tool and its potential. Throughout our process to gain user feedback we had several meetings with data experts to get their feedback about the design decisions and to keep the prototype aligned with the data expert tasks. Especially since we were trying to create a tool that supports exploration, we needed data experts to see the interface at different stages and talk about their possible new exploring questions in relation to the design. The initial design process, after the meetings to gather the tasks and

understand their data better involved brainstorming, sketching, and reviewing existing work in the domain. Based on the information provided, we made a sample dataset and used the 10+10 design sketching approach [18] to have some initial sketches. In the next step, we had a sketching meeting with the two data experts, in which we first asked everyone to do some quick sketches [73, 88]. Then, we showed the early sketches and discussed them in a group.

4.3 Early Design Choices

After the session, a follow-up design meeting was conducted with my supervisor to discuss potential design choices and ideas. After examining the sketches, we identified two main features in the dataset that could serve as the foundation for our design: time and location. However, we found that the time data offered richer and more diverse information, and it was closely related to the historical context, which formed the basis for most of the domain experts' inquiries. As a result, we decided to prioritize time as the primary aspect of our visualization.

To assess the effectiveness of visual exploratory tools for data experts, we developed an early prototype using a sample dataset. This prototype included zoom interaction to allow for deeper exploration of the data. The data points were positioned along the timeline, with those having partially missing data placed at the most detailed data point available, and those with entirely missing data positioned at the beginning of the timeline (DG2). Color coding was employed to facilitate the identification of missing data and to provide additional information, such as the role of individuals or the genre of the book.

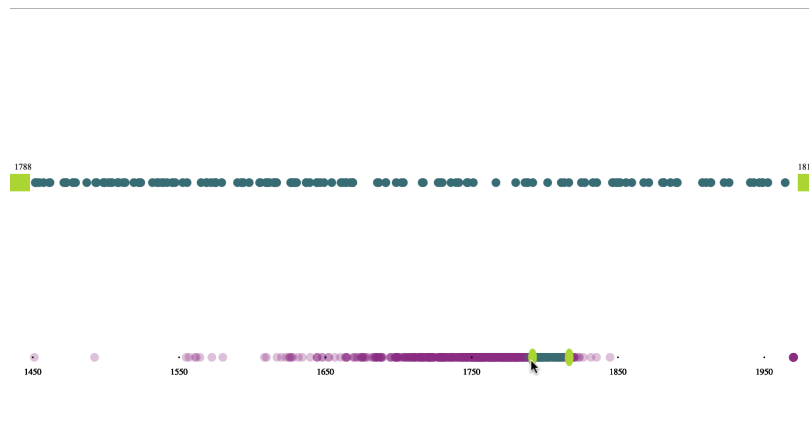


Figure 4.1: In the initial designs, we attempted to stretch timelines for zooming purposes; however, we observed that numerous data points coincided at the same positions, and stretching the timeline did not alter this overlapping issue.

Based on the domain expert's feedback, the prototype was going to have three different timelines according to firms and the persons and books. Having three timelines gives us the

opportunity to show the connections in the database. However, after getting the complete data in this stage, we noticed that the dates are mostly just years, therefore, on timelines, most of the data was packed, and we had to change the design and interaction to fit the data. The prototype helped identify specific questions that experts may ask using the exploration tool. Some of these questions aligned with the research questions discussed in initial sessions, while others provided additional insight into the tool. Also, it helped data experts to have a better idea of what visualization can provide them, and bring the team closer in terms of mutual language.

Considering the challenges of the number of items in each year and the need to show missing data on the screen. That caused the issue of having many points in the same location. We decided to use abstraction showing all of the data in one year in a rectangle and using the length to represent the number of items in the year. We kept time as the main component in the new design, and we used three timelines, each with the same name, for the three different main entities in the dataset, so that we could also show the connections between them to answer the relationship questions (having three timelines is also based on the nature of the data). We decided to have a vertical timeline instead of a horizontal one, as the expansion of the timeline will increase the length to the extent that it will exceed the screen even if it was horizontal, and make scrolling necessary for users. In terms of horizontal and vertical scrolling for users, experiments show that they prefer vertical scrolling [109]. The missing data that we could not place on the timeline, was not removed from the data. Instead, we placed them on top of the timelines, which can represent the importance of missing items in this dataset (DG2).

4.4 Final Design Description

In general, the prototype consists of two main sections (Fig 4.2). On the left side, there is a timeline, while the right side contains three distinct types of filters and a designated area where the information box will appear. Let's delve into each of these components in greater detail.

4.4.1 Timeline and Information Box

In this section, I will elaborate on the design and interactions concerning the main timelines and the information box. I will also provide justification for how these designs align with my established goals. Explaining the information box in this context is important, as it emerges as a result of interactions with the timeline.

The initial design step involves a vertical timeline on the left side of the screen, which essentially spans the entire screen. Within this timeline, we have three distinct timelines positioned side by side, each representing the three main types of entities present in the

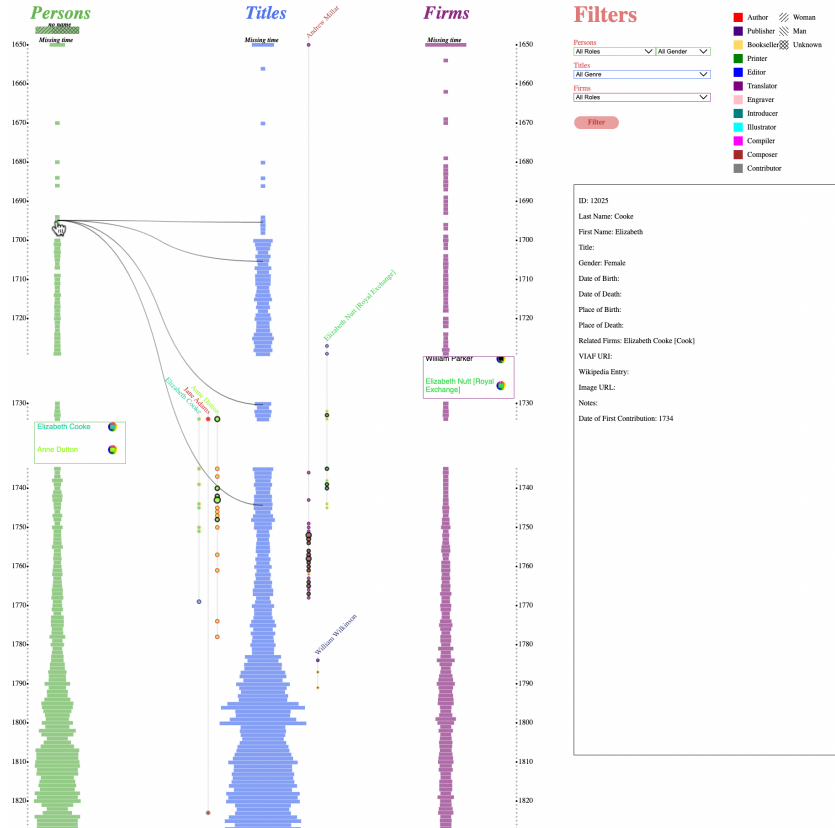


Figure 4.2: Overview of the prototype

database. These three entities were chosen based on the underlying structure of the dataset, which includes people, titles, and firms that contribute to these titles.

In our design, we positioned the entities on the timelines based on specific criteria. For people, we placed them according to their first contribution, which we calculated as the first time they appeared in the database. For firms, their placement is determined by their start date (established date), and for publications, it is based on the date of that specific version of the book being published. Notably, a book may appear multiple times on the timeline if it was reprinted at a later date. This decision was made to cater to the interests of data experts who find the reprinting process intriguing.

The data positions on the timelines were chosen after discussions with data experts, ensuring that the dates chosen truly represent events related to publication in that historical era. Additionally, for the personal timeline, calculating the first available contribution helps mitigate missing data (Challenge 1) and provides an overview of how many people started their work each year. I have calculated the starting date of persons by checking for their first contribution in the dataset and adding a field to the persons table based on their first contribution that exist in the dataset.

Due to the considerable number of records for each entity in each year, displaying them all as separate objects on the screen was impractical. Moreover, we often only have the year available as the date, without exact data. Consequently, many objects end up having the same position on the timeline, making it challenging to utilize zoom-based interactions for detailed examination. To address this, we followed the design principle of overview context [92], abstracting the data into rectangles in different years. The size of each rectangle serves as a visual variable representing the number of items in that year, facilitating users in gaining an overview of historical occurrences.

However, some titles have missing time information (Challenge 1), preventing them from being placed directly on the timeline. To handle this, we created a separate rectangle at the top of the screen to represent missing time data and ensure consistency in interactions. Additionally, we encountered another type of missing data, referred to as "Unknown people." These are title contributors for whom we lack information. In some cases, we may have information about their gender. Consequently, we categorized these unknown individuals into three groups: unknown gender, male, and female, each represented by a distinct pattern on top of the person's timeline (DG2). We utilized three different colors across the timelines to enable easy distinction and highlight relevant information in the information box (when containing title-related data). To ensure continuity during scrolling, header titles float at the top and bottom of the timelines, remaining visible for users even as they scroll.

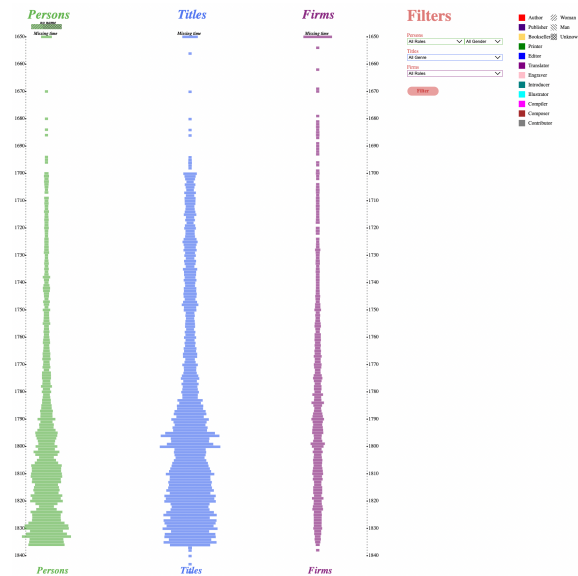


Figure 4.3: Overview of the prototype

Timeline Interactions

Continuing with the "detail on demand" concept, the initial interaction involves hovering over the rectangles (Fig 4.4). By simply placing the pointer over a rectangle, users can

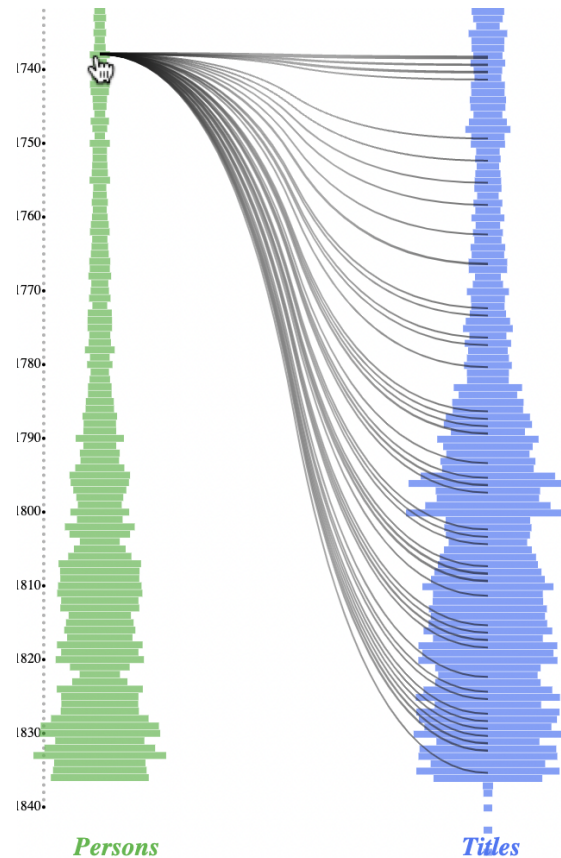


Figure 4.4: Hovering on a rectangle in different years reveals connections between two timelines in the same year. This figure is about persons and title timelines demonstrating the years in which individuals initiated their involvement in publications and the subsequent years of their contributions.

observe the connections between the three timelines. Connecting lines are utilized to offer an overview of the relationships between different years in terms of various events, aggregated together. Some of the questions that overviews may be able to answer include:

1. What is the distribution of publication from when persons start their work in the same year?
2. For publications in one year, what is the start working year of persons and firm contributors?
3. When usually firms that have started in a specific year publish a book?

For making the connecting lines easier to follow, we used curves with two points on their path. This decision was made after trying different kinds of lines, and the two bending points let the endings of the curve be almost straight and easier to follow.

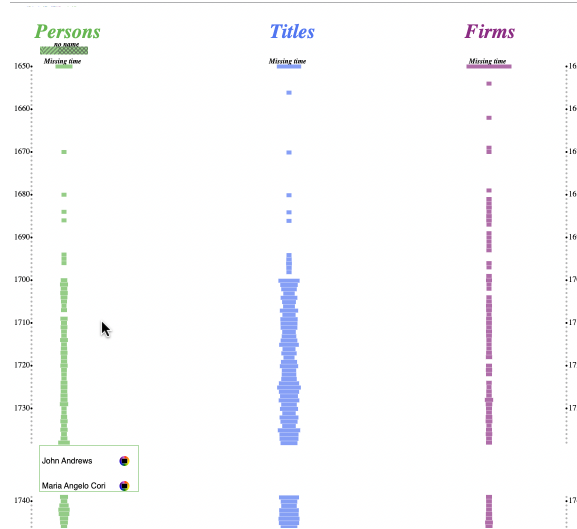


Figure 4.5: clicking on lists will show the items as a scrollable list, and will push all of the items down

To delve deeper and access more detailed information, users can click on each rectangle, which provides a comprehensive and scrollable list of items within a specific year (Fig 4.5). However, the scrollable list and interactions differ between the firm-person timeline and the title timeline.

Person and firm timelines follow a logic based on the titles within the database, as titles establish the relationship between persons and firms. Consequently, the positions of these timelines are fixed, and their interactions are similar in both the person and firm timelines.

The list of Firm and Person timelines displays the names of persons and firms, alongside a color picker tool whose specific usage will be discussed later. The names in both lists are arranged alphabetically. Hovering over each name reveals the connections to different years in which the specific firm or person has contributed to any title (Fig 4.6). The connecting lines share similar characteristics as the lines between timelines; they appear thicker and darker when there is a higher level of contribution from the person or firm in a particular year, and they consist of paths with two points.

Clicking on a name allows users to access further details of the firm or person (Fig 4.7). In such cases, the publication timeline of the clicked item appears close to the title timeline, and an information box appears on the right, containing details about the selected person or time.

The people/firm timeline is represented by circles with the corresponding name tag of the person or firm on top of the timeline. The circles have a middle color that is consistently similar to the name tag's color, while the border color can vary based on the role that the firm/person had in a particular book (color coding will be explained). If the border color is black and the size of the circle is larger, it indicates that the person/firm contributed

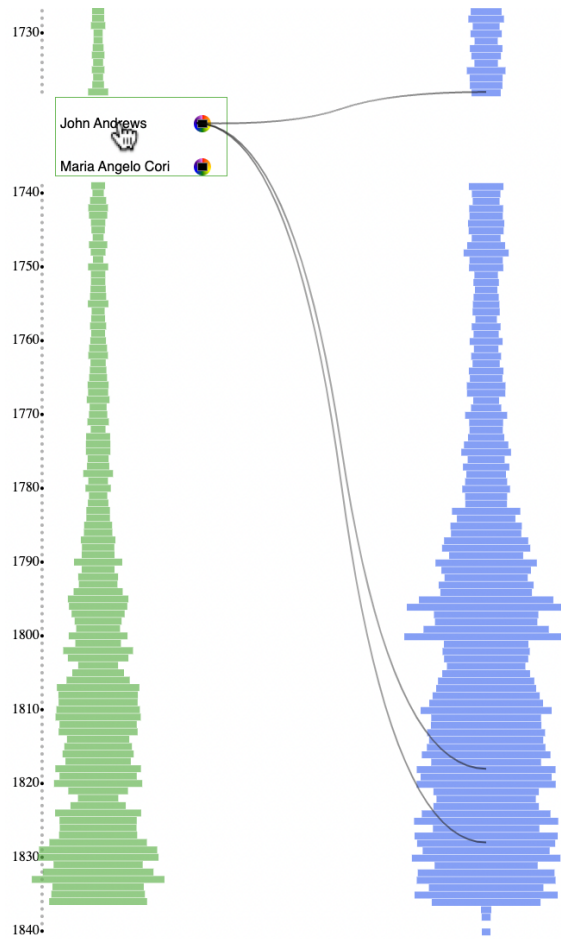


Figure 4.6: Hovering on the names in the scroll-able will show the links to the person’s publications

to more than one publication in that year (Fig 4.8). When clicking on circles with black borders, a spiral of the contributed titles appears as a circle with the same color as the main timeline, with the center of the circle corresponding to the clicked item on the main timeline.

Additionally, when a contribution(circle)is on the screen via the timeline of a firm or person, that contribution title will be highlighted in the title timeline using the same color as the timeline. The main color of the timeline can be changed using the color picker tool beside the name of the item. This feature proves useful when users want to open multiple people/firms simultaneously on the screen, as they can assign different colors to easily distinguish them on the title timeline(DG1).

As more timelines are opened, they position themselves closest to the title timeline, filling the space between the three timelines and moving closer to the person/firm timeline, depending on the type (firm or person) of the timeline (Fig 4.9).

To make intersecting items more distinctive, the prototype allows zooming, and the color of the items has an opacity of less than 1. On the screen, there are at most two intersecting items to ensure clarity and readability.

There are two ways to close a timeline: the first is to click again on the name of the firm/person in the list, and the other one is to click and hold on the timeline's name tag and then click on the trash can icon that will appear beside it (Fig 4.10). When a timeline is closed, the color of the title's name in the title timeline will return to the default (black) color, and the other timelines will move closer to the middle timeline. When the user opens a timeline, the color of the title in the title timeline will change, there are two possibilities. If the scrollable list of that year is open, it will auto-scroll to the last book of that year to which the firm/person corresponding to that timeline contributed. Otherwise, if the year was closed and there was no list open, the user can open the list, and it will similarly auto-scroll to the last colored book. This auto-scroll feature is added to the prototype as the list can be extensive, making it easier to find specific items. Despite the ability to find the item in the scrollable list for each year, the circle representation of titles in the timelines is also clickable. Clicking on any of the circles will trigger the appearance of the information box with related details (DG5). Simultaneously, black triangles will appear beside the persons/firms' timelines, pointing to the years in which the firm/person contributed to that particular title (Fig 4.12). Using those triangles, the user can open the specific year in the persons/firms timeline and get the name of the firm/person highlighted in bold (the list will also scroll there itself) (Fig 4.11). The titles timeline in the middle is similar in terms of being able to access a scrollable list by clicking on the rectangle. However, there is no color picker tool beside the title's name. If you hover over each title, the connecting lines will appear, showing the different years in which the contributors of that book exist on the side (firms/persons) timeline. Additionally, if you click on the title name, the information box will appear on the side, and the triangle points will continue showing the location of the firm and personal contributor on the timeline for further exploration (DG1).

Updating the Timeline's Time scale Opening and Closing a Year

To open a rectangle in any of the three timelines, the rectangles below it will move down accordingly, ensuring that all the data remains on the screen. Additionally, the entire timeline scale will adjust, meaning that if a timeline for a person/firm is open on the screen from the year 1650 to 1655 and the user opens the year 1653, the distance on this timeline between these two contributions will increase. Closing a timeline will restore everything to its previous state, maintaining a consistent timescale throughout the screen.

4.4.2 Filters

Considering the data experts' requirement to identify patterns from specific data categories, we have incorporated filters for the timelines (Fig 4.13). These filters enable them to gain

insights from various subcategories, providing a broader scope to answer different questions. It is also worth mentioning that we decided on the filters at this stage based on our discussion with PI. Filters:

For persons: Gender and Role filters.

For firms: Filters based on the role of the firm.

For titles: Filters based on the genre of the book.

4.5 Implementation

This prototype was developed using Python, HTML, CSS, and JavaScript(JS). Given the large size of the dataset and the need for quick response times in our interactions, we adopted a design approach that separated the logic and database from the interface. To store the data, we utilized MariaDB SQL server. The backend was implemented using Django, a powerful Python web development library. Most interactions involve sending requests to the backend, thereby avoiding complex calculations at the interface level. For creating the interface, we utilized JavaScript and D3[14], a visualization library based on JS that directly models the data and is known for its capabilities in data visualization.

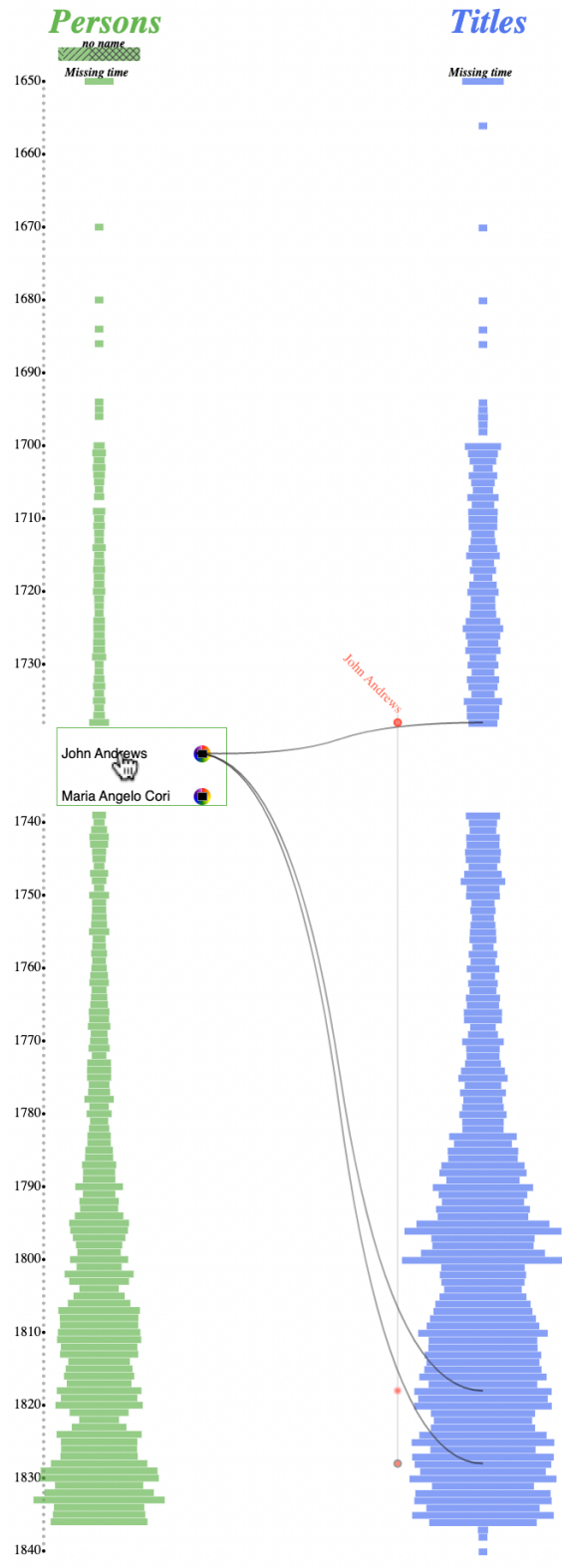


Figure 4.7: Clicking on the names in the scroll-able list, a timeline for one person will appear with circles on the timeline representing the publications

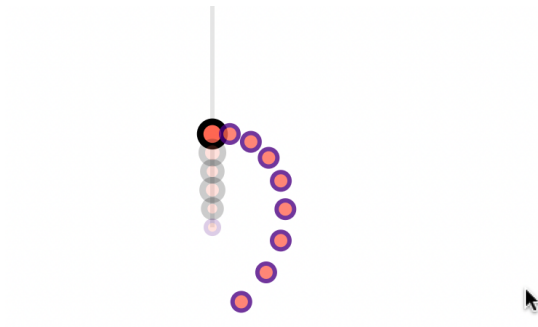


Figure 4.8: Clicking on a circle with a black border on the timelines, a spiral will appear with the publications inside that circle around it

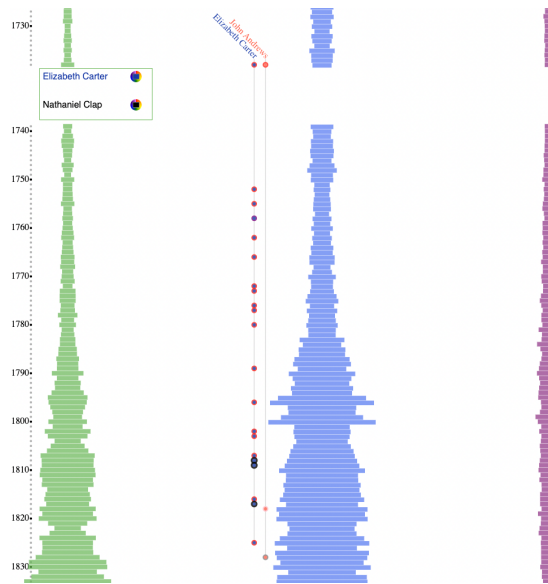


Figure 4.9: It is possible to change the color of the timeline using a color picker. In this image, the red timeline opened before the blue one because of that it is closer to the title timeline

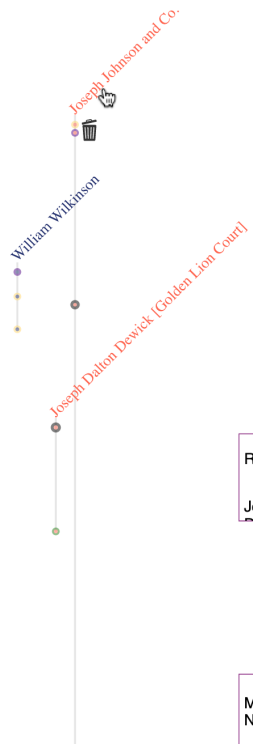


Figure 4.10: After a long click on the timeline name-tag the trashcan button will appear that will delete the timeline

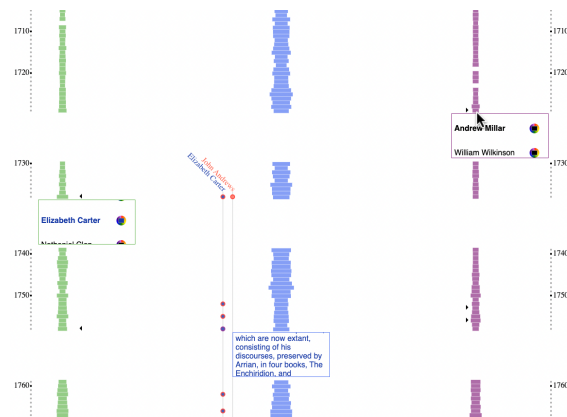


Figure 4.11: If there is a circle representing a firm or person on the screen the name of the person in the scrollable list will be bold

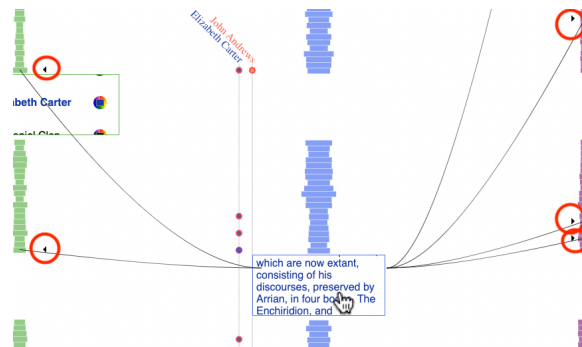


Figure 4.12: Clicking in the titles, the triangles will appear to help the user find the related firm and person

Filters

Persons

All Roles All Gender

Titles

All Genre

Firms

All Roles

Filter

Figure 4.13: Filters

Chapter 5

Qualitative Study and Results

I now describe my qualitative study, explaining the procedure and presenting the results. I decided to do a qualitative design study to get feedback about the design and interaction, as well as to get some understanding of the effectiveness of this tool for the data experts' research process. I was lucky enough to have 6 domain experts take part in this study. Simultaneously I decided to also study the system with non-data experts to learn more about the usability of the tool. For evaluating design studies [51][96] there are different methods among which case study and usage study are the most used [114].

5.1 Study methodology

This study was conducted with 12 participants in total (Domain experts: 6, non-domain experts: 6). The 6 domain experts were invited to participate in the study (1 Male and 5 Female), including four graduate students and 2 undergraduate students from our university's Department of English. All of them had experience working with this dataset in different ways, and their experience ranged from 6 months to 5 years. Also, 6 people who were not domain experts were invited, using email and word of mouth, to offer opinions about the user experience and about their attempts to answer the domain questions in order to get to know the tool. Non-data experts were from diverse backgrounds, including visualization, archaeology, cognitive science, psychology, English, and environmental science. Their education level was also varied (1 post-doc, 1 Ph.D., 4 Undergraduate).

The experiments were conducted in ixLab (Interactive Experiences Lab), using MacBook Pro with a 16-inch display. The interface was open in Chrome browser (we decided to run the study on the laptop as most of the data experts use their laptop to do research, and it has smaller screen; thus the experiment will be more realistic). Prior to each experiment, participants were asked to sign the consent form and fill in the questionnaire (A.0.1). The experimenter provided a short explanation of the experimental procedure and the tool. Then, participants were then instructed to complete the five tasks outlined in Section 5.1.1. Throughout the tasks, participants were encouraged to verbalize their thoughts [99], and the

experimenter engaged them in dialogue by asking questions. They could also ask questions and get help while doing the tasks, as the main purpose of the tasks was to engage them with the tool and teach them about interactions. After finishing the tasks, each participant could freely use the tool, generate their own questions and try to find their answers. This part of the study was proposed to observe what are the tasks that participants would like to perform, and how they use the features to perform them. After this part, each participant underwent a fifteen-minute semi-structured interview about their experience with the experiment and the tool, and they were asked to answer a questionnaire (A.0.2). Finally, experimentees were asked to sketch their ideas on paper, if they wanted to[83]. The experiments were recorded in both audio and screen formats, and the average duration of each session was approximately one hour.

5.1.1 Tasks (1)

In the first part of the study, I gave the participants 7 questions to answer that needed different interactions and aspects of the tool to answer them. Participants could ask the experimenter to re-explain different features. Some of the tasks had a repetition of using some interactions. As the visualization had some unfamiliar designs, I wanted to see how different they will perform the similar interaction after performing it once. The tasks were designed to instruct various interactions on the visualization.

Here are the tasks:

1. Find a firm that published for more than 20 years.
2. Find a person with more than 10 pubs.
3. Find an author in the year 1780 who published in genre fiction.
4. Find an author who was active for more than 7 years.
5. Find how many persons were authors in 1750 who had more than one publication.
6. Find a person in 1760 who had participated in publications with more than one role.
How many publications did the person contribute to in any role across her career?
7. Which firms did she work within her second publication?

5.1.2 Tasks (2)

In the second task, they could use the tool freely based on their own desires. In this task, we asked data experts to focus on utilizing the tool for different scenarios that they encountered during their work with the data. This way we could both check the effectiveness of the tool as we familiarize ourselves with the tasks they are performing in more detail. Considering this task for non-data experts, we wanted to see if they could explore the data using this

tool and see how this tool can give them questions to explore the dataset that they have not worked with previously.

5.1.3 Semi-structured Interview

In the interview, I started by asking some pre-prepared questions. However, the conversation was followed up by questions based on participants' answers. The main set of questions were as follows:

1. Has the tool led to any unexpected or remarkable findings?
2. Was there something you would have liked to be able to do?
3. What were the challenges you faced working with this tool to explore the data?
4. Do you have suggestions for changes? Additions? New parts to the Vis new parts to the interactions? Could you sketch your ideas?
5. I am curious about your thoughts on this visualization and this tool.

5.2 Data Analysis

To analyze the results of our study I listened to the audio/video records of each interview and transcribed them. Then, I carefully reviewed and sorted the written records, identifying important ideas by doing thematic analysis [16] and putting them into categories. I also referred to the notes I took during the study, and tried to give context to the findings.

5.3 Results

Based on the post-study questionnaire, all participants found the system useful in identifying time-related patterns and easy to use for these tasks. The experts, with the exception of one neutral response, found the system helpful in discovering relationships between persons and firms. Similarly, the non-data experts, except for two neutral responses, also noted that the system was valuable in identifying relationships between persons, firms, and titles.

All participants, with the exception of one neutral response, indicated that they were able to obtain the desired information from the dataset using the tool. Moreover, all data experts expressed a strong agreement that they would utilize this system to explore the dataset further and mentioned their intention to recommend the tool to other researchers on their team. Furthermore, among non-data experts, all participants stated that they would use this tool with similar interactions on other datasets. Some data experts even expressed their enthusiasm for future use, as exemplified by P12 who said, *"This is so neat. I could play around with this for hours."* P2 also mentioned *"I thought it was really fun to use. I think it would be when it's like I don't think it's still in progress, but I think it will be*

really useful, for, like other researchers working on their projects.”. Some of the experts were impressed by what the visualization can do; for example, P9 verbalized her feelings by saying *“First of all I want to say that, we’ve been talking about visualization for so long, and I had no idea it could be so informative and good.”*.

5.3.1 How did Participants encounter tasks?

I have defined the tasks such that they did not have an exact answer. They could randomly try to find the answers. However, participants generally did not encounter any major difficulty, except for the last task. For T1 and T2 participants needed to first hover over the firm timeline and used the line to estimate the number of years a firm was active or the number of publications a person had. Most of the participants did this task without any problem. I noticed that based on the first year they hover over, they see different line patterns, and they can finish this task at different times. During these tasks, I noticed that they were understanding the general features of visualization.

For T3, almost all of the participants understood that they could use the filters really fast, and then they opened the year to check the names in the list after the filter. Some of them filtered for the role (Author), and Genre of the title (fiction) in the first place. Others, only filtered for the genre and they later needed to check the rounding of the circle for the person role. Some of the participants also checked the information box in this task to double-check if they were correct.

For T4, almost all of them were now comfortable working with timelines and opening and closing the boxes, and using the lines. Some of the data experts, such as P11 and P12, opened one list, and since they know the people there, because of prior knowledge. Others opened the timeline and checked for the circles on the timeline and found people randomly.

For T5, participants needed to open the year and they had to either check the lines or timelines to understand the number of people with more than one related title. Most participants just opened the timelines and counted the ones with the property after opening all of them.

Tasks T6 and T7 were related and they were the most challenging tasks, the participants needed to find a person, anyone, who based on circle rounding had contributed to more than one role and count all of the person’s contributions, and then they had to find the firm contributor of the person’s second publication, they had to either use the triangles that would appear after clicking on the person, or open the list where the title would be highlighted and follow the lines to the firm timeline, and open it, where they could find the firm in bold. Participants, mostly, did not have a problem finding the firm name in the firm name based on the information in the box. However, most of them (8/ 12) did not remember the triangles and needed help to find them and to find the firms in the lists.

5.3.2 Overview helps in different ways

During the initial exploration phase, we requested participants to attempt to answer their own questions, which proved to be challenging, particularly for non-data experts and data experts with limited familiarity with the dataset. The difficulty lay in knowing where to start and which specific aspects to investigate. However, we observed that the timeline overview played a crucial role in assisting individuals who were less familiar with data and potential tasks, as it provided a helpful starting point for their explorations. For instance, P8, with an English background, shared, *“I am very curious about these really kind of prolific years.”*. In another use case, most used filters on the overview several times to compare the overview for different roles.

On the other hand, there were some data experts who were also doing research on WPHP, who tended to begin by formulating their own specific questions. They initiated their exploration with a focus on detail, using filters and specific criteria to narrow down the dataset. Subsequently, the overview visualization aided them in contextualizing their findings and allowed them to continue exploring other parts of the dataset. For instance, P12, a data expert, started by filtering for publishers known to the researcher and then sought to investigate further, seeking insights from the visualization tool. Besides data experts with their own research questions, there were some casual users that later in their exploration started to look for people in the publication industry that they know, for example, Shakespeare and Jane Austen.

Based on the need to find a specific starting point for some questions, when the participants were going to give up, I gave them the year that was the first publication of person, and they could continue from that point. Therefore my conjecture is that for the scenarios that the participant wants to start from a person, it might be useful to have a search function that will provide the location of individual person on the timeline. This means, that the user could just type the name of the person as input, and search engines could show the year needed to find the person.

5.3.3 Learning Curve

Our main goal was to create interactions that effectively conveyed data and allowed the end-user to explore it, even if the design was unfamiliar to our participants. Based on participants' responses to the questionnaires, Most of them (9/12) have little to no experience working with visualizations. As a result, during the initial task (5.1.1), many participants had questions about the interactions. However, as they engaged with the system and completed the tasks, they became more comfortable using it to find the answers they were seeking. The participants verbalized their learning experiences and acknowledged the learning curve associated with the new interactions and design. Despite the initial challenges, they were able to adapt and successfully utilize the system as they progressed with the

tasks. For example, P1 said, *“So once a person is kind of immersed into this sets, then it’s a lot easier for them to understand.”*

There were some suggestions to mitigate the learning curve, for example, scenario-based handbooks given to users to find out how they can interact with the system. P9 suggested, *“So I’m just thinking if there is, anyway, or if there will be, I don’t know what the process is, but maybe having like a guideline or something that says so... for example, if you want to compare the number of books published by unknown women to books published by unknown men, you could do this.”* [P5, Visualization Expert] also suggested having a tooltip that will be like the games, this will appear in the tool for first-time users and they will keep coming up until the user indicates they don’t want to see them.

5.3.4 Data Experts See More in the Data

As expected, data experts utilize their knowledge to find answers more quickly. We observed that during these tasks, data experts leverage their understanding of the data to optimize their starting points. For instance, in T2, P1, a data expert, began in the year 1750 because they knew it was a prolific period for the print industry. Similarly, in T4, they sometimes focused on individuals they were already familiar with and knew to have many contributions. P11, used the main title, instead of the information box, to find the firm contributor because she was aware that titles contain the firm names sometimes.

Data experts also count on their own knowledge to interpret data beyond taking it as a fact [15]. Non-data experts tend to accept visuals and relations as straightforward facts about the data. Data experts, on the other hand, approach the same visuals more skeptically, questioning the underlying information. For example, when looking at the overview showing a concentration of publications in later years, non-data experts may interpret it as a flourishing era for print. However, data experts consider other possibilities, such as the likelihood of focusing more on publications from that specific time period. P11 expressed this uncertainty, mentioning the need to examine whether the data is representative of the actual publishing industry during that time or if it was influenced by the data imported.

Besides interpreting the overview differently, data experts have uncertainty about the way data is inputted into the database. During the task completion, data experts frequently used the information box to double-check the title information to address their uncertainty. Based on the WPHP approach of saving the titles as they are, they can check if the fields have been filled properly. This step provided them with an added layer of confidence in their results.

5.3.5 Data Discrepancies Prompt Exploration

One of the goals of having an explorable interface was to identify data discrepancies and mistakes during the data exploration process. Although it was possible for all of the participants to see patterns, data experts’ knowledge plays a significant role in recognizing

inconsistencies in the data (section 5.3.4), as they have more knowledge about the data gathering process and the structure it should have [59]. For instance, P2, a data expert, mentioned, *“It’s funny I keep opening titles and seeing like, oh, this needs to be fixed.”* Or P11, a different data expert, noticed that the person was a woman when she opened a record from missing data with an unknown gender timeline.

We created a specific interaction that allowed data experts to see the timeline of unknown people, separated based on if their gender is known or not. They used this feature to identify patterns of missing data. P1 mentioned, *“When I clicked on missing persons, it showed me that many of the titles, for example, from the later eighteenth century, and that’s really useful to know because then that could be a place where we might start to continue my recovery work rather than starting all the way from like the seventeenth century. Right?”*

During the study, we noticed a particular issue in the dataset related to firms with linked contributions. There was the discrepancy in firm start dates, being later than the firm’s first recorded publication. Before commencing the study, we were already aware of this problem and decided to investigate further. We wanted to understand how data experts would react to this issue and whether they would notice it. To answer this, we observed how many participants noticed the problem on their own. Additionally, we engaged in conversations with data experts who did not notice the issue themselves. We showed them the specific part in question and observed how they tried to figure out the reason behind it.

When discussing this issue with data experts, their first reaction was to check the information box to check if the record is actually starting the fact, or if it is a programming error. They checked the firm-related data by opening the records, clicking on the firm’s name, and then examining the first publication start date. Based on their knowledge and understanding, different data experts had varying insights. P11 first thought it could be like this because of the firms moving, and For instance, P13, who had extensive experience in the project, initially considered the possibility of reprints and moving firms. However, upon further inspection, she realized that the firm name in the imprint did not match the firm assigned to the title, indicating a possible issue of multiple records for the same firm with different addresses. She explained the potential problem, stating, *“So somebody who was probably working on this and they said, oh, James Wallace is in the imprint... Okay, and they typed in James Wallace and they just picked which everyone came up first. Or if the imprint said that it was printed at 46 pattern, even though he was no longer at the address.....You know what I mean? Then it’s like. If the title page says it was that address, but we know he moved.....Which one do we attach it to? Right. Does that make sense”*. This is showing that it is an error with the firm that has been attached to the titles.

Overall, this system proved helpful in identifying data errors, some of the data experts mentioned that the way lines were used made it easy to spot such issues. P4 stated, *“we noticed that there’s clearly like problems with the firm data.”*, and continued, *“the way that you show firms and titles over time makes that insight really easily easy to see and also easy*

to find because I can just scroll along firms and everywhere the line is kind of curving up expect missing title.”

5.3.6 Data Experts Encountering Things They Usually do not See

All of the data experts mentioned that it is interesting for them to see the data in this format. Hanging around the interface they learn about things that they usually do not see. For example, P1 mentioned *“I guess I didn’t really know who Draper was, which was nice, because I could see her name on the page... So in the sense when I’m playing around with it, I’m picking on random names that I might not otherwise encounter.”*, and P2 opened a timeline from a firm that had two different roles in its career, and explained, *“So it’s nice to have that visualization to have everything together so often when I’m entering data, I’m not taking the time to scroll through someone’s entire complete history of things they worked on, and I’m not seeing the fact that they did other things other than printing.”* P9 also pointed they usually do not visually interact with data and that is different from their daily tasks which makes it interesting for them to use the tool. On the overview level, extreme dates of data also grabbed their attention; for example, P11 got interested in the latest publications and said, *“These ones go all the way to 1840, which is interesting.”* Then she tried to have some explanation for the data points that are really late in the dataset. For example when they are writing the spotlights.

The other part of the view that was giving data experts a new perspective on their data was the possibility of opening several timelines at the same time. We noticed that the possibility of having timelines of different people open beside each other is helpful for data experts’ research; for example, P2, after pointing out that it is something that she could not see without using this visualization, mentioned, *“It was really interesting to get to see which, especially women publishers because we focus on women, which ones were working in close proximity, time-wise, publishing at the same time.”* or P12 opened two timelines for firms and continued, *“Oh interesting, cause he was more spread out over more years. And like less per year, whereas this guy, Thomas H. was like, published a ton in like 5 years specifically. Well, that’s one address for him. I think his other addresses might look the same actually, but. Wow, so cool.”* P9, another data expert, also states, *“Yeah, being able to compare these two lines, I think, is very important for someone who’s working with this database”*

Moreover, the new views resulting from the study sometimes led to research conjectures or hypotheses. For instance, one of the data experts(P12) observed lines connecting the title to other columns, prompting them to ponder which firms collaborated and at what stage they were involved. This observation intrigued them, and they expressed their thoughts as follows: *“So you get 2 lines. So that tells me. That there’s that the people who printed this. Started they were active in. They started being active in like 1745 or whatever but there were also a firm involved that became active in 1725. so that’s almost you can likeHow*

established the firm is. Who published it? That's cool. I really like that. That's really needed. Oh, that's really needed. I like it a lot. Okay." Afterward, their focus shifted to the title column, and they remarked:

"I'm really, really interested in how you can see. The different firms like this is fascinating to me because this tells me that you've got firms working together who've been active at very different times, right? So like You've got a firm who started in 1765 working with a firm that started in 1785 right so like you've got someone really new to the business working with someone who hasn't been in the business all that long. That's really cool. That's really, really cool to know. And like throughout, like you've got kind of people at all sorts of stages. We have a lot of different firm groups that would work together, who like knew each other and they would publish together over like a really long time. So I'm I thought I would see more where they're like all going to like the same year. They all started around the same time where they all, you know. So that's really interesting.

The expert also started to have a hypothesis that Firms usually have various roles during their life in comparison to persons. They verbalized the idea as *"He was a printer. And a bookseller, is that right? The yellow and the green... fascinating So that's interesting because it means that our firms kind of show up more easily as having multiple roles. Because they more commonly had multiple rules. Sorry, that's cool."*

5.3.7 New Interesting Insights about Experts' Tasks

We were also observing data experts and monitoring their needs and extracting more different activities related to their data, especially about the relations between persons, titles and firms.. With P12, for example, we talked about different records that exist in the database according to the movement of the firm to different locations, therefore we think merging those firms in some a way that they are separable and also relatable could be interesting for data experts and help them explore better. Some of the data experts also mentioned that there are different prints of books in different places during the time, and being able to visually see the location or filter it can be useful, especially since they recently started to work separately on different locations of prints. They pointed out that currently they have separate people to work on different publications based on location and it will be helpful to add a location filter.

5.3.8 Non-Data Experts Suggesting Possible Use Cases

All of the data experts mentioned some other use cases for using the interactions and designs in other fields. Some non-experts, such as P6 [psychology background] mentioned the use case in psychology for categorizing patients in an organized way with a lot of data, saying *"This is definitely a good tool for a lot of information, I think, can be very useful, for even*

cases – I mean, I do. Psychology. And it’s ... I could see like this kind of tool being used for maybe kind of categorizing and having a very large set of data. And maybe patients, and with different variables here because I think it’s sometimes it’s very hard to find a visualization that it’s somewhat, you know, clean and useful when it comes to a lot of data.”

P3 [Archaeology background] suggested the use case for showing the bones and their information and also mentioned the possible use case as a tool that can be adjusted to different first-level courses *“I’m thinking that it would be really easy to apply this to like something that’s not like ... to different datasets so like not people and literature, but like, I don’t know something in my own discipline so like....The name of the bone, and then the features of the bone, and then what is the reference? Where that define?, or something like that? So it’ll be really easy, I think. Maybe not be that easy, but like to change just have the same visualization once you’re familiar with it. You could use this for different things like... or if you’re studying a class in like first-year biology, and you want to study for your classrooms like you make one thing here, and then you next year ...you have another class of like an app, and then you use the same tool, but just for different things.”*

Two participants mentioned they can see the use case of this tool in research, such as by showing the researchers and their research interests. One participant also suggested making it phone-friendly so that it will be more accessible.

5.3.9 User Experience and Suggestions

During the post-study interview, participants were asked about potential changes and additional features they would like to see in the tool. One main topic that arose was the size of certain elements. Although participants found the triangle indicators helpful, as P2 mentioned, *“I really like the arrows that come up, cause then I can look away. I don’t have to track where it’s going,”* They also expressed that these indicators were too small and hard to notice on the screen. To address this, a simple solution would be to increase the size of the triangles, particularly horizontally.

Another size-related issue was with the thickness of the connecting lines. In the study version, the line thickness was proportional to the number of connections. However, it turned out that people mainly used the lines to understand if there was only one connection or more. The current version made it difficult to distinguish small numbers, so the plan is to use two thicknesses for the lines: one for one item and another for more than one. Similar to the connection lines, the number of circles in one year on people’s timelines will also be changed to only two sizes since users have to click on the year to view the spiral anyway.

Besides the thickness of the lines, the connection lines were found to be helpful in providing insights quickly and adding to the aesthetic appeal of the tool, as P2 stated, *“publishing, it was nice to have that visualization of the lines like I could see that it was, say, like a spread of 15, 20 years, rather than having to scroll through a bunch of titles and*

like look at the actual days coming up, I think it was more visually appealing for sure," and "it almost looks like art, with all the pretty lines."

Regarding color coding, most participants liked it for people's roles and timelines. However, they mentioned that the color legend in the corner would be out of sight if they scroll down, so they prefer it to be floating. Additionally, they wanted some sample color options besides the broad choices available in the color picker. In terms of color, two aspects were discussed during the interviews: color blindness and the consistency of color to convey the role of color on the border of the circle.

When it comes to using different shapes for different elements, almost everyone agreed that they were distinct enough. However, there was a suggestion from P5 to use a different shape instead of the black border to show more than one book in timelines.

Another topic that emerged during the discussions was related to spirals. Two participants mentioned that they find spirals visually appealing, with one of them even expressing interest in exploring a timeline with a larger spiral. However, P5 and P4, who are visualization experts, shared the perspective that spirals might not be the most user-friendly option. P5 suggested using straight diagonal lines instead, as they are less mentally demanding and easier to comprehend for users.

The final part that is worth mentioning is the timeline and time scale that is changing with user interaction. I observed that participants could estimate the time correctly. One of the factors that they gained help from was the helper lines. They liked that they could have the line have an exact scale and some of them even used the scale to partition the time and only focus on part of it. One of the suggestions related to this time was making the process of their appearance automatic by the time the user opens a specific year in the timeline. This way it would be faster to check if they have opened the correct year.

5.4 Study Summary

In this chapter, I described our qualitative study. Our primary aim of this study was better understand our tool in use. The qualitative data we gathered has provided valuable insights about both the data and the data experts' thoughts and ideas. However, we acknowledge that while qualitative data is rich, and there are other aspects of the data that warrant exploration to attain a deeper understanding and uncover more valuable insights.

Chapter 6

Conclusions

In this Chapter, I will review my work by first reflecting on the research and lessons learned throughout this thesis. I will then discuss the limitations encountered during the study and outline potential future directions. Finally, I will conclude this chapter by summarizing the key findings and contributions of the thesis.

6.1 Discussion on our Design Goals

In this section, I am going to talk about my design goals, and how much the design achieved them.

The first design goal (DG1) was to make an explorable design. Based on the feedback of the users in the questionnaires, they told us that they could find the time-related patterns and they could find the connections between titles and contributors (both persons and firms). We also observed that data experts used the tool to explore the inconsistencies that they noticed in the data. However, there are some areas that need more research with regard to this design goal. For example, being able to start exploring the data starting from details needs the starting point. The search option was suggested the study, however, it will also affect other aspects of explorability such as encountering new parts of data. Therefore, more research needs to be done in this regard.

The second design goal (DG2) was to keep all of the data on the screen, so we positioned the data on the timeline. Due to this choice, we encountered two types of data that couldn't be displayed on the timeline: records in each table that lack a date, and unknown individuals. To address this, we positioned both of them above the timelines, allowing the end-user to access them easily. In the regular zoom level also it is possible to fit all of the data on the screen. However, if the user open the timeline or increase the zoom level, the timeline won't fit on the screen, and the user need to scroll to have the full timeline. Therefore, I can say we reached this design goal for the current stage of dataset, but expanding the time range will effect this design goal

As for the third design goal (DG3), no records were altered during the pre-processing phase. Only for records with dates that were not in a recognizable format, we manually identified the year and categorized them accordingly. To elaborate on the process, we have a section of the code check if the year is in the proper format, if the format of the year is wrong I was printing the years, then I saw which formats are they, and I had code to based on each situation find the year that we need for visualization purpose. However, it's important to note that the details of each record in the dataset remain unchanged. Based on this, we reached this design goal; However, to show more detail about the time of the records on the screen, for example when we have the exact date of some records, more design exploration is possible to do.

As for the fourth design goal (DG4), our design is centered around the three primary data tables. This framework is flexible, allowing for the incorporation of additional data and the extension of the years covered by the visualization. There is an option to reach this goal better with providing an option for end-users to modify the data in the primary database in case they encounter any errors. Also, although this design is flexible, there are some options to increase the flexibility by letting the user to rotate the timelines, for example dragging the person timeline to be in the middle instead of corner, this will make this design more flexible.

Lastly, in regard to (DG5), to provide data experts with comprehensive record details, we have implemented an information box. This box displays all fields associated with any selected record. This design goal is completely achieved in the current design; one way to improve it can be to show some information if the user hover on the data point, without changing the information box.

6.2 Lessons Learned

This section is our reflection on the research process framed as lessons learned. Reflecting on the process is an important part of design studies as it will bridge between applying research to the current corpora of visualization [90] [45].

6.2.1 Data Expert Collaboration Needs Effort from Both Sides

In our project, I collaborated with a data expert from the English department to develop an explorable visualization that would be helpful in the research process. This cross-disciplinary approach was crucial, as the literature suggests that involving data experts in the design process could lead to better visualizations that better serve their specific needs. However, we encountered some challenges during this collaboration that provided valuable lessons for future projects.

One of the most discussed challenges in working with data experts is that they have limited availability due to their other commitments, making it difficult to schedule exten-

sive sessions to gather their needs fully [90]. Secondly, there is a language barrier between visualization specialists and data experts that obstructs effective communication. In many applied work situations also the data experts often struggle to articulate their requirements precisely, and we, as visualization specialists, had difficulties understanding their domain-specific terminology.

To reach a mutual language with the data expert and understand the needs, we tried to reach an optimum approach. Going through the suggestions from other research it is not always the best idea to have extensive meetings with data experts, especially if the problem is not exactly defined on the expert side and the time that the expert can put on the project is limited. Some researchers recommend immersing themselves in the experts' research groups to better grasp their needs and preferences [89]. Unfortunately, in our case, we couldn't adopt this approach because the data experts' research was spread across various phases, happening at different times.

Given the time constraints and the challenge of clear communication, we decided to explore a different strategy. After the first meeting about their needs and understanding of their data, we started to learn as much as we could from their website. Later, instead of having lengthy meetings, we organized sketching sessions with the experts. During these meetings, we provided a small sample dataset, and shared it with the group encouraging data experts to express their visualization requirements more informally (for the full explanation of the sketching session you can check the section 4.2). Sketching sessions have been used to spark their interest in the early stages when we had no concrete designs. In these sessions, both our research team and the data experts sketched their visualization ideas, allowing for a more interactive and understandable exchange of concepts.

Maintaining the data experts' interest and engagement was critical to the success of the project. To achieve this, adopting an iterative approach, and continuously seeking feedback from the data experts throughout the design and implementation process was necessary. By involving them at every stage and incorporating their insights, it is possible to ensure that the final visualization tool came at least close to meeting their needs. In this process, I noticed that progressing in the implementation and design at the same time will simplify the process of explaining the design to experts and will make them more interested in the project. Our data experts had the opportunity to observe the visualization and its use cases, making it easier for them to evaluate and provide feedback.

In conclusion, my experience, in this project, highlighted the significance of overcoming communication challenges in working with data experts from different fields. By simplifying communication, and focusing on increasing the data experts' interest it is possible to involve data experts from the beginning and foster continuous feedback. Sketching sessions with experts and having a simple implementation of the design in the design meetings can be helpful in this process.

6.2.2 Using Familiar Designs is Not Always the Key

In visualization research, two distinct approaches emerge when designing for special use cases. Some emphasize the importance of simplicity and maintaining familiarity with the visualizations for the data experts, and name this as crucial for a visualization’s success. On the other hand, researchers who focus on explorability in visualization tend to advocate for designs that cater to non-data experts and encourage the use of creative concepts and interactions to make explorable visualization.

Our research is designing a visualization that is both explorable and usable for data experts in their research process. In our design study, we chose to prioritize exploratory aspect of the visualization. By doing so, we aimed to facilitate dataset exploration, allowing users to delve deeply into the data utilizing interactions. Subsequently, we worked on enhancing the usability of the visualization. We did not limit ourselves to using familiar visualisations because, while they are effective at showing results, they were ineffective in our use case given the challenges.

While we observed the existence of a learning curve for users, we found that after comprehending the visualization and receiving proper explanations, users were able to use it independently and effectively explore the data. Our approach was successful in providing usability as well as explorability, acknowledging the needs of data experts while also ensuring accessibility for non-data experts.

The lesson learned from this design study is that by focusing on exploratory capabilities initially and then improving usability, we can create a visualization tool that empowers users to efficiently explore and analyze data on their own, after overcoming the learning curve.

6.3 Limitation and Future Work

One limitation of our prototype is the restricted number of timelines that can be opened on the screen simultaneously. While we had already checked this limitation with our data experts during our initial design, some scenarios related to the dataset emerged during the study. To address this, we believe it would be beneficial to make it possible to expand the distance on the screen to accommodate more timelines.

Initially, our approach in creating this tool focused on promoting direct interaction with the data to discover unexpected information. Hence, we intentionally avoided incorporating a search engine. However, after consulting with data experts and gaining insights into their perspectives, we recognized that adding search functionality could greatly enhance the speed of information retrieval. Thus, we now aim to explore a balanced combination of search capabilities and the tool’s inherent features to find an optimal solution for this issue. As P1 expressed, *“So this is interesting because it gives me the visual component, but in terms of finding information quickly for me, I would maybe struggle with this a little bit.”*

In the future, based on data experts feedback, our goal is to deploy the tool on the WPHP website. By doing so, we can conduct a Multidimensional In-depth Long-term Case study [105] using this tool. While our current study helped us better understand the tool, we understand that no study is flawless [111]. Therefore, we believe that conducting a long-term study will provide us with further insights into the design and help us enhance its exploratory capabilities. Through more research and improvement, we aim to make our tool even more effective and valuable for data exploration and analysis, not only with our use case but also with more general use cases. As we received feedback from non-data experts on use cases from other domains, one potential direction would be to adapt the visualization to make it suitable for a wider range of datasets with varying characteristics.

6.4 Conclusions

I have designed, implemented, and studied a visualization that is explorable and can help research. I collaborated with researchers from the English department who are working on Women's Print History Project during my research process to make a visualization that can be useful in their research process. The primary contribution of this research lies in identifying and extracting the specific needs and tasks within the domain in Chapter 3. The second significant contribution is the successful design and implementation of a tool tailored to address the requirements of data experts as described in Chapter 4. Throughout the process, we elucidated our design goals and provided justifications for the design choices made in alignment with these goals. Furthermore, we conducted a qualitative study of our design with two distinct groups of participants: data experts and general users. This study allowed us to gather feedback on two important aspects - the tool's effectiveness in addressing data experts' research needs and its overall user experience for a broader audience. We have reflected on the process and things we have learned in the lesson learned section.

Bibliography

- [1] Eighteenth century collections online. Gale.
- [2] English short title catalogue. London : British Library. Title from opening screen (publisher's Web site, viewed Apr. 6, 2007).; Also available on CD-ROM.
- [3] Undergraduate research in the arts & humanities. <https://artshumanities.berkeley.edu/resources/undergraduate-resources/undergraduate-research-arts-humanities#:~:text=Research%20in%20the%20Arts%20and,films%20or%20other%20human%20productions>).
- [4] Voyant Tools. <https://voyant-tools.org/>.
- [5] Literature is not data: Against digital humanities. <https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/>, Oct 2012.
- [6] Clemens Arbesser, Florian Spechtenhauser, Thomas Mühlbacher, and Harald Piringer. Visplause: Visual data quality assessment of many time series using plausibility checks. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):641–650, 2017.
- [7] Gregg Skip Bailey. Iterative methodology and designer training in human-computer interface design. *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, 1993.
- [8] Rahul C. Basole, Trustin Clear, Mengdie Hu, Harshit Mehrotra, and John Stasko. Understanding interfirm relationships in business ecosystems with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2526–2535, 2013.
- [9] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1):361–362, Mar. 2009.
- [10] Marcia J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5):407–424, Jan 1989.
- [11] Leilani Battle and Jeffrey Heer. Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. *Computer Graphics Forum*, 38:145–159, 06 2019.

- [12] Alejandro Benito-Santos and Roberto Therón Sánchez. A data-driven introduction to authors, readings, and techniques in visualization for the digital humanities. *IEEE Computer Graphics and Applications*, 40(3):45–57, 2020.
- [13] Marcus Bingenheimer, Jen-Jou Hung, and Simon Wiles. Social network visualization from tei data. *Lit. Linguistic Comput.*, 26:271–278, 2011.
- [14] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011.
- [15] Nadia Boukhelifa, Marc-Emmanuel Perrin, Samuel Huron, and James Eagan. How data workers cope with uncertainty: A task characterisation study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 3645–3656, New York, NY, USA, 2017. Association for Computing Machinery.
- [16] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
- [17] Matthew Brehmer, Stephen Ingram, Jonathan Stray, and Tamara Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *Visualization and Computer Graphics, IEEE Transactions on*, 20:2271–2280, 12 2014.
- [18] Bill Buxton, Sheelagh Carpendale, and Nicolai Marquardt. *Sketching user experiences the workbook*. Elsevier Science, 2012.
- [19] Nan Cao, J. Sun, Yu-Ru Lin, David Gotz, Shixia Liu, and Huamin Qu. Facetatlas: Multifaceted visualization for rich text corpora. *Visualization and Computer Graphics, IEEE Transactions on*, 16:1172 – 1181, 01 2011.
- [20] Stuart K. Card and David Nation. Degree-of-interest trees: A component of an attention-reactive user interface. In *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '02*, page 231–245, New York, NY, USA, 2002. Association for Computing Machinery.
- [21] Chaomei Chen. Citespace ii: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57:359–377, 02 2006.
- [22] Isaac Cho, Wenwen Dou, Xiaoyu Wang, Eric Sauda, and William Ribarsky. Vairoma: A visual analytics system for making sense of places, times, and events in roman history. *IEEE Transactions on Visualization and Computer Graphics*, 22:210–219, 2016.
- [23] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.*, 41(1), jan 2009.
- [24] Christopher Collins, Fernanda Viegas, and Martin Wattenberg. Parallel tag clouds to explore and analyze faceted text corpora. pages 91 – 98, 11 2009.
- [25] T. F. Cox. Multidimensional scaling., 2001.

- [26] Anamaria Crisan and Tamara Munzner. Uncovering data landscapes through data reconnaissance and task wrangling. In *2019 IEEE Visualization Conference (VIS)*, pages 46–50, 2019.
- [27] Antonina Dattolo, Marco Corbatto, and Marco Angelini. Authoring and reviewing bibliographies: Design and development of a visual analytics online platform. *IEEE Access*, 10:21631–21645, 2022.
- [28] Anthony Don, Elena Zheleva, Machon Gregory, Sureyya Tarkan, Loretta Auvil, Tanya Clement, Ben Shneiderman, and Catherine Plaisant. Discovering interesting usage patterns in text collections: Integrating text mining with visualization. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, CIKM '07*, page 213–222, New York, NY, USA, 2007. Association for Computing Machinery.
- [29] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. The information flaneur: A fresh look at information seeking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, page 1215–1224, New York, NY, USA, 2011. Association for Computing Machinery.
- [30] Cody Dunne, Nathalie Henry Riche, Bongshin Lee, Ronald Metoyer, and George Robertson. Graphtrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. 04 2012.
- [31] Marian Dörk, Sheelagh Carpendale, Christopher Collins, and Carey Williamson. Visgets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1205–1212, 2008.
- [32] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. Visualizing explicit and implicit relations of complex information spaces. *Information Visualization*, 11:5–21, 01 2012.
- [33] Marian Dörk, Nathalie Henry Riche, Gonzalo Ramos, and Susan Dumais. Pivotpaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2709–2718, 2012.
- [34] Marian Dörk, Christopher Pietsch, and Gabriel Credico. One view is not enough: High-level visualizations of a large cultural collection. *Information Design Journal*, 23:39–47, 07 2017.
- [35] Isabella (Belle) Eist. Hidden in the imprints: Introducing Ann Vernor, bookseller and publisher, active 1793-1807. <https://womensprinthistoryproject.com/blog/post/117>, 2023. [last access: Apr-15-2023].
- [36] Niklas Elmqvist and Jean-Daniel Fekete. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE Transactions on Visualization and Computer Graphics*, 16(3):439–454, may 2010.
- [37] Stefania Forlini, Uta Hinrichs, and John Brosz. Mining the material archive: Balancing sensate experience and sense-making in digitized print collections. *Open Library of Humanities*, 4(2):35, 2018.

- [38] Sohaib Ghani, Bum Chul Kwon, Seungyoon Lee, Ji Soo Yi, and Niklas Elmqvist. Visual analytics for multimodal social network analysis: A design study with social scientists. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2032–2041, 2013.
- [39] Fred Gibbs and Trevor Owens. Building better digital humanities tools. *DH Quarterly*, 6(2):1–14, 2012.
- [40] Doron Goldfarb, Max Arends, Josef Froschauer, and Dieter Merkl. Revisiting 3d information landscapes for the display of art historical web content. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, ACE '11*, New York, NY, USA, 2011. Association for Computing Machinery.
- [41] John D. Gould and Clayton Lewis. Designing for usability—key principles and what designers think. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '83*, page 50–53, New York, NY, USA, 1983. Association for Computing Machinery.
- [42] Hua Guo and David H. Laidlaw. Topic-based exploration and embedded visualizations for research idea generation. *IEEE Transactions on Visualization and Computer Graphics*, 26(3):1592–1607, 2020.
- [43] Uta Hinrichs, Beatrice Alex, Jim Clifford, and Aaron Quigley. Trading consequences: A case study of combining text mining visualisation to facilitate document exploration. In *Proceedings of Digital Humanities 2014 (DH 2014)*, 2014.
- [44] Uta Hinrichs, Stefania Forlini, and Bridget Moynihan. Speculative practices: Utilizing infovis to explore untapped literary collections. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):429–438, 2016.
- [45] Uta Hinrichs, Stefania Forlini, and Bridget Moynihan. In defense of sandcastles: Research thinking through visualization in digital humanities. *Digital Scholarship in the Humanities*, 34(Supplement_1):i80–i99, 10 2018.
- [46] Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale. 2009.
- [47] Matthew Jockers. *Macroanalysis: Digital Methods and Literary History*. 04 2013.
- [48] Hyunmo Kang, Catherine Plaisant, Bongshin Lee, and Ben Bederson. Netlens: Iterative exploration of content-actor network data. pages 91–98, 10 2006.
- [49] David Knoke and Song Yang. *Social Network Analysis*. [Place of publication not identified], 2008.
- [50] Bill Kules and Robert Capra. Designing exploratory search tasks for user studies of information seeking support systems. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '09*, page 419–420, New York, NY, USA, 2009. Association for Computing Machinery.
- [51] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE transactions on visualization and computer graphics*, 18(9):1520–1536, 2011.

- [52] H. Lamqaddam. When the tech kids are running too fast: Data visualisation through the lens of art history research. *3rd Workshop on Visualization for the Digital Humanities at IEEEVIS*.
- [53] Houda Lamqaddam, Andrew Vande Moere, Vero Vanden Abeele, Koenraad Brosens, and Katrien Verbert. Introducing layers of meaning (lom): A framework to reduce semantic distance of visualization in humanistic research. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1084–1094, 2021.
- [54] Craig Larman and Victor Basili. Iterative and incremental development: A brief history. *Computer*, 36:47 – 56, 07 2003.
- [55] Bongshin Lee, Mary Czerwinski, George Robertson, and Benjamin B. Bederson. Understanding research trends in conferences using paperlens. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, page 1969–1972, New York, NY, USA, 2005. Association for Computing Machinery.
- [56] Bongshin Lee, Greg Smith, George Robertson, Mary Czerwinski, and Desney Tan. Facetlens: Exposing trends and relationships to support sensemaking within faceted datasets. pages 1293–1302, 04 2009.
- [57] Michelle Levy. Ann Williams: Postmistress, poetess, sericulturist. <https://womensprinthistoryproject.com/blog/post/106>, 2022. [last access: Apr-15-2023].
- [58] Michelle Levy and Kandice Sharren. The women’s print history project. <https://womensprinthistoryproject.com/>, 2023. [last access: Apr-15-2023].
- [59] Haihan Lin, Derya Akbaba, Miriah Meyer, and Alexander Lex. Data hunches: Incorporating personal knowledge into visualizations. 29(1):504–514, 2022.
- [60] Timothy Luciani, Andrew Burks, Cassiano Sugiyama, Jonathan Komperda, and G. Elisabeta Marai. Details-first, show context, overview last: Supporting exploration of viscous fingers in large-scale ensemble simulations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1225–1235, 2019.
- [61] Kashumi Madampe, Rashina Hoda, and John Grundy. A faceted taxonomy of requirements changes in agile contexts. *IEEE Transactions on Software Engineering*, 48(10):3737–3752, 2022.
- [62] Timothy Major and Rahul C. Basole. Graphicle: Exploring units, networks, and context in a blended visualization approach. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):576–585, 2019.
- [63] Lev Manovich. What is visualisation? *Visual Studies*, 26(1):36–49, 2011.
- [64] Ji-Ye Mao, Karel Vredenburg, Paul W. Smith, and Tom Carey. The state of user-centered design practice. *Commun. ACM*, 48(3):105–109, mar 2005.
- [65] Gary Marchionini. Exploratory search: From finding to understanding. *Commun. ACM*, 49(4):41–46, apr 2006.
- [66] Vivien Marx. The big challenges of big data. *Nature*, 498(7453):255–260, 2013.

- [67] Zainab Masood, Rashina Hoda, and Kelly Blincoe. Real world scrum a grounded theory of variations in practice. *IEEE Transactions on Software Engineering*, 48(5):1579–1591, 2022.
- [68] Justin Matejka, Tovi Grossman, and George Fitzmaurice. Citeology: Visualizing paper genealogy. *Conference on Human Factors in Computing Systems - Proceedings*, 05 2012.
- [69] Nina Mccurdy, Julie Gerdes, and Miriah Meyer. A framework for externalizing implicit error using visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):925–935, 2019.
- [70] K. T. McDonnell and K. Mueller. Illustrative parallel coordinates. *Computer Graphics Forum*, 27(3):1031–1038, 2008.
- [71] Christofer Meinecke, Ahmad Dawar Hakimi, and Stefan Jänicke. Explorative visual analysis of rap music. *Information*, 13(1), 2022.
- [72] Franco Moretti. *Graphs, maps, trees abstract models for a literary history*. Verso, 2007.
- [73] Michael J. Muller and Sarah Kuhn. Participatory design. *Commun. ACM*, 36(6):24–28, jun 1993.
- [74] Tamara Munzner. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.
- [75] Tamara Munzner and Eamonn Maguire. *Visualization Analysis amp; Design*. CRC Press, 2015.
- [76] Brad Myers. Challenges of hci design and implementation. *Interactions*, 1(1):73–83, jan 1994.
- [77] Georgia Panagiotidou, Ralf Vandam, Jeroen Poblome, and Andrew Vande Moere. Implicit error, uncertainty and confidence in visualization: An archaeological case study. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4389–4402, 2022.
- [78] Daniel Park, Steven M. Drucker, Rodrigo Fernandez, and Niklas Elmqvist. ATOM: A Grammar for Unit Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3032–3043, December 2018.
- [79] Christian Partl, Alexander Lex, Marc Streit, Hendrik Strobelt, Anne Wassermann, Hanspeter Pfister, and Dieter Schmalstieg. Contour: Data-driven exploration of multi-relational datasets for drug discovery. *Visualization and Computer Graphics, IEEE Transactions on*, 20:1883–1892, 12 2014.
- [80] Robert Pienta, James Abello, Minsuk Kahng, and Duen Horng Chau. Scalable graph exploration and visualization: Sensemaking challenges and opportunities. In *2015 International Conference on Big Data and Smart Computing (BIGCOMP)*, pages 271–278, 2015.

- [81] Catherine Plaisant, Brett Milash, Anne Rose, Seth Widoff, and Ben Shneiderman. Lifelines: Visualizing personal histories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, page 221–227, New York, NY, USA, 1996. Association for Computing Machinery.
- [82] Catherine Plaisant, James Rose, Bei Yu, Loretta Auvil, Matthew G. Kirschenbaum, Martha Nell Smith, Tanya Clement, and Greg Lord. Exploring erotics in emily dickinson’s correspondence with text mining and visual interfaces. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL '06, page 141–150, New York, NY, USA, 2006. Association for Computing Machinery.
- [83] Jonathan C. Roberts, Chris Headleand, and Panagiotis D. Ritsos. Sketching designs using the five design-sheet methodology. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):419–428, 2016.
- [84] George Robertson, Kim Cameron, Mary Czerwinski, and Daniel Robbins. Polyarchy visualization: Visualizing multiple intersecting hierarchies. 12 2001.
- [85] Stan Ruecker, Milena Radzikowska, Piotr Michura, Carlos Fiorentino, and Tanya Clement. Visualizing repetition in text. *Digital Studies / Le champ numérique*, 1(3):1–9, 2009.
- [86] Maximilian Schich, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing. A network framework of cultural history. *Science*, 345(6196):558–562, 2014.
- [87] m.c. schraefel, Max Wilson, Alistair Russell, and Daniel A. Smith. Mspace: Improving information access to multimedia domains with multimodal exploratory search. *Commun. ACM*, 49(4):47–49, apr 2006.
- [88] D. Schuler. Participatory design: Principles and practices, 1993.
- [89] Michael Sedlmair. *Visual analysis of in-car communication networks*. PhD thesis, lmu, 2010.
- [90] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.
- [91] Rachel Shadoan and Chris Weaver. Visual analysis of higher-order conjunctive relationships in multidimensional data using a hypergraph query system. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2070–2079, December 2013.
- [92] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. *IEEE Symposium on Visual Languages, Proceedings*, 03 2000.
- [93] Ben Shneiderman and Aleks Aris. Network visualization by semantic substrates. *IEEE transactions on visualization and computer graphics*, 12(5):733–740, 2006.
- [94] Ben Shneiderman and Aleks Aris. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):733–740, 2006.

- [95] Ben Shneiderman, David Feldman, Anne Rose, and Xavier Ferré Grau. Visualizing digital library search results with categorical and hierarchical axes. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, DL '00, page 57–66, New York, NY, USA, 2000. Association for Computing Machinery.
- [96] Ben Shneiderman and Catherine Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, page 1–7, New York, NY, USA, 2006. Association for Computing Machinery.
- [97] James Slack, Kristian Hildebrand, and Tamara Munzner. Prasad: A partitioned rendering infrastructure for scalable accordion drawing. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 41–48. IEEE, 2005.
- [98] Annelien Smets. Designing for serendipity: a means or an end? *Journal of Documentation*, 79(3):589–607, Jan 2023.
- [99] Michael Smuc, Eva Mayr, Tim Lammarsch, Wolfgang Aigner, Silvia Miksch, and Johannes Gaertner. To score or not to score? tripling insights for participatory design. *IEEE computer graphics and applications*, 29:29–38, 05 2009.
- [100] Robert Spence. Information visualization. *Harlow, Addison-Wesley*, 01 2006.
- [101] John Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, apr 2008.
- [102] Moritz Stefaner, Thomas Urban, and Marc Seefeldler. Elastic lists for facet browsing and resource analysis in the enterprise. pages 397–401, 10 2008.
- [103] Laura Swietlicki and Pierre Cubaud. Overview visualizations for large digitized correspondence collections: A design study. In Gianmaria Silvello, Oscar Corcho, Paolo Manghi, Giorgio Maria Di Nunzio, Koraljka Golub, Nicola Ferro, and Antonella Poggi, editors, *Linking Theory and Practice of Digital Libraries*, pages 266–273, Cham, 2022. Springer International Publishing.
- [104] Paola Valdivia, Paolo Buono, Catherine Plaisant, Nicole Dufournaud, and Jean-Daniel Fekete. Analyzing dynamic hypergraphs with parallel aggregated ordered hypergraph visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(1):1–13, 2021.
- [105] Eliane R. A. Valiati, Carla M. D. S. Freitas, and Marcelo S. Pimenta. Using multi-dimensional in-depth long-term case studies for information visualization evaluation. In *Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '08, New York, NY, USA, 2008. Association for Computing Machinery.
- [106] Frank van Ham and Adam Perer. “search, show context, expand on demand”: Supporting large graph exploration with degree-of-interest. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):953–960, 2009.

- [107] Fernanda B. Viegas, Martin Wattenberg, and Jonathan Feinberg. Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1137–1144, 2009.
- [108] Fernanda B. Viegas, Martin Wattenberg, and Jonathan Feinberg. Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1137–1144, nov 2009.
- [109] Pawan Vora. *Web application design patterns*. Morgan Kaufmann, 2009.
- [110] Angela Wachowich. (Unidentified) woman not inferior to man: ‘Sophia,’ proto-feminism, and the anonymous female writer. <https://womensprinthistoryproject.com/blog/post/97>, 2022. [last access: Apr-15-2023].
- [111] Emily Wall, Cindy Xiong, and Yea-Seul Kim. Vishikers’ guide to evaluation: Competing considerations in study design. *IEEE Computer Graphics and Applications*, 42(3):29–38, 2022.
- [112] Martin Wattenberg. Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’06, page 811–819, New York, NY, USA, 2006. Association for Computing Machinery.
- [113] Chris Weaver. Cross-filtered views for multidimensional visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):192–204, 2010.
- [114] Gunther H. Weber, Sheelagh Carpendale, David Ebert, Brian Fisher, Hans Hagen, Ben Shneiderman, and Anders Ynnerman. Apply or die: On the role and assessment of application papers in visualization. *IEEE Computer Graphics and Applications*, 37(3):96–104, 2017.
- [115] Ryen W. White, Bill Kules, Steven M. Drucker, and m.c. schraefel. Introduction. *Commun. ACM*, 49(4):36–39, apr 2006.
- [116] Mitchell Whitelaw. Generous interfaces for digital cultural collections. *Digit. Humanit. Q.*, 9(1), 2015.
- [117] Florian Windhager, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2311–2330, 2019.
- [118] Ka-Ping Yee, Kirsten Swearingen, Kevin Li, and Marti Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’03, page 401–408, New York, NY, USA, 2003. Association for Computing Machinery.
- [119] Ruojin Zhang, Vimukthi Jayawardene, Marta Indulska, Shazia Wasim Sadiq, and Xiaofang Zhou. A data driven approach for discovering data quality requirements. In *International Conference on Interaction Sciences*, 2014.
- [120] Torre Zuk and Sheelagh Carpendale. Visualization of uncertainty and reasoning. pages 164–177, 06 2007.

Appendix A

Study Material

A.0.1 Pre-Study Questionnaire

Exploring using of Visualization in exploring the historical datasets

Study Number: 30001698 Pre-Study Information

1. 1. Pick the right age range for you?

Mark only one oval.

- 18-24 years old
- 25-34 years old
- 35-44 years old
- 45-54 years old
- 55-65 years old

2. What is your gender?

Mark only one oval.

- Man
- Woman
- Non-binary
- Prefer not to answer
- Other: _____

Background: Name the postsecondary majors that you have studied or you are studying in each of the following:

3. Bachelor's degree:.....

4. Master's degree:.....

5. PhD degree:.....

6. Other:.....

7. What is your current position? (Sample answer: student, software developer, math teacher, etc.. Please **do not** specify where you work or study.)

8. How much experience do you have working with visualization tools to explore data?

A.0.2 Post-Study Questionnaire

Exploring using of Visualization in exploring the historical datasets

post -Study questionnaire:

1. I found the system easy to use...

Check all that apply.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

2. I was able to find time-based patterns in the dataset...

Mark only one oval.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

3. I was able to find which people worked with what firms using this tool...

Mark only one oval.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

4. This system helped me to get the information I wanted from the dataset...

Mark only one oval.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

5. After using this system I feel confident about the usefulness of visual tools to explore my datasets...

Mark only one oval.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

6. On a scale of 1-7 how likely are you to reuse this tool for knowing more about this dataset?

7. On a scale of 1-7, how likely are you to reuse this tool for your future data-related project?

8. On a scale of 1-7, how likely are you to use similar interactions for your future data-related project?

9. How likely are you to recommend this tool to your friends?

This content is neither created nor endorsed by Google.

Google Forms