Universität
Bremen

Multimodal and Collaborative Interaction for Visual Data Exploration

Gabriela Molina León

Supervisors:
Andreas Breiter, University of Bremen
Petra Isenberg, Inria & Université Paris-Saclay

Reviewers:
Prof. Dr. Andreas Breiter, University of Bremen
Prof. Niklas Elmqvist, PhD, University of Aarhus

Examination Board:
Prof. Dr. Andreas Breiter, University of Bremen
Prof. Niklas Elmqvist, PhD, University of Aarhus
Prof. Dr. Tanja Schultz, University of Bremen
Prof. Dr. Yvonne Rogers, University College London
Anna Ricarda Luther, University of Bremen
Nicolás Ruiz, University of Bremen

PhD Defense:
April 24th, 2024

Multimodal and Collaborative Interaction for Visual Data Exploration

A dissertation presented by
Gabriela Molina León
to the Faculty of Mathematics and Computer Science

in partial fulfillment of the requirements
for the degree of
Dr.-Ing.

University of Bremen
Bremen, Germany
March 2024

Abstract

As we generate and encounter vast amounts of data every day, the need to support human-data interaction increases. This dissertation investigates how different interaction modalities and devices can support data experts to visually explore and make sense of data, individually and collaboratively. Through a series of empirical studies applying mixed methods, I study how experts interact and wish to interact with spatio-temporal data on tablets and large vertical displays at the workplace.

While data exploration and sensemaking usually take place on a desktop computer, there is a diverse range of computing devices that provide novel ways of interacting beyond the standard mouse and keyboard input devices used for WIMP (windows, icons, menus, pointers) interfaces. Therefore, I explore how different interaction modalities, such as touch, speech, pen, and mid-air gestures, can support exploratory and sensemaking tasks on interactive surfaces. The starting point of this dissertation is a visualization design study with social policy researchers to study data-driven work in the context of a real-world scenario. Through this design study, the dissertation contributes the first formal evaluation of co-creation as a methodology for visualization design as well as the data and task abstractions derived from the social policy research domain. After defining the design requirements of the domain experts, the dissertation focuses on the interaction design of visualization systems that support the data-driven tasks. A comparative evaluation on visual data exploration between desktop and tablet-based workplaces revealed that experts apply different interaction strategies to solve similar tasks across devices, making use of different views and interaction techniques.

Following up on the single-user scenario, I examine how pairs of experts interact with visualization systems, and with each other, in the context of co-located and synchronous work. The dissertation presents how experts wish to interact on large vertical displays through an elicitation study with touch, pen, speech, and mid-air gestures. The dominance of speech interaction among user preferences leads to an in-depth exploratory study on the role of speech in collaborative work. Despite its challenges, speech interaction benefits awareness and is evenly present in closely

coupled and loosely coupled collaboration. Overall, participants prefer interacting unimodally, changing modalities depending on the task and the distance from the display. The dissertation contributes the characterization of interaction patterns and strategies on multimodal visual systems, in individual and collaborative scenarios. Based on the findings on performance, user experience, and interaction choices, the dissertation provides a series of design considerations for multimodal and collaborative systems to support data exploration and sensemaking, together with two systems that enable individuals and small groups to perform such tasks.

This thesis is dedicated to the loving memory of my grandmother, Elena.

Acknowledgments

First of all, I would like to thank my supervisor, Andreas Breiter, for being a supportive mentor throughout my PhD journey. Thank you for giving me the freedom to pursue what I am passionate about and for encouraging me to work interdisciplinarily. I would also like to thank Petra Isenberg for becoming my second supervisor. Thank you for your mentorship, for introducing me to the larger visualization community, and for inviting me to the Aviz lab, where I had a wonderful time.

I would also like to thank my co-authors for their collaboration and for teaching me to look at my work from different perspectives. Thank you to the participants of all my studies for joining my experiments, and allowing me to learn from their work. Thank you to my former and current colleagues from the Collaborative Research Center 1342 and the Institute for Information Management Bremen, especially Michael Lischka, Nils Düpont, Hendrik Heuer, Ivo Mossig, Gabriella Skitalinska, Irina Zakharova, Juliane Jarke, Keonhi Son, Helen Seitzer, and Jakob Henninger, for creating a supportive and joyful environment to work at.

I am also grateful to my friends from home and from Weimar, who supported me from wherever they were, and who made me laugh about the good and the bad moments. Special thanks to Alechandrina Pereira, César Hernandez, Adrian Lozano, Veronika Krauß, Frederik Schreiber, Cindy Molina, and Andrea Holmqvist.

Finally, thank you to Henning for his patience and encouragement through the ups and downs of this journey. Thank you to my family, especially my mother, for raising me to believe I can achieve any goal I set for myself. And to Fritz, my cat, for cheering me up during the long writing days.

Contents

Publications

This dissertation is partly based on previously published content from the following peer-reviewed publications. If applicable, the reuse of content is indicated at the beginning of the corresponding chapter.

Journal Articles

- Gabriela Molina León, Petra Isenberg, and Andreas Breiter. "Eliciting Multimodal and Collaborative Interactions for Data Exploration on Large Vertical Displays". In: *IEEE Transactions on Visualization and Computer Graphics* 30.2 (2024), pp. 1624–1637. ISSN: 1941-0506. DOI: `10.1109/TVCG.2023.3323150`.
- Gabriela Molina León, Michael Lischka, Wei Luo, and Andreas Breiter. "Mobile and Multimodal? A Comparative Evaluation of Interactive Workplaces for Visual Data Exploration". In: *Computer Graphics Forum* 41.3 (2022), pp. 417–428. DOI: `10.1111/cgf.14551`.
- Gabriela Molina León and Andreas Breiter. "Co-creating Visualizations: A First Evaluation with Social Science Researchers". In: *Computer Graphics Forum* 39.3 (2020), pp. 291–302. DOI: `10.1111/cgf.13981`.

Conference Publications

- Gabriela Molina León, Michael Lischka, and Andreas Breiter. "Mapping the Global South: Equal-Area Projections for Choropleth Maps". In: *2020 IEEE Visualization Conference (VIS)*. 2020, pp. 91–95. DOI: `10.1109/VIS47514.2020.00025`.

Workshop Publications

- Gabriela Molina León, Petra Isenberg, and Andreas Breiter. "Talking to Data Visualizations: Opportunities and Challenges". In: *VIS Workshop on Multimodal Experiences for Remote Communication Around Data Online*. 2023. arXiv: 2309.09781 [cs.HC].
- Gabriela Molina León. "Advancing Inclusive Design in Multiple Dimensions". In: *CHI Workshop on Inclusive Design of CUIs Across Modalities and Mobilities*. 2023. arXiv: 2303.16790 [cs.HC].

Poster Publications

- Gabriela Molina León, Gabriella Skitalinskaya, Nils Düpont, Jonas Klaff, Anton Schlegel, Hendrik Heuer, and Andreas Breiter. "Co-Creating a Research Data Infrastructure with Social Policy Researchers". In: *Proceedings of 20th European Conference on Computer-Supported Cooperative Work*. European Society for Socially Embedded Technologies (EUSSET), 2022. DOI: 10.48340/ecscw2022_p03.

To Be Submitted

- Gabriela Molina León, Anastasia Bezerianos, Olivier Gladin, and Petra Isenberg. *Talk to the Wall: The Role of Speech Interaction in Collaborative Work*. To submit. 2024.

The content of Chapter 6 is to be submitted soon for publication.

1 Introduction

We generate data every day to abstract physical and social phenomena around us. We abstract the world into datasets to generate knowledge about all kinds of topics, such as economic development, climate change, and pandemics. Accordingly, data sensemaking and exploration have become essential in many jobs. Professionals apply computational approaches to examine large amounts of data, gain insights, and make decisions based on their findings. Although many automated methods are available, human intervention is still necessary, given that, for example, transforming social phenomena into data can lead to discrimination against individuals from disadvantaged groups (Mejías and Couldry, 2019). To consider potential biases and make informed decisions, professionals need to filter, manipulate, and inspect the data from multiple perspectives: to interact with it. Accordingly, interaction is an essential aspect of the data analysis process.

One way to interact with data is to interact with their visual representations. Computer-based visualizations serve as data representations that make use of human visual capabilities to amplify cognition (Card et al., 1999). Interactive data visualizations not only help us to detect data patterns visually and expand our working memory, but also allow us to manipulate large amounts of data to examine it from different perspectives. As such, visualization is an important instrument to facilitate data exploration. Considering the information seeking mantra "Overview first, zoom and filter, then details-on-demand" of Shneiderman (1996), systems designed for visual data exploration should support different exploratory tasks, such as identifying items of interest, filtering, and lookup.

Furthermore, while data work can be performed individually, it is often part of teamwork within a larger collective. Professionals work together to draw data-driven conclusions in line with the goals of their organization, combining their knowledge and skills. Yet, interactive systems designed for individual users do not necessarily support collaborative work, where communication, awareness, and territoriality are important factors (Scott et al., 2004). At the same time, technology affects how people work in groups (Grudin, 1994). Thus, we should think of social norms and collaborative interaction when designing visual interactive systems.

We generate billions of gigabytes of data every day worldwide through diverse computing devices (World Economic Forum, 2022). These devices can be stationary or mobile, have displays of different sizes, and support multiple interaction modalities. Traditionally, data exploration has taken place on desktop computers with a mouse and a keyboard through a WIMP (windows, icons, menus, pointers) interface. However, nowadays, *beyond the desktop* devices (Lee et al., 2012) have become common at the workplace. Tablets are used as portable computers to perform personal tasks, while large vertical displays have become a regular element of meeting rooms to assist group work. Today, at least 1.28 billion people own tablets worldwide (Alsop, 2022). The tablet is a promising device type for individual work, as tablet adoption has significantly grown to support remote and hybrid work due to the COVID-19 pandemic (Corporation, 2023). Still, we have not yet fully explored and taken advantage of all the interaction possibilities that these devices offer to work with data. Accordingly, this dissertation focuses on investigating multimodal and collaborative human-data interaction in professional settings. Given that the design of interactive tools must take the context of use into account, I define the context of my dissertation research according to the five Ws (*What, Why, Who, Where, and When*) of Tominski and Schumann (2020) for describing the context of interactive visual analysis: My goal is to understand how we can design multimodal and collaborative interactive systems for exploration and sensemaking tasks (*why*) with spatio-temporal data (*what*) by data experts (*who*) at their workplace (*where*), at their first encounter with said data (*when*). We need to facilitate visual data exploration by supporting different goals, such as discovering patterns and anomalies, querying details, and generating hypotheses to later test through statistical analyses.

An interactive system can be composed of different input devices and interaction modalities*. Aside from using a mouse and a keyboard as input devices, there are other options to consider, such as our hands, our voices, and our body movements. These modalities can be used individually or combined, by a person or a group. Touch supports a *natural* experience, as it enables direct manipulation of the data or its representation through the fingers. However, it suffers from the *fat-finger problem* which refers to providing less spatial accuracy while interacting (Drucker et al., 2013). Similar to touch, using a handheld pen (stylus) is also considered a form of direct manipulation that tends to be more accurate than touch, due to the physical structure of the pen. However, this type of interaction requires the pen as an additional input device. Using our voice to perform speech commands in natural language is another option that allows to directly express our intentions. However, the common appearance of recognition errors for speech interaction with commercial voice assistants often leads to discarding this modality, due to the loss of trust. Another modality for distant interaction is mid-air gestures, which has proved to be successful in combination with proxemics, i.e., the use of body movements (Badam et al., 2016). However, using gestures continuously can be physically demanding, and it requires a tracking system integrated into the visualization system.

Overall, each interaction modality has advantages and disadvantages. The research on multimodal interaction (i.e., the combination of two or more modalities) with data visualizations has so far focused on scenarios of single users. While some combinations of interaction modalities have been studied in this context, we do not yet know how we should map or combine specific interaction modalities with different exploratory tasks, and even further, collaborative work. Through my dissertation, I aim to extend the research on multimodal interaction design in different contexts for visual data exploration and sensemaking, considering the needs of individuals and groups in professional settings. I build upon previous work by investigating how different interaction modalities can support and change collaborative work. Accordingly, my research aims to help us better understand how designers of interactive systems could make use of different interaction modalities for specific tasks, investigating the challenges and opportunities of the modalities and their

*The terms *interaction modality* and *input modality* are often used interchangeably. I use *interaction modality* to differentiate it clearly from the input device.

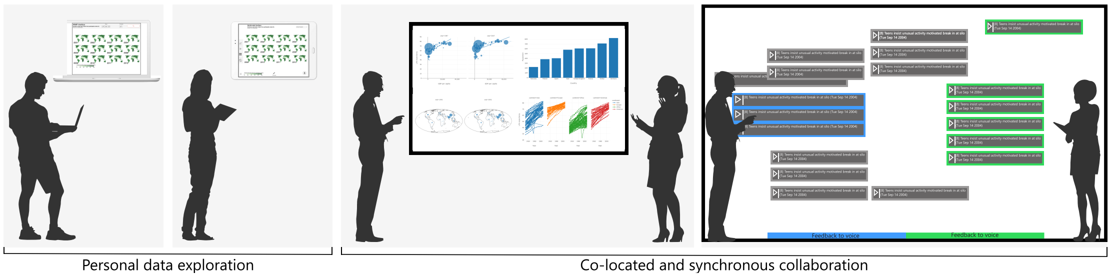


Figure 1.1: The human-data interaction scenarios covered in this dissertation: (a) Personal data exploration scenario on desktop computers and (b) on tablets, (c) collaborative exploration on large vertical displays (LVDs), and (d) collaborative sensemaking on wall displays (second example of LVD).

combinations.

This dissertation is anchored in the co-creation of technology. Co-creation is a design methodology that stems from the tradition of participatory design and refers to the deep involvement of the end-users to apply their collective creativity throughout the design process, emphasizing the collaboration between people from different domains (Sanders and Stappers, 2008). As my goal was to design interactive systems for the work environment of data experts, I collaborated with researchers who interact with data in their everyday work lives to investigate multimodal and collaborative interaction in a real-world context. The starting point of my dissertation is a design study with social policy researchers (as an example of real-world data experts) interested in co-designing an interactive data exploration system. We worked together in a collaborative research center whose goal was to provide new data on the evolution of social policies worldwide. Given that participation in the design of technology is often restricted to a one-time consulting, this collaboration was a unique opportunity to deeply involve the experts and encourage them to become co-owners of the design (Delgado et al., 2023).

Accordingly, the data and tasks I focused on are part of the outcomes of the design study with social policy researchers. Their design requirements for interactive data systems led me to work with spatio-temporal data and the representations and visualizations suitable to explore said data. In terms of collaboration, I conducted studies on co-located and synchronous work as the researchers organized themselves in small groups to collect and analyze data about specific topics within their research field.

1.1 Thesis Statement

This dissertation investigates how different devices and interaction modalities can support individuals and groups in visual data exploration and sensemaking. In Figure 1.1, I present the different interactive scenarios covered in this dissertation. I started by researching what kind of data visualizations can support domain experts in exploring spatio-temporal data. I applied the co-creation methodology with domain experts to have a solid human-centered base on data and visualizations to focus on. Then, I shifted the focus to how interaction can support the experts. I started by comparing how their experience differs between working on a desktop computer and a tablet. Later, I examined group work as an essential part of data work and investigated how group members make different interaction choices depending on the task and the visualization at hand. I elicited interaction design requirements from the end-users and conducted an elicitation study to learn more about their interaction preferences. The human-centered methods I used informed the design of two interactive systems I implemented for visual data exploration and sensemaking.

1.2 Research Scope

The scope of this dissertation lies within the intersection of the research fields of human-computer interaction (HCI) and visualization (VIS), partly zooming into computer-supported cooperative work (CSCW). Figure 1.2 presents an overview of the related topics within those fields. From a VIS perspective, I focus on supporting domain experts in conducting exploratory and sensemaking tasks on visualizations of spatio-temporal data. While most of my work is within the sub-field of information visualization (InfoVis), I also consider visual analytics (VAST) in Chapter 6. From an HCI perspective, I focus on investigating the use of interaction modalities on interactive surfaces, using tablets as an example device to support individual work, and large vertical displays (including wall-sized displays) as an example for collaborative work. From a CSCW perspective, I narrow down the collaboration scenario to co-located and synchronous work, and I apply co-creation as a participatory approach to designing tools for domain experts. While CSCW emerged as a sub-field of HCI, nowadays they are considered overlapping fields (Fitzpatrick and Ellingsen, 2013).

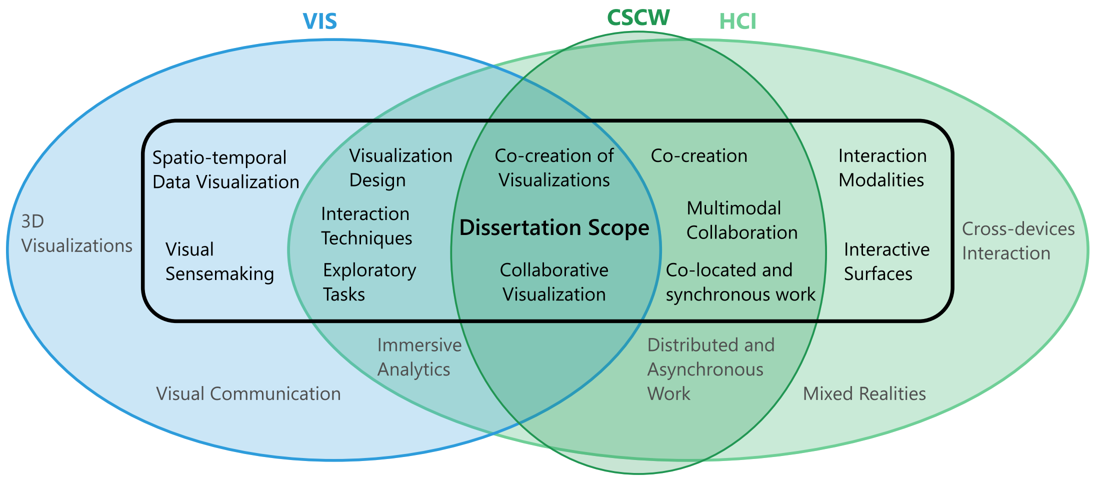


Figure 1.2: The research scope of the dissertation is at the intersection of the research fields of visualization, human-computer interaction, and computer-supported cooperative work. The topics in gray are related but not within the scope. The oval sizes were simplified for the figure and are not meant to indicate the proportional size of the fields.

Visualization of spatio-temporal data I focus on spatio-temporal data and the design of 2D visualizations that can support domain experts in exploring said data, as a result of collaborating with social science researchers. I selected the visualization techniques (e.g., line chart, choropleth map, scatterplot matrix, etc.) based on findings on how to explore spatio-temporal data, taking strategies such as animation and small multiples into account (e.g., Andrienko et al., 2003b). Authoring and designing new visualizations was not within the scope.

Data exploration and sensemaking I focus on exploration and sensemaking tasks, closely connected to the case study with social science researchers. According to the data type, I designed experiments and developed a custom system to support exploratory tasks with spatio-temporal data. Then, I extend the scope to sensemaking in the context of collaborative work, to also investigate how people interact while foraging for data and communicating findings (Card et al., 1999).

Interaction modalities My work investigates what interaction modalities can support data experts in their work, focusing on the combination of different interaction modalities to complement each other (Saktheeswaran et al., 2020). I examine user preferences, interaction choices, and patterns with touch gestures, speech commands, pen interaction, and mid-air gestures on interactive surfaces, in comparison to mouse and keyboard interaction on desktop computers. I leverage these modalities to facilitate interaction techniques related to exploration and sensemaking (e.g., lookup, selection, search). Other relevant modalities such as proxemics and gaze interaction are not included to keep the scope manageable.

Interactive surfaces My work investigates how different interactive surfaces can support data exploration and sensemaking tasks *beyond the desktop* (Lee et al., 2012). My contributions include design recommendations for interactive visual tools on tablets, large vertical displays, and wall-sized displays, to support individual and collaborative interaction. My research examines scenarios where people only make use of physical space. Other environments that introduce virtual elements, such as augmented and virtual reality, are out of the scope.

Co-located and synchronous work After investigating single-user data exploration, my research looks at collaboration by groups of two people (pairs) in the same location (co-located) and at the same time (synchronous), according to the classification of Johansen (1988). While some findings could transfer to other scenarios, I focus on pairs working together with a single display, and I investigate how using specific interaction modalities (alone or in combination) influences their collaboration.

1.3 Research Questions

Considering the benefits of human-centered design methods and my focus on designing tools for professional workplaces, the first goal of my dissertation is to involve real-world experts in the design of interactive tools to support their work, and to determine what tasks are relevant for them. As the experts will be the end-users, it is important to ensure their involvement and assessment of the designed solutions. Hence, I worked closely with social science researchers to support them in their work with social policy data over several years. Accordingly, my first research question is:

RQ1 What data and tasks are relevant to support real-world experts in their data-driven work through the design of interactive visualizations?

My second goal was to extend the research on interaction design for data exploration to devices *beyond the desktop*. Although, for instance, tablets already support specialized applications such as Tableau Mobile and Power BI, we do not yet know how conducting data exploration on them differs from the standard process in desktop computers. Previous research has investigated how to design touch-based visualizations on tablets (Baur et al., 2012; Drucker et al., 2013; Sadana and Stasko, 2016) and I build upon that work to understand how we could leverage multiple interaction modalities to perform exploration tasks. As tablets have a smaller physical screen size than the desktop computer and support other interaction modalities than the mouse and keyboard, it is important to investigate how the performance and experience of the users differ in comparison to the classical desktop setup to learn how we could leverage the capabilities of each device in the design of responsive interactive systems. Therefore, the second research question asks:

RQ2 How do different devices and interaction modalities affect performance and user experience during data exploration tasks?

The following research questions consider individuals and groups, as data work is often part of collaborative work. For instance, data science projects are highly collaborative, and organizations create data science teams to explore large datasets, combining different professional skills (Zhang et al., 2020). How an individual explores data to learn from it may differ from how a team explores the same dataset. Accordingly, this dissertation also aims to extend the body of research on interaction design for collaborative visualization in the context of co-located and synchronous work (i.e., when at least two people work together in one location at the same time). While research on post-WIMP interaction has mainly focused on single-user scenarios so far, there are devices, such as large vertical displays, that can support multiple interaction modalities and are meant for collaborative work in meeting spaces. The standard mouse and keyboard input devices are not ideal for large vertical displays due to the screen size and the varying distance between the users and the display. Moreover, the way people work together and move in the 3D space in front of the screen may influence what interaction modalities and combinations may be appropriate for specific tasks. To find answers, this dissertation addresses the following research question:

RQ3 What interaction modalities do end-users favor for visually exploring data, individually and collaboratively? What modality combinations are preferred?

After determining user preferences about interaction modalities, it is necessary to empirically evaluate them with an implemented solution to assess whether and how the corresponding interaction design can support collaborative work. To extend the range of devices considered at the workplace for collaborative work, this dissertation also takes wall-sized displays into account, which are specifically designed for groups to make sense of large datasets. In that context, the final research question of the dissertation is:

RQ4 How do different interaction modalities (and their combination) affect collaborative work? How do they relate to collaboration styles?

Figure 1.3 shows the connection between each research question and the corresponding dissertation chapters. I covered the scope of my research questions by first selecting an application scenario to focus on, and then, investigating and proposing different ways of interacting with data on devices and workspaces relevant to that scenario. Afterward, I examined more general but related scenarios to draw conclusions. The first phase of my dissertation research consisted of conducting a design study with real-world experts who explored data daily as part of their job, i.e. social science researchers. Social scientists are increasingly working with datasets provided by governments and international organizations to conduct their research on social and political phenomena, and are in the process of adopting advanced computational methods for data exploration. Therefore, they are ideal partners to collaborate with to design interactive data tools. The co-creation design study of Chapter 3 presents how I applied the co-creation methodology to involve and empower the domain experts during the design process of visualizations suitable to their work. This design study corresponds to answering **RQ1**, where I evaluated using co-creation as a visualization design methodology. The human-centered approach involved conducting surveys, interviews, and workshops to understand the workflow and needs of the scientists. Another goal was to understand what data and tasks are relevant in a real-world data-driven scenario and what kind of devices are appropriate for supporting individual and collaborative work in that context. After defining the data, tasks, and suitable visualizations to work with, I conducted a series of empirical experiments to investigate how leveraging different interaction modalities could support their work on other devices than desktop computers.

The findings from the design study served as the base of the first empirical study, comparing two versions of a visualization system, implemented for exploring data on desktop computers and tablets, as these two devices were already used by the experts at their workplace. Based on the capabilities of each device, the system supported interaction techniques performed with a mouse and a keyboard, and equivalent interactions with touch, pen, and speech input. In Chapter 4, I compared these two interactive workspaces to answer **RQ2**, by studying how the combination of devices and modalities affected the performance and user experience of the experts during the visual data exploration. The comparative study also investigated the interaction strategies of the experts across devices and modalities.

While the comparative study focused on personal exploration, I proceeded to investigate collaborative work in the second and third empirical studies. In Chapter 5, I examined how people wish to interact with data on a large vertical display, regardless of any limitations imposed by current technologies. I applied a human-centered approach to answering **RQ3** by conducting an elicitation study where the interaction design ideas came from the participants. Participants could propose interactions with either touch, speech, pen, or mid-air gestures, or a combination of them, performed by one or two persons. In contrast to the previous experiment, I went beyond the case study with the social science researchers and aimed to work with any professionals who perform data exploration in their everyday work lives, taking both the individual and the collaborative scenarios into account. This combination served as a preparation for answering **RQ4** in the third empirical study. In the final study, I built upon the outcome of the elicitation study and continued to investigate collaborative work by examining how the elicited interactions could be used in a real-world system on a wall display, designed to integrate the modalities and interactions proposed by the data experts of the previous study. In this final phase, I focused specifically on touch and speech as interaction modalities, given their dominance in the results of the elicitation study. In the following section, I describe the main research contributions that resulted from these studies.

1.4 Methodological Approach

I applied a variety of research methods to answer the formulated research questions. These include quantitative methods such as statistical analyses and questionnaires, as well as qualitative methods such as observation and interviews. In the following, I describe the methods according to the corresponding objectives.

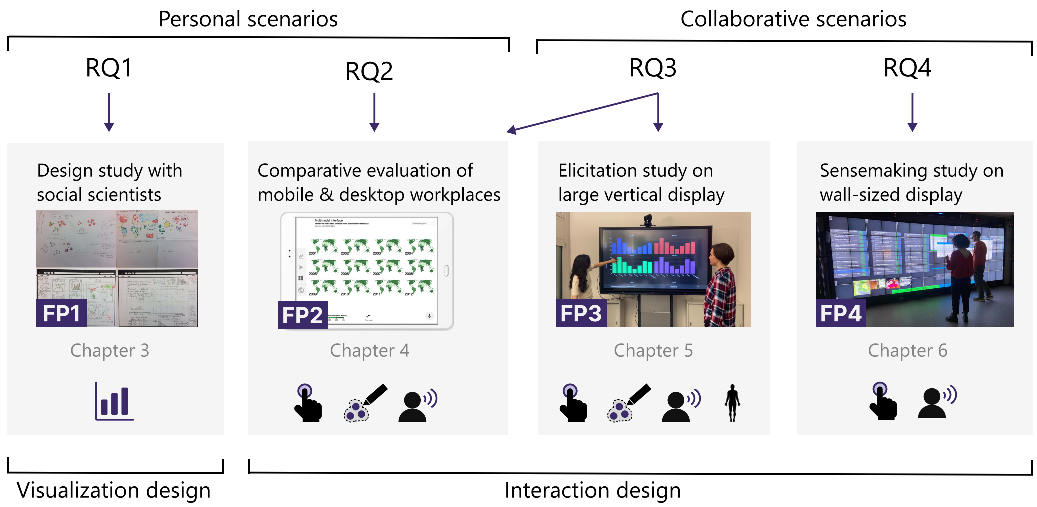


Figure 1.3: Dissertation structure around four main research questions (RQ1-4) and four projects developed to answer them, presented as full papers (FP) in Chapters 3–6.

1.4.1 Visualization and Interaction Design

In Chapter 3, I conducted a visualization design study with social policy experts. I applied the co-creation methodology, to elicit the design requirements and empower the experts to shape the design of visualization tools meant to support their work. The design process included co-creation workshops, surveys, semi-structured interviews, and a user study. In the workshops, we used design methods such as wishful thinking, paper prototyping, and reflective discussions (Kerzner et al., 2019). I conducted a formative and summative evaluation of co-creation as a visualization design methodology. The formative evaluation comprised a survey, semi-structured interviews, and a reflective group discussion, while the summative one was the user study. In that study, we also observed how participants interacted with the prototype and asked them to think aloud.

To abstract the data and tasks of the experts, I followed the analysis framework of Munzner (2014). To operationalize the visual data exploration process, I applied a task-centered approach and carefully selected the exploratory tasks of the experiments according to existing task typologies, which cover the exploration of spatio-temporal data (see more details in Chapter 2).

In Chapter 5, I conducted an elicitation study to investigate preferences regarding multimodal and collaborative interactions. As elicitation studies help reveal user preferences in a hypothetical situation without technology limitations, I applied the Wizard of Oz technique to simulate how the device would react to the interactions. For designing the *Modality Explorer* system presented in Chapter 4, I surveyed previous research on multimodal visualization systems and commercial products that involved the visualization techniques relevant to our study that were not present in the research work.

1.4.2 Prototyping

I applied paper prototyping in the co-creation workshops and software prototyping in all projects. Software prototyping was necessary to investigate multimodal interaction because speech commands are rarely supported on commercial visualization systems, and pen input is often considered equal to touch without distinction. Although time-consuming, the development of the web-based artifacts, *Modality Explorer* and *TouchTalkInteractive*, allowed me to evaluate the relevant interaction techniques with all modalities. For the iterative development of the *Modality Explorer*, I conducted an expert review with two HCI researchers and a usability study. The system involved in the final user study of Chapter 3 was the first prototype of WeSIS, the Global Welfare State Information System, which will be released to the social policy research community in late 2024. The latest version is the product of the collaborative work of researchers and developers from the CRC 1342 (2023).

1.4.3 Controlled experiments

I conducted four controlled experiments: the user study at the end of the co-creation process in Chapter 3, a within-subjects experiment to compare performance and user experience across two interactive workplaces in Chapter 4, the elicitation study in Chapter 5, and the collaborative study of Chapter 6.

In Chapter 4, I conducted a statistical analysis of the response time and accuracy metrics to test our hypotheses regarding significant differences in performance, including their relation to task type and tablet ownership. In Chapter 5, I analyzed the interaction proposals of the elicitation study to determine the consensus set (i.e., the set of interaction proposals with the highest agreement) and the *max-consensus* and *consensus-distinct ratio* metrics. In Chapter 6, I statistically analyzed the relationship between the use of speech interaction and personality traits.

In the studies of Chapters 4 and 5, I performed log analyses to study the interaction choices of the participants regarding task, modality, and visualization technique. The analyses helped identify interaction patterns and strategies across devices.

In each study, I used questionnaires with Likert scales to collect subjective feedback, such as user preferences. For measuring user experience in Chapter 4, I used the SUS questionnaire. I conducted semi-structured interviews in Chapter 3 to understand the workflow of the researchers and in Chapter 6 to obtain qualitative feedback about the participant experiences. I recorded and analyzed videos in Chapters 4–6 to determine interaction strategies, document interaction proposals, and understand collaborative work (qualitative analysis). In all experiments, participants gave their informed consent (in line with the GDPR), and we followed the requirements regarding the rights and privacy of the participants.

1.5 Research Contributions

While each of the publications involved in this dissertation has its own contributions to the fields of VIS, HCI, and CSCW (presented in Chapters 3–6), I present the main contributions of the dissertation in this section.

1.5.1 Data and task abstractions from social policy researchers for visualization design, together with the evaluation of co-creation as a visualization design methodology

The first contribution of the dissertation refers to the two main outcomes of conducting a design study with social science researchers, as an example of data experts, and of applying the co-creation methodology in the study. The first outcome is abstracting the data and tasks of the researchers through the co-creation design study. We contribute a detailed characterization of the visualization requirements of social policy researchers and, following the analysis framework of Munzner (2014), the data and task abstractions of the researchers derived from the requirements. We describe the spatio-temporal data and exploratory tasks that we will focus on to investigate multimodal and collaborative interactions. The second outcome is a methodological contribution: the first formal evaluation of co-creation, a participatory methodology, for visualization design. We provide a series of recommendations for applying co-creation in visualization design.

1.5.2 Two multimodal systems: the *Modality Explorer* for individual data exploration and *TouchTalkInteractive* for collaborative sensemaking

The second contribution is an artifact contribution (Wobbrock and Kientz, 2016) and comprises two interactive systems designed and implemented to support visual data exploration and sensemaking. Given that multiple interaction modalities are rarely supported on off-the-self systems, I developed custom prototypes to investigate my research questions and to demonstrate how multimodal systems can be designed to facilitate individual data exploration on tablets, and collaborative sensemaking on wall displays. The first is *Modality Explorer*, a web-based visualization system designed to support different types of exploratory tasks for spatio-temporal data through touch, pen, and speech interaction on tablets, with equivalent interactions using the mouse and keyboard on desktop computers. The second is *TouchTalkInteractive*, a web-based visual analytics system that supports touch and speech interactions elicited from end-users, to assist pairs of users in co-located visual sensemaking on a wall-sized display. Both systems integrate carefully selected interaction techniques, based on a survey of consistent interaction techniques for spatio-temporal data and the elicited consensus set from Chapter 5. The *Modality Explorer* was designed for the comparative study (Chapter 4) and *TouchTalkInteractive* was based on the elicitation study and developed for the final collaborative study (Chapter 6). Both systems are open-source and publicly available to guarantee the reproducibility of the corresponding research projects (see Chapters 4 and 6 for details).

1.5.3 Characterization of interaction patterns and strategies on multimodal visual systems across devices

The third contribution results from the quantitative and qualitative findings from comparing desktop and mobile workplaces (Chapter 4), and the collaborative sensemaking study (Chapter 6). The video analyses and interaction logs, together with the questionnaires and performance measurements, helped to understand how the experts explored and made sense of the data, individually and collaboratively. We contribute the characterization of different strategies to solve exploratory tasks, comparing desktop and tablet-based workplaces for personal work. Participants were significantly faster but not significantly more accurate on the desktop than on the tablet and chose specific modalities and views for solving different tasks. Through an interaction analysis, we found that the smaller screen size of the tablet did not lead them to zoom more often but rather to approach the tasks differently (e.g., favoring maps and bar charts instead of line charts to solve tasks related to temporal development). In the context of collaborative work, we extend the understanding of how people interacted during different collaboration styles, using touch and speech interaction. We found that speech is a viable option for performing global tasks at a distance, and has the potential to contribute to awareness within the team.

1.5.4 Modality preferences across devices for visual data exploration and sensemaking in individual and collaborative scenarios

The fourth contribution extends our understanding of interaction modality preferences based on the findings of the three empirical studies. In the single-user scenario, participants preferred using the tablet-based interface over the WIMP interface for visually exploring data and appreciated pen input on the tablet due to its high precision. While the pen was used most for all tasks, the qualitative feedback revealed that speech commands were preferred to search for data items whose location on the screen was unknown. In terms of collaborative work, we contribute the consensus set as well as the analysis of the associated 1015 interaction proposals from the elicitation study (Chapter 5). The consensus set revealed that people prefer unimodal and personal interactions (one modality, one person) over multimodal or collaborative interactions when touch, pen, speech, and mid-air gestures are available. Speech was the preferred modality, and in multimodal proposals, participants often combined touch gestures with speech commands in a sequence. When interacting collaboratively, the pairs interacted either unimodally with simultaneous actions or multimodally in sequence. We introduce the concept of *collaborative synonyms* to describe distinct interactions that are identical except for being done by one or two people. Furthermore, in the sensemaking study, we found that participants preferred speech commands for performing global tasks, such as sorting, searching, and selecting all elements. We also found that the likelihood of using speech commands relates to personality traits.

1.5.5 Design considerations for facilitating multimodal and collaborative interaction on interactive surfaces

The fifth and final contribution involves the design considerations derived from the findings of the comparative evaluation, the elicitation study, and the collaborative study. Overall, participants appreciated the expressiveness of multimodal interaction but favored unimodal interactions per task. Our findings contrast with previous work suggesting the preference for multimodal over unimodal interactions (Saktheeswaran et al., 2020). Multimodal systems for visual data exploration on tablets should support all primary interactions with the pen, while still leveraging touch for equivalent actions, as well as speech commands for tasks with no or unknown spatial component. During collaborative work, participants leaned towards assigning modalities to each other and standing at different distances from the display. Speech interaction was favored for interacting from a distance and performing global tasks, while touch was leveraged for most other tasks. For visual exploration on large vertical displays, participants required touch, pen, and speech interactions, favoring touch for manipulating views, pen for selection and annotations, and speech commands for the remaining tasks. Through the interaction analyses, we found that the pen was the most used modality on the tablet, while touch gestures were most used on the wall display, but speech commands were favored for global tasks. We argue, based on related research results on Drucker et al. (2013) that choosing a modality depends on the visualization technique and actions intended, as well as on the device. Moreover, the personality traits of the person may predict which modality they choose to interact with.

1.6 Outline and Publications

This cumulative dissertation contains previously published content from peer-reviewed publications. The four main manuscripts are presented in Chapters 3–6 and reproduced without changes, except for table adjustments. Figure 1.3 presents how those chapters connect to the research questions. In the following, I present how the dissertation is organized together with the connection to the research questions, contributions, and papers. Although I am the first author of the papers and led all the projects, the dissertation is the result of my collaboration with my supervisors, as well as with fellow researchers from multiple institutions. Thus, I use the form "we" to refer fairly to the research I conducted together with my co-authors in the corresponding chapters. I will describe what the contribution of each author was at the beginning of each chapter.

Chapter 2 reviews the literature relevant to our work on visualization and interaction design, multimodal interaction, and collaborative work. The chapter highlights the research gaps the dissertation addresses. Chapter 3 presents the design study with the social science researchers and answers **RQ1**. While the collaboration continued over the next years and covered multiple topics (e.g., interface design), Chapter 3 covers a one-year period focused on visualization design. There, we introduce the co-creation methodology, its potential benefits for visualization design, and the lessons learned from the collaboration with the social scientists. The chapter presents the first research contribution on co-creating data visualizations (subsection 1.5.1) and was published as a peer-reviewed paper in the Computer Graphics Forum journal, accepted at the EuroVis 2020 conference:

Full paper 1 Gabriela Molina León and Andreas Breiter. "Co-creating Visualizations: A First Evaluation with Social Science Researchers". In: *Computer Graphics Forum* 39.3 (2020), pp. 291–302. DOI: 10.1111/cgf.13981.

I initiated the research project with Andreas Breiter, as part of our interdisciplinary collaboration with social scientists. We designed together the co-creation process, I selected the workshop methods, and led the workshops. I conducted both surveys, all interviews, and the user study. I performed the qualitative analyses and wrote the paper. Andreas Breiter provided detailed feedback on the draft.

Beyond the four papers included in the dissertation chapters, I led other peer-reviewed publications derived from their findings. There were two direct follow-up from the co-creation design study presented in Chapter 3. First, we conducted an experiment investigating the visualization design choices of the domain experts for choropleth maps, as this visualization technique was relevant to identify geographical patterns. The experts wished to emphasize the collected data from the Global Souths[†], as those world regions are often missing in welfare studies (Schmitt et al., 2015). Our experiment revealed that the map projection was the most important design choice to achieve that and the Equal Earth projection (Savric et al., 2019) was the preferred projection. The resulting short paper was presented and published at the IEEE VIS 2020 conference:

Short paper 1 Gabriela Molina León, Michael Lischka, and Andreas Breiter. "Mapping the Global South: Equal-Area Projections for Choropleth Maps". In: *2020 IEEE Visualization Conference (VIS)*. 2020, pp. 91–95. DOI: 10.1109/VIS47514.2020.00025.

[†]I pluralize "South" to intentionally highlight the diversity in the regions usually named Global South (Citational Justice Collective et al., 2023).

In addition, we presented and published a poster at the ECSCW 2022 conference discussing the interdisciplinary and collaborative context of the study and sharing our recommendations for similar data infrastructure projects:

Poster 1 Gabriela Molina León, Gabriella Skitalinskaya, Nils Düpont, Jonas Klaff, Anton Schlegel, Hendrik Heuer, and Andreas Breiter. "Co-Creating a Research Data Infrastructure with Social Policy Researchers". In: *Proceedings of 20th European Conference on Computer-Supported Cooperative Work*. European Society for Socially Embedded Technologies (EUSSET), 2022. DOI: 10.48340/ecscw2022_p03.

Returning to the dissertation chapters, Chapter 4 reports on the comparative study of two interactive workplaces to answer **RQ2** and to answer **RQ3** for the single-user scenario. From this study, we learned how people apply different strategies to exploratory tasks on desktop and tablet computers with different interaction modalities (mouse, keyboard, touch, pen, and speech). The chapter presents the first of the two implemented systems (subsection 1.5.2) and contributes the first findings to understand how the interaction patterns, performance, and user experience of domain experts vary across workspaces while exploring data individually. The content was published as the following peer-reviewed paper:

Full paper 2 Gabriela Molina León, Michael Lischka, Wei Luo, and Andreas Breiter. "Mobile and Multimodal? A Comparative Evaluation of Interactive Workplaces for Visual Data Exploration". In: *Computer Graphics Forum* 41.3 (2022), pp. 417–428. DOI: 10.1111/cgf.14551.

As the first author, I led this research project and implemented the system. Michael Lischka contributed to the system design, with his perspective as geographer and social scientist, and he conducted the study together with me. Wei Luo contributed to the final version of the system interfaces. Andreas Breiter contributed to the research questions and the experiment design. I performed the analyses of the study results and wrote the paper. Michael Lischka and Andreas Breiter provided detailed feedback on the draft.

Chapter 5 presents the elicitation study where we examined user preferences for interacting with data visualizations on a large vertical display, given the choice of interacting with any of four modalities (touch, pen, speech, and mid-air gestures). With this study, we considered not only single-user scenarios but also collaborative ones to address **RQ3**, which led to the identification of user preferences (subsection 1.5.4) and design considerations according to the task and number of persons involved (subsection 1.5.5). The publication on which this chapter is based was published in the IEEE Transactions on Visualization and Computer Graphics journal and has been awarded the Replicability Stamp of the Graphics Replicability Stamp Initiative (GRSI).

Full paper 3 Gabriela Molina León, Petra Isenberg, and Andreas Breiter. "Eliciting Multimodal and Collaborative Interactions for Data Exploration on Large Vertical Displays". In: *IEEE Transactions on Visualization and Computer Graphics* 30.2 (2024), pp. 1624–1637. ISSN: 1941-0506. DOI: 10.1109/TVCG.2023.3323150.

Regarding author contributions, I designed and conducted the elicitation study, and performed the quantitative analysis of the results. Andreas Breiter contributed to the experiment design, and Petra Isenberg contributed to the introduction, background, and discussion sections. The three of us interpreted the study results. I wrote the first complete draft of the paper and led the revisions.

Following the elicitation study, I shared my reflections on how to leverage multimodal interaction for inclusive design in the context of conversational user interfaces (CUI) on a paper presented at the CUI workshop of the CHI 2023 conference:

Workshop paper 1 Gabriela Molina León. "Advancing Inclusive Design in Multiple Dimensions". In: *CHI Workshop on Inclusive Design of CUIs Across Modalities and Mobilities*. 2023. arXiv: 2303.16790 [cs.HC].

Given the dominance of speech commands in the interaction proposals of the elicitation study, we reflected on the design challenges and opportunities associated with speech interaction. We presented those reflections on a paper at the MERCADO workshop of VIS 2023.

Workshop paper 2 Gabriela Molina León, Petra Isenberg, and Andreas Breiter. "Talking to Data Visualizations: Opportunities and Challenges". In: *VIS Workshop on Multimodal Experiences for Remote Communication Around Data Online*. 2023. arXiv: 2309.09781 [cs.HC].

After reflecting on the results of the elicitation study, we conducted a second collaborative study, this time on a wall display. In Chapter 6, we present the design of the system to support collaborative sensemaking, which is the second implemented prototype (subsection 1.5.2). The design is based on the elicited interactions from Chapter 5 to assess how different interaction modalities affect collaborative work (**RQ4**). We adapted the interaction techniques supported according to the necessary actions for the task and focused on touch and speech as they were the dominant modalities of the elicitation study. The study extended the findings on interaction patterns across devices (subsection 1.5.3), user preferences (subsection 1.5.4), and design considerations for multimodal and collaborative systems (subsection 1.5.5). The chapter is based on the following article to be submitted soon.

Full paper 4 Gabriela Molina León, Anastasia Bezerianos, Olivier Gladin, and Petra Isenberg. *Talk to the Wall: The Role of Speech Interaction in Collaborative Work*. To submit. 2024.

This research project was mainly developed during my research stay at Inria Saclay — Île-de-France, hosted by Petra Isenberg. Together with Anastasia Bezerianos, the three of us developed the research questions and designed the experiment. I led the system design and conducted the study. Olivier Gladin contributed to the system implementation. I performed the statistical and qualitative analyses of the study results. Anastasia Bezerianos, Petra Isenberg, and I discussed, and interpreted the results. I was the lead writer of the paper, while the others performed optimizations and individual additions.

Chapter 7 concludes the dissertation, revisiting and discussing the research contributions, providing final design considerations for multimodal and collaborative interaction, as well as pointing to three future research directions on interaction design and collaborative work.

2
Research Background

The research contributions of this dissertation build upon previous work on visualization design, interaction design for visualizations, and co-located collaborative work. This chapter presents relevant work on these topics to situate the research projects presented in Chapters 3–6. While each of those chapters contains related work specific to that project, this chapter presents fundamental concepts and additional literature relevant to the complete dissertation, while referring to the related work sections of other chapters when pertinent.

2.1 Visualization Design

This dissertation engages with visualization research, covering both the design process of visualizations and the design of interactions for visualization systems. Different definitions of visualization exist. According to Ware (2004), visualization refers to "a graphical representation of data or concepts". Card et al. (1999) define it as "interactive visual representations of data to support human cognition". However, the more recent definition of Munzner (2014) fits best to the scope of my work:

> Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

This definition mentions two important aspects that set the scope of my research. It refers explicitly to helping *people* carry out *tasks*, as my work aims to help experts carry out tasks associated with data exploration and sensemaking.

The visualization research field has traditionally been separated into three focal points according to specific topics and former conferences: Information Visualization (InfoVis), Scientific Visualization (SciVis), and Visual Analytics (VAST) (Isenberg et al., 2017). Information visualization focuses on abstract and non-physical data (e.g., total of daily website visits), scientific visualization focuses on physical data (e.g., brain imaging), and visual analytics, although often confused with InfoVis, is the younger sub-field focusing on analytical tasks with large datasets (Cook and Thomas, 2005; Keim et al., 2008). My dissertation focuses mainly on InfoVis (Chapters 3, 4, and 5), extending its scope towards VAST in Chapter 6.

2.1.1 Visualization Design Studies

The starting point of this dissertation is a design study to develop visualization solutions for a specific domain. Sedlmair et al. (2012) define a visualization design study as:

> A project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines.

This type of problem-driven research is only possible through collaborating with real-world experts, such as the social policy researchers I worked with, in order to analyze their needs and design appropriate solutions. Such design studies require models and methodologies to guide the process (Meyer and Dykes, 2020). I apply the analysis framework of Munzner (2014) that asks three questions to systematically explore the visualization design space:

- *What* is shown? (data)
- *Why* are users looking at the visualization? (tasks)

- *How* is it shown? (visualization techniques)

I answer these questions in Chapters 3 and 4. In Chapter 3, I present the design study with the domain experts, describing their workflow and design requirements, together with the lessons we learned. In Chapter 3, I examine the data and tasks further and evaluate different interaction techniques based on the appropriate visualization techniques. Except for the work of Henry and Fekete (2006b) on creating network visualizations for social scientists and the work of Coelho et al. (2020) on designing a system to analyze dominance hierarchies, there was little visualization design research with social sciences as an application domain previous to our study.

Through the design process, we learned that all the social policy indicators (i.e., datasets) the researchers were working on are spatio-temporal data, i.e., policy values with spatial (country) and temporal (year) components. These two variables represent the referents, while the main value of the indicator was the abstract attribute (e.g., type of maternity leave). Through the co-creation workshops, interviews, and surveys, we abstracted their tasks successfully and used the task typology of Andrienko and Andrienko (2006) to analyze them and map them to appropriate visualization techniques. Therefore, I operationalize the data exploration process with a task-centered approach. I selected carefully the exploratory tasks of the empirical studies according to the task typologies of Andrienko and Andrienko (2006) and Brehmer and Munzner (2013), combined with the interaction taxonomy of Yi et al. (2007). Section 4.2.1 presents further related work on the exploration of spatio-temporal data.

2.1.2 Co-creation of Visualizations

According to the Interaction Design Foundation - IxDF (2021), co-creation is "the practice of collaborating with other stakeholders to guide the design process". Sanders and Stappers (2008) consider that co-creation is "any act of collective creativity" and co-design is "collective creativity as it is applied across the whole span of a design process". While the terms *co-creation* and *co-design* are often used interchangeably, all the definitions suggest that co-creation refers to applying the creativity of a group to a product of value; in our case, a visualization solution. Derived from participatory design, co-creation has been proven to successfully support the design of technical solutions in multiple domains, such as learning analytics (Dollinger and Lodge, 2018) and digital public information services (Jarke, 2021).

In the context of software development, co-creation refers to the deep involvement of future users not only in the requirement elicitation but also in the design decision-making process (Delgado et al., 2023).

For visualization design, Jänicke et al. (2020) suggest applying participatory approaches based on a series of visualization use cases, arguing that the acceptance of visualization projects depends, at least partly, on the degree of involvement of the domain experts. While there have been multiple design studies involving co-creation and creative methods in visualization design (e.g., Landstorfer et al., 2014a; Lloyd and Dykes, 2011), there has been no formal evaluation of co-creation as a visualization methodology. Chapter 3 presents the formative and summative evaluation of co-creation as part of our design study, in the form of a methodological contribution. More related work on co-creating visualizations is presented in Section 3.2.

2.2 Interaction Design for Visualization

As the complexity and the amount of data increases, being able to interact with its visual representation becomes more relevant. Interaction makes visualizations flexible and allows people to examine the data from different perspectives (Tominski and Schumann, 2020). Good interaction maximizes the directness, simplicity, and naturalness of the dialogue between the person and the computer (Hornbæk and Oulasvirta, 2017). Dimara and Perin (2020) define interaction for visualization as:

> The interplay between a person and a data interface involving a data-related intent, at least one action from the person and an interface reaction that is perceived as such.

Interaction techniques allow us to modify the visualization or query new data according to our goal and task (e.g., panning to make new data points visible). In this dissertation, I focus first on the design of interactive visualization systems that support visual data exploration by individuals. According to Keim (2002),

> Visual data exploration aims at integrating the human in the data exploration process, applying its perceptual abilities to the large data sets available in today's computer systems.

Accordingly, I conducted a series of experiments to determine how to provide diverse interaction techniques to support people in exploring their data visually.

2.2.1 Interaction beyond the desktop

As the diversity and performance of non-desktop computers have increased (e.g., smartphones, tablets, tabletops, wall displays, head-mounted displays, etc.), there is a need to develop visualization systems that people can interact with on these devices. In 1993, Nielsen (1993) argued for designing user interfaces that do not limit themselves to using WIMP elements (windows, icons, menus, and pointers). However, in 2012, Lee et al. (2012) considered that most visualization research was still based on the assumption that interaction would happen on desktop computers with a WIMP interface. Thus, the authors encouraged the community to think of post-WIMP interfaces (also known as Natural User Interfaces) considering multiple dimensions, such as input type, interaction distance, and collaboration context. My dissertation aims to address this gap, at least partially, by investigating different interaction modalities with devices that support variable interaction distances, in the context of individual and collaborative work (Chapters 4–6).

Non-desktop devices come with a set of challenges and opportunities according to the size and capabilities of the display. While smartphones, smartwatches, and tablets have a smaller display to present the data, they provide touch interaction and possess multiple sensors to leverage for interacting, besides allowing people to interact on the go. Accordingly, researchers have created custom visualizations for tablets (Baur et al., 2012; Sadana and Stasko, 2016), designed interaction techniques for interacting on smartphones (Eichmann et al., 2020; Kim et al., 2021), and investigated performance and preferences on smartwatches and fitness trackers (Blascheck et al., 2019; Islam et al., 2022). We build upon their work to compare the performance and user experience between tablet-based and desktop-based visualizations in Chapter 4. While Drucker et al. (2013) compared WIMP-based and gesture-based interfaces on tablets, we contribute a comparative evaluation across devices to contrast against the traditional workplace and to include the different interaction modalities supported by each device. Chapter 4 discusses more related work on post-WIMP interactions with a focus on tablets in Section 4.2.2.

Meanwhile, large devices, such as tabletops and wall displays, provide a large surface to display larger amounts of data but require people to look at the visualizations from distorted angles and to walk around or look from a distance to have an overview. Langner et al. (2019) designed an interaction vocabulary for interacting with multiple coordinated views at different distances. While the authors focused on touch interaction on the wall or through a smartphone, we investigated interaction distances using different interaction modalities. Researchers have also proposed to include additional mobile devices for interacting with a large vertical display from a distance (e.g., Horak et al., 2018) but that is out of the scope of this dissertation. We focus on interactions with a single display, shared with others if necessary. Section 5.3.1 discusses previous work on large vertical displays further.

2.2.2 Interaction modalities

One of the main characteristics that differentiate computing devices is the ways of interaction they support. While the mouse is the primary input device for desktop computers, touch is the most common interaction modality on interactive surfaces. Each interaction modality has advantages and disadvantages. A common proposal to address the limitations is to combine multiple modalities in one system so that they can complement each other, and to increase the degrees of freedom. For instance, the pen can facilitate better spatial accuracy than touch, and thus, is offered as a second modality on tablets. While both usually work independently on commercial solutions, researchers have proposed to combine them, e.g., using a thumb press to indicate how pen actions should be interpreted (Pfeuffer et al., 2017). Another reason to support multiple interaction modalities is the potential distance of the person from the display. Touch and pen interactions require to be close to the device, while speech and mid-air gestures can be performed from a distance. Regarding interaction with data visualizations, Saktheeswaran et al. (2020) found that people preferred multimodal interaction combining touch and speech over unimodal interaction when comparing the two options with unit visualizations. However, their study considered only one visualization technique and a series of tasks, and the result may not apply in other scenarios. According to Oviatt (1999), even if people have the possibility of interacting multimodally, that does not mean that they will take it. We explore these interaction choices in Chapter 5. Our results rather suggest that people prefer unimodal interaction for low-level tasks, but also show that there are patterns regarding which modalities can be combined, and in what order, for specific exploratory tasks. We discuss further work on interaction modalities in Sections 4.2.2 and 5.3.1.

Multimodal visualization systems are still rare, and thus, their research requires custom prototypes, such as Orko (Srinivasan and Stasko, 2018) and InChorus (Srinivasan et al., 2020). The two interactive systems we developed help extend that system design space. Furthermore, the research on multimodal interaction with data visualizations has so far focused on scenarios of single users. In the next section, I discuss interaction in the context of collaborative work.

2.3 Collaborative visualization

When two or more people come together to work and interact with data, new factors come into play for the design of interactive systems. Johansen (1988) proposed to classify collaborative work across two dimensions: time and space. His time-space matrix is the most common classification of groupware in CSCW (Neumayr et al., 2018) and distinguishes between synchronous (same time) and asynchronous (different time) collaboration, as well as co-located (same place) and remote work (different place). This dissertation focuses on synchronous and co-located scenarios. However, we discuss the potential use of speech interaction in other scenarios on one of the related workshop papers (Molina León et al., 2023).

The main research field that focuses on collaborative work is CSCW, intending to understand the connection between groups and technology (Grudin, 1994). However, collaborative work can also involve data visualizations, and accordingly, Cook and Thomas (2005) considered the design of collaborative visualization tools a grand research challenge. Isenberg et al. (2011a) define *collaborative visualization* as:

> The shared use of computer-supported, (interactive,) visual representations of data by more than one person with the common goal of contribution to joint information processing activities.

When thinking of interaction design in collaborative systems, we should think not only of the interaction with the system but also of the interaction among people (Lee et al., 2012). Thus, collaboration style, awareness, territoriality, and record-keeping become important factors to consider (Mahyar et al., 2012; Scott et al., 2004). In Chapter 6, we focus on collaborative styles and the interplay between them and interaction modalities, as well as awareness. Although relevant, territoriality and record-keeping are not within the scope of this dissertation. Tang et al. (2006) proposed to classify collaboration styles into six coupling styles, that Isenberg et al. (2010) extended to eight codes for taking multiple representations of the same data source into account: active discussion (DISC), view engaged (VE), sharing of the same view (SV), sharing of the same information with different data views (SIDV), same specific problem (SSP), same general problem (SGP), different problems (DP), and disengaged (D). We applied the later set of codes (excluding SIDV because there were no duplicated views) to classify the collaboration between our participants in Chapter 6. Moreover, we focus on the process of collaborative sensemaking. Qu and Hansen (2008) define collaborative sensemaking as

> a process where a group of people (...) seeks or creates a shared representation collaboratively to accomplish a shared task.

When thinking of collaborative interactive systems, people may have the option to either interact individually or collaboratively. There is a body of research on designing visualization systems supporting either interaction type (Chegini et al., 2017; Langner et al., 2019; Liu et al., 2017) and we build upon their work to examine user preferences in Chapter 5. We found that participants preferred personal interactions to explore data in a large vertical display, and identified relevant interaction patterns to understand when they decide to interact collaboratively. We discuss more related work on collaborative data exploration in Section 5.3.2 and on collaborative visual analytics in Section 6.2.2.

Regarding interaction modalities, Tse et al. (2008) investigated multimodality in collaborative work on tabletops and found that having more than one modality may influence how people collaborate. Inspired by their work, we examine this relationship on wall displays for the first time in Chapter 6. Badam et al. (2016) proposed to combines proxemics and mid-air gestures to control interactive lenses on large displays. They found that while each modality may fit better specific tasks (e.g., proxemics for navigation), a combination of the two modalities enabled participants to benefit from the advantages of each and to switch seamlessly between implicit and explicit interaction. While previous HCI work has proposed systems that incorporate multimodal interaction in collaborative activities (Cohen et al., 1997; MacEachren et al., 2005), it is still unknown how using more than one interaction modality may influence team work and the interaction among team members.

Jakobsen and Hornbæk (2014) investigated collaborative work involving only touch interaction and found that participants shifted easily between loosely coupled and closely coupled work. Based on their findings, we seek to investigate whether the same holds for collaborative work involving speech interaction. As speech commands require switching attention towards the screen, we investigate whether that disrupts the workflow. We found no evidence of speech commands being an obstacle to collaborative work, and participants used speech interaction in the same proportion across loose and close collaboration (see details in Section 6.5).

In the four following chapters, dedicated sections present the related work relevant to the specific research contributions of each chapter.

3

Co-creating Visualizations: A First Evaluation with Social Science Researchers

This chapter is based on the first full paper of the dissertation, published at EuroVis 2020 (Molina León and Breiter, 2020). The published paper and the chapter are identical. The information system designed in the study was the first prototype of the Global Welfare State Information System (WeSIS) of the Collaborative Research Center "Global Dynamics of Social Policy" (CRC 1342, 2023).

3.1 Introduction

Visualization designers have multiple approaches on how to organize the design process of custom visualizations. In recent years, there has been a rise in human-centered design approaches under the umbrella of user-centered design (Lloyd and Dykes, 2011; Mckenna et al., 2015) to elicit user requirements, accompannied by an increasing interest in creative methods to discover visualization design opportunities (Goodwin et al., 2013; Kerzner et al., 2019). Participatory methodologies have been mentioned (Henry and Fekete, 2006a; Kahng et al., 2018; Landstorfer et al., 2014a) but are still rare, and therefore, understudied in visualization design research. *Co-creation* is a design methodology that proposes to design not only *for* the future users, but also *with* them (Sanders, 2008). It is based on the principles of mutual learning, empowerment, openness and diversity, involvement and ownership, transparency, and effectiveness (Bossen et al., 2016; Jarke et al., 2019). The future users are valued as co-creators by the design experts, and the aim is to increase satisfaction and long-term adoption by including the validation of the people affected early on.

We present a visualization case study of using co-creation as a design methodology to create an information system *with* and *for* social science researchers. Nowadays, there is an increasing availability of large datasets for social science research, thanks to the open government initiative and the rise of social media. Therefore, social scientists are integrating programming into their workflow to analyze their data. However, the technical skills of the community are not growing as fast as the amount of data available. Consequently, they are increasingly collaborating with computer scientists to develop tools that support their data exploration and analysis. In our case, we collaborated with a group of social science researchers to design a visual information system that supports them in their research on social policy data. We are all part of an interdisciplinary project, in which political scientists, sociologists, geographers, economists, and computer scientists collaborate to describe and explain social policy phenomena.

We considered co-creation a suitable approach for this scenario because experts of diverse disciplines needed to develop a shared understanding of their goals across domains, and co-creation aims to support mutual learning towards developing a product that fulfills the tasks. We hypothesized that the co-creation principles (mutual learning, empowerment, openness and diversity, involvement and ownership, transparency, and effectiveness) would be achieved by applying the co-creation methodology to the design of the system. Achieving the principles meant that the co-creators would learn from each other, feel empowered, consider the process open and diverse, feel involved and in ownership of the designed system, feel that the process was transparent and that it led to an effective design. Although participatory approaches have an inbuilt evaluation embedded (Bossen et al., 2016), we used specific methods such as surveys and interviews to assess how the co-creators perceived

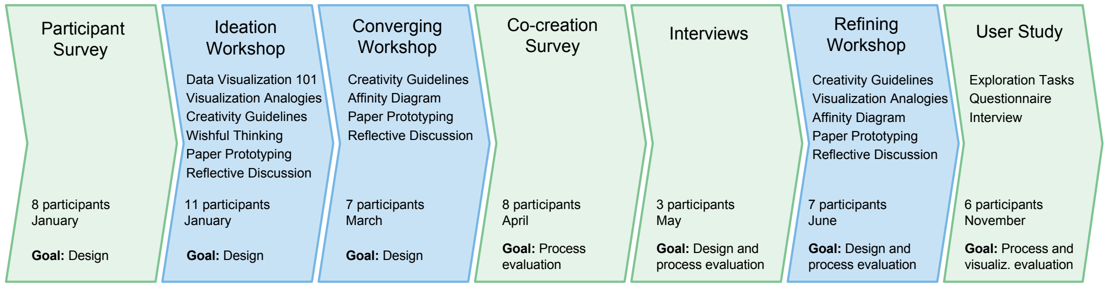


Figure 3.1: Overview of the co-creation process. The workshops are in blue and include the methods used. Although each step has specific goals, the evaluation of the design is also inbuilt throughout the whole process (Bossen et al., 2016). In total, 14 social science researchers participated.

the process according to the principles.

Regarding the domain goals of the experts, our aim was to analyze their research tasks to identify visualization opportunities. Through the co-creation process, it became clear that the social scientists want to explore the data to discover temporal and geographical patterns which, in combination with their domain expertise, lead them to generate research hypotheses that can be later tested with the data. We elicited the requirements to design such visualizations, and developed multiple prototypes of the visual system in an iterative process. The resulting system includes more than 400 indicators about social policies (e.g. "Government expenditure on health") applied worldwide over the last 140 years. It allows the social science researchers to explore their data through topic pages (e.g. "Health and long-term care") and country profiles, as well as by combining and comparing multiple indicators. We released a first version of the system within our project and continue working on new features based on the researchers' feedback. They share their data across the project through the system and plan to cite it in their publications.

In this paper, we focus on describing the co-creation process we have conducted, and a first evaluation of the design methodology. We conducted three co-creation workshops, two surveys, a round of interviews, and a user study with 14 social science researchers. An overview of the process is shown in Figure 3.1. After the second workshop, we conducted a formative evaluation of the process with an evaluation survey, interviews, a group discussion, and the documented reflections of the facilitators. After the last workshop, we conducted a summative evaluation by means of a user study, in which the researchers interacted with the system and were interviewed to learn more about their experience, not only with the visualizations, but also with the whole co-creation process. Through co-creation, participants felt listened to, and felt empowered by learning about data visualization. However, in the formative evaluation, they were not confident about the effectiveness of the methodology to develop visualizations that support their research. They preferred to receive a finished product within a shorter time frame, instead of committing to a long-term research process they may not benefit from right away. After testing the first prototype of the system in the user study, researchers were more positive about the usefulness of the visualizations for their research. Overall, they most valued working together with their peers in the workshops, as well as creating a system tailored to their research collaboratively. To our knowledge, this is the first evaluation of co-creation as a visualization design methodology. We share our insights to guide visualization researchers and practitioners who are considering using co-creation as a methodology for visualization design.

3.2 Related Work

The origin of co-creation goes back to the participatory design (PD) movement that emerged in Scandinavia in the 1970s. This movement started with a political agenda that claimed that everyone affected by a decision should have the opportunity to influence it (Lee et al., 2018; Schuler and Namioka, 1993). In the case of information technology design, the motivation was the transformation of the workplace due to the introduction of computers, and the aim was to ensure that people who use this technology play a critical role in the design (Simonsen and Robertson, 2012). We apply this participatory mindset to our scenario to design data visualizations that facilitate the work of social science researchers.

According to Sanders and Stappers, co-creation is any act of collective creativity, and co-design is collective creativity applied throughout a design process. However, these terms are often used interchangeably (Sanders and Stappers, 2008). Sanders (2008) argues that design researchers with a 'participatory mindset' design *with* the people because they value the future users of the tool as *co-creators*. The opposing 'expert mindset' refers to designing only *for* the people. In the private sector, research has shown that co-creation has a positive effect on the customer's satisfaction and loyalty (Grissemann and Stokburger-Sauer, 2012). In the public sector, co-creation is considered a cornerstone for social innovation and one of its main goals is citizen involvement (Jarke et al., 2019; Voorberg et al., 2015). Dörk and

Monteyne (2011) define the involvement of citizens in urban planning and design as urban co-creation. Furthermore, co-creation has been used to work with specific audiences, such as blind people (Ugulino and Fuks, 2015) and elderly individuals (Bull et al., 2017). In the field of learning analytics, Dollinger and Lodge (2018) claim that co-creation has the potential to increase flexibility as well as the chances of long-term adoption. Drawing on these insights, we chose to co-create with our domain experts to increase the chances of success and adoption of the system, given that we work in an academic context and have the possibility to work closely with them.

In the field of visualization research, Landstorfer et al. (2014a) co-created a visualization with network security engineers to support the inspection of large log files. Based on their experience with those domain experts, they proposed several principles for co-creation such as defining and refining requirements with the stakeholders face-to-face, using real data from the stakeholders as early as possible, brainstorming and sketching with the users, and rapid prototyping. In a geo-visualization case study, Lloyd and Dykes (2011) concluded that using real data is key to understanding the needs and design possibilities, and that paper and virtual prototyping enable successful communication. Paper prototyping is particularly good at eliciting suggestions for novel visualizations. We followed these principles and insights in our design study. Furthermore, Kerzner et al. (2019) recently proposed a framework for creative visualization-opportunities workshops with multiple guidelines and methods. We used and adapted their methods of Creativity Guidelines, Wishful Thinking, Visualization Analogies, and Reflective Discussion in our co-creation workshops.

Regarding the effects of co-creation and similar participatory methods, Frishberg (2011) and Kahng et al. (2018) reported that facilitating participatory design sessions helped them identify important needs in their case studies. According to Steen et al. (2011), further research is necessary to evaluate the co-design method according to its intended benefits and to assess its costs and risks, for instance, comparing the organization costs with the outcome of the project. Mitchell et al. Mitchell et al. (2016) investigated the impact of co-design when generating sustainable travel solutions and found that co-design promotes idea generation and a holistic view of the problem. In contrast, we investigate the co-creation with social science researchers, to report its benefits and limitations for visualization case studies.

3.3 METHODOLOGY

This work presents a qualitative case study based on a year-long collaboration of the authors with four research groups of social science researchers working together in a collaborative research center. Fourteen researchers decided to take part in the co-creation process to design a visual information system led by the authors. As co-creation focuses on the use of workshops for collaborative design, we conducted three co-creation workshops combined with surveys, interviews and a user study. Collaborative design means that the domain experts would take part not only to help elicit the requirements but also to propose their design ideas. Based on successful examples of related literature (Goodwin et al., 2013; Kerzner et al., 2019), we applied multiple workshop methods, such as Wishful Thinking and Paper Prototyping, described in Section 3.4. Furthermore, we observed the participants and documented each activity. Before and after each workshop, the facilitators filled out a reflection form to document their plans, impressions and experience. They wrote down the planned activities for each workshop, then what happened during the workshop, and afterward, what they learned from it. In particular, the form served to compare the workshop goals with the outcome and to reflect on the events to plan the next steps.

With the co-creation principles in mind, we wanted to assess whether applying co-creation as a design method would lead participants to feel empowered, involved, and to learn from each other, among others. To evaluate the use of co-creation, we conducted a formative and a summative evaluation. The formative evaluation included a survey after the second workshop, a round of interviews, and a reflective group discussion in the third workshop. Although our sample was small, the survey results gave us an overview of the participants' experience through the co-creation process. Then we conducted semi-structured interviews to better understand the workflow of our co-creation partners and their experience through the process. With the written consent of the participants, we recorded and transcribed the interviews. We used the co-creation principles to develop a coding scheme for analyzing the interviews, that we later adjusted during the analysis. The analysis helped us to assess whether the principles were being achieved and to plan the next workshop accordingly. In the last workshop, we organized a reflective discussion that we also transcribed and analyzed.

Once the prototype of the visual information system was ready, we conducted a summative evaluation through a user study. The study consisted of a series of tasks, a second survey and a final interview. Interviews were both about the designed system and about the co-creation process. We recorded, transcribed and qualitatively analyzed them, producing a coding scheme that led us to identify the core topics. Based on the documentation of the workshops, the observations, and the other methods just mentioned, we then proceeded to critically reflect on what we learned from this case study. In Section 3.4, we present each step of the process together with the evaluation results, and the critical reflections on our experience. In Section 3.5, we present the lessons learned.

3.4 Our Co-creation Process

In the context of a large interdisciplinary project, we worked together with 14 social scientists to analyze how technology can support their work, and to design a visual information system accordingly. These researchers are investigating the global evolution of social policies over the last century, and work in different research groups according to the topics they specialize in: social security, labor, economic relations, health care, education, and family policy. During the preliminary meetings with each group, visualization came up as one of the main topics they were interested in. They want to visualize their time series and network data in order to explore it with the goal of finding patterns that lead them to formulate their research hypotheses, or help them to work on these hypotheses. We started by studying the data portals of international organizations the researchers were familiar with, such as the World Bank Group, 2022 and the Organisation for Economic Co-operation and Development, 2021. However, we also had to consider that the main goal of the social scientists was to focus their research on new data that they planned to collect, and our collaboration was going to happen parallel to the data collection process.

Based on this information, we planned a co-creation process to explore the possibilities and opportunities for data visualization to support their work. Throughout the process, we iteratively co-designed and developed visualization prototypes, following the double diamond design process model (Design Council, 2015). This model first encourages divergent thinking to define the problem, and then convergent thinking to find the best solution. We organized a series of workshops that combined creativity techniques from related literature (Goodwin et al., 2013; Kerzner et al., 2019) with a focus on paper prototyping and rapid software prototyping, to elicit novel ideas and easily adjust the design, as suggested by previous case studies (Lloyd and Dykes, 2011; Sedlmair et al., 2012). Designing collaboratively in the workshops aimed to ensure the deep involvement of the co-creation partners, mutual understanding, and a continuous 'built-in' validation of the visualizations (Landstorfer et al., 2014a). In each iteration, we started designing together on paper – either proposing new ideas or annotating printed prototypes – and then produced or refined corresponding software prototypes. The core design sessions took place in the workshops, while the software implementation was done by the visualization researchers after each workshop. Our method choices were based on design processes for creative visualizations (Kerzner et al., 2019), and previous co-creation projects (Jarke et al., 2019; Lee et al., 2018). In each workshop, there was a main facilitator leading the methods, a second facilitator taking care of the organizational aspects, and a documenter. Each workshop took half a day and was documented with a co-creation reflection form (Jarke et al., 2019). In the form, we first wrote down our intended goals and planned activities for the workshop, then what actually happened in the workshop, and afterward, we reflected on whether we reached our goals and what we learned.

To help participants learn from each other and to counter the influence of power hierarchies, we asked participants to work with people from different research groups whenever possible. Besides the workshops, we conducted surveys and interviews to get input not only from the group activities, but also at the individual level. Overall, we conducted two surveys, three workshops, a round of interviews, and a user study, complemented by the preliminary meetings and informal discussions. The process is shown in Figure 3.1. Each step is explained in the following sub-sections. The whole process took over a year and had 14 participants: two full professors, five postdocs, six doctoral students, and one master's student.

We evaluated the co-creation process through the co-creation survey, the interviews, and the discussion in the refining workshop, as well as with our continuous reflection through the co-creation reflection form. Since we asked participants to write down their answers in color-coded cards in the different workshop activities, we were able to then read and analyze them, to reflect on what happened. Based on related literature on the evaluation of participatory design and co-creation (Bossen et al., 2016; Jarke et al., 2019), we evaluate the co-creation process according to the following criteria: (1) mutual learning, (2) empowerment, (3) openness and diversity, (4) involvement and ownership, (5) transparency, and (6) effectiveness. We apply these principles to our case, not only based on related literature, but also because a successful interdisciplinary collaboration includes developing a shared understanding. Effectiveness is evaluated through the qualitative feedback of the participants. We present our evaluation findings in more details in Sections 3.4.4, 3.4.6 and 3.4.7.

3.4.1 Participant Survey

First, we conducted a survey to learn more about the stakeholders and their previous experience with data visualization. We sent it to the researchers who expressed interest in participating in the workshops, and used the results to prepare the content of the workshop. We asked nine questions about the researchers' background, the data they work with, their use of visualizations at work (if any), and the reasons for using them. The survey is included in the supplementary material of this paper.

Eight researchers took part in the survey: four political scientists, two geographers, one sociologist and one economist. All of them work with tabular data — six of them specifically with time series. Four work additionally with networks. When asked about how often they use visualizations at work, four of them replied 'usually', and one 'always'. Regarding their motivation, six have used visualizations for presentations with their colleagues, while five have used them for data exploration. Based on these results, we planned the first workshop focusing on the use of networks and tabular data for the tasks of presentation and data exploration.

3.4.2 Ideation Workshop

Keeping in mind the goal of developing a collective shared understanding across disciplines, we started the first workshop with an introduction called Data Visualization 101. We presented Munzner's analysis framework (Munzner, 2014) to explain how we can systematically explore the visualization design space, specifically the data, the tasks, and the visualization and interaction techniques. We then showed diverse examples related to their data and their tasks in Visualization Analogies, to inspire and to give an overview of the possibilities. After the visualization-focused content, we used Wishful Thinking (Goodwin et al., 2013) to learn about the social science perspective. The participants expressed their research aspirations by answering three questions: (1) "What would you like to *know*?", (2) "What would you

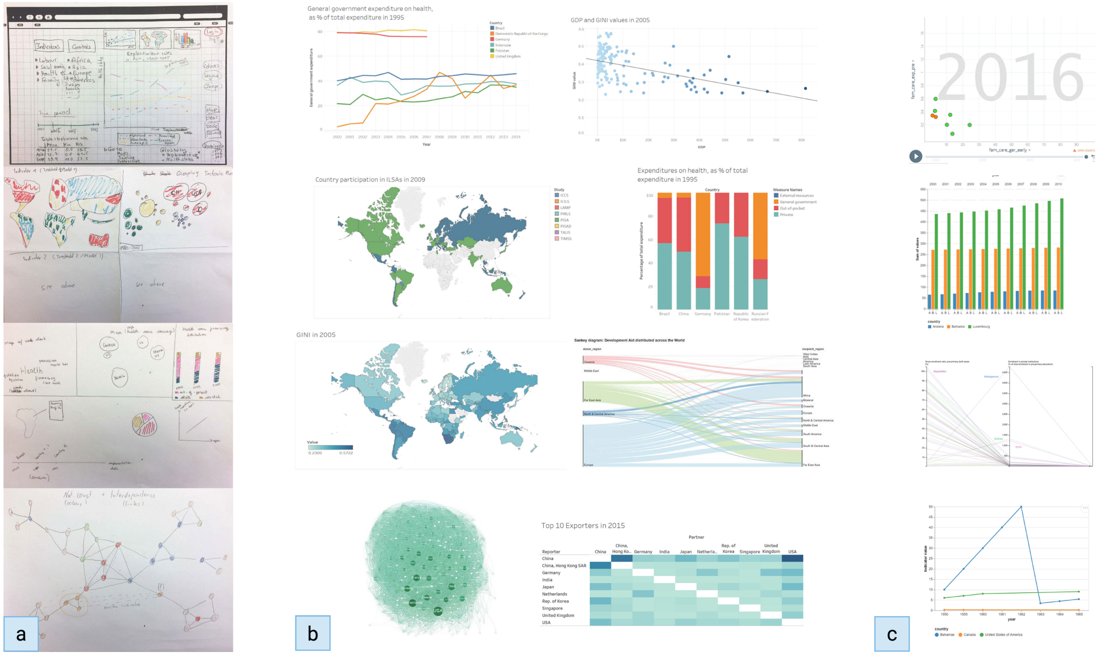


Figure 3.2: (a) A selection of the paper prototypes from the ideation workshop. (b) Software prototypes from first iteration. (c) Second iteration. We used Tableau, Vega-lite, D3.js, and other tools to translate the core ideas of the paper prototypes into interactive visualizations, complemented with other relevant techniques. For example, we translated the drawn line charts and scatterplots first into interactive equivalents (on the top), and then created an animation of the scatterplot according to the participants' feedback.

like to *be able to do*?", and (3) "What would you like to *see*?". Answering these questions individually and discussing the answers with the group helped participants to make their goals and expectations more concrete. In previous meetings, we had already received initial visualization suggestions from the researchers. However, as Sedlmair et al. (2012) suggest, a common pitfall in design studies is that domain experts think about the solution before thinking about the problem. This is why we asked our co-creation partners to first focus on their wishes.

Furthermore, we did early paper prototyping following the recommendation of Koh et al. (2011), allowing the experts to sketch their design ideas early on. The goal was to start collecting visualization ideas that would correspond to their wishes. The main workshop facilitator took the position of an active collaborator, as described in the co-creation framework of Lee et al. (2018). This means that the facilitator actively participated in the idea generation and decision making. Each participant drew at least one prototype. A selection of the prototypes is shown in Figure 3.2(a). The most common visualization techniques used were maps and networks, and five participants drew and described multiple coordinated views. Each participant explained orally what they would expect to happen if interaction was included, and the most common interaction technique was the selection of a particular indicator or time frame. Two participants wished to save their queries. After the workshop, we applied *parallel prototyping* (Dow et al., 2011) to design multiple low-fi prototypes covering the design space as broadly as possible. We analyzed and clustered the paper prototypes according to the common topics and ideas. For each one, we documented the type of data, the tasks, and the techniques considered. Then we used this information to create software prototypes to present in the next workshop (see Section 3.4.3). In the reflective discussion, participants expressed that the most valuable activity was Data Visualization 101.

Reflections

Starting with Data Visualization 101 encouraged some participants to focus on visualization solutions too soon. We chose to give the introduction first to provide an overview of the topic before brainstorming, and to lead by example on exchanging domain expertise. Although participants were very positive about learning the basics of visualization, Wishful Thinking answers were focused on designing visualizations. Doing Wishful Thinking before would have allowed the participants to first focus on the task. Participants seemed to have a hard time developing their own ideas from scratch. They felt more comfortable discussing existing examples, and their ideas were often a reflection of the visualizations they have interacted with on the websites of the World Bank Group, 2022 and other international organizations. Furthermore, they focused on describing scenarios to explore their data. This indicated that their main interest was data exploration.

3.4.3 Converging Workshop

In our co-creation process, we aimed at following the double diamond design process model Design Council, 2015. In this model, co-creation starts with divergent thinking, i.e. imagining possibilities, and then proceeds to convergent thinking, i.e. agreeing on alternatives. Since the ideation workshop focused on divergent thinking by enabling participants to discuss initial ideas and to explore visualization possibilities, the converging workshop aimed to define the problem more specifically by reaching agreement. Accordingly, we had a group discussion to agree on the tasks that everyone needed to tackle, and this led to creating an Affinity Diagram about the researchers' workflow. This activity resulted in agreeing on four steps that describe the researchers' workflow: (1) discover time and geographical patterns in the indicators, (2) generate hypotheses, (3) test the hypotheses through statistical data analyses, and (4) explain the patterns. Discussing the workflow with the whole group led to developing a shared understanding of the common problems.

When we invited the participants to this workshop, we asked them to send us examples of their data. In the workshop, we presented 10 visualization prototypes that used their data, based on the ideas we had discussed in the ideation workshop. Some of them are presented in Figure 3.2(b). We converted the paper-based ideas into software prototypes and complemented them with more visualization techniques according to their data and tasks, following Munzner's framework Munzner, 2014. For instance, we translated the paper-based node-link diagrams of country relationships into a network prototype and then included a Sankey diagram because the researchers sent us data about development aid and were interested in the distribution of aid across world regions. The prototypes included standard interaction techniques such as selection and showing tooltips on hover to help users explore the data. We asked the researchers to analyze the prototypes in small groups, and then to share their thoughts for the next iteration. Our goal with rapid prototyping was to quickly get feedback from our co-creators, understand them better, and refine the prototypes, as well as the requirements accordingly Fuks et al., 2012. Participants were finally able to see how their ideas can be combined in a concrete product. In the reflective discussion, one participant was positively surprised that the researchers had more in common than expected, and two mentioned that they appreciate that co-creation is a continuous process where they get a chance to provide feedback quickly.

Reflections

The steps of the research workflow that we agreed on almost perfectly match the description of the *discover* task by Munzner Munzner, 2014. Experts want to use visualizations to discover data patterns that help them to generate hypotheses and later test them. Therefore, the definition of the workflow led us to successfully abstract their main goal. Preliminary requirements originated from the workflow, such as comparing indicators, combining time and space attributes of related indicators, and identifying similarities and differences across them. Showing initial visualizations with data provided by the researchers made a big difference for the participants. This made the progress of our collaboration concrete and visible.

3.4.4 Co-creation Survey

To perform a formative evaluation of the co-creation process, we conducted a survey with the social scientists after the converging workshop. The questionnaire had 16 questions: eight questions with statements to rate from 1 (strongly disagree) to 5 (strongly agree), two open questions about the workshop content, two open questions about their personal involvement, two questions about participation and two about demographics. The questions were meant to assess whether the co-creation principles had been followed. For instance, we asked what did they learn from participating in the process to find out if mutual learning took place. The complete questionnaire is included as supplementary material of this paper.

Eight participants took part in the survey. An overview of the answers concerning participant agreement with the eight statements is shown in Figure 3.3. Although the sample is small, the results yield some meaningful insights. The three researchers who participated in both workshops had a higher level of agreement with every statement than the other participants. All participants agreed with the statement "I feel listened to". On the other hand, participants were least convinced about having an equal chance to be part of the decision making, and about the co-creation leading to visualizations that assist their research. Three participants answered that creating the affinity diagram about their workflow was the most useful activity, while paper prototyping was most useful for two. The preferred workshop activities were paper prototyping and the group discussion that led to create an affinity diagram of the researchers' workflow. Others pointed out that the co-creation process has helped them to better understand what their colleagues are doing. When asked how to improve the workshops, there was no consensus but multiple suggestions were given, such as teaching how to create visualizations with the tools they are most familiar with, and spending more time on discussing existing tools and how to improve them.

Figure 3.3: Survey answers for the eight statements about the co-creation process up until the converging workshop.

3.4.5 Interviews

We conducted and recorded semi-structured interviews to better understand the researchers' workflow and their thoughts on co-creation. To this end, we organized a series of open questions in two parts. The first one concerned the researchers' role in the project, the research tasks, and the artifacts used by the researcher within those tasks. It also included discussing the workflow from the converging workshop. The second part was about the motivation to participate in the process, as well as the expectations about the co-creation process and about the visualizations, at the beginning of our collaboration.

We conducted a pilot interview, followed by three interviews with researchers from different research groups: one doctoral student, one postdoc, and one professor. Later we transcribed and analyzed the interviews. The first part of the interview reflected that the current task of the researchers is data collection. They pointed out that the research workflow of the converging workshop had not fully started yet, because the data collection takes most of their time. The artifacts they presented were data sources and programming scripts to collect data. Two researchers pointed out that they will never collect all the data they wish to have, due to a lack of data from specific world regions and time frames.

The second part of the interview was about evaluating the process so far. We asked participants a series of open questions about their motivation to participate, their expectations of co-creation and of the visualizations before the process started, as well as their opinion about the results so far. Their main motivation to participate and initial wish for the co-creation process was to learn how to create visualizations on their own. They were satisfied with the introduction in the first workshop, but they would like to know more about tools they can use to create visualizations with. According to two participants, data visualization is a major opportunity for better communicating social science research, especially to the general public:

"*If we could get better visualizations, our work would become more useful. That I'm sure of*" — P3

"*Visualization is the stepchild of social sciences... there's a lot to be done. Therefore, this project is a great opportunity to make steps further*" — P10

However, P3 pointed out that social science researchers may be interested in complex visualizations, while the general public may be not. On the one hand, P3 and P10 often reflected on what would be best to show to the general public, although the system is meant to be used by social scientists. On the other hand, P7 defined the usefulness of the visualizations according to whether they are ready in time for the next publication. It was not clear when the visualizations will be delivered to them.

Regarding the start of the design process, P10 would have preferred starting with existing visualizations from the domain instead of brainstorming from scratch. He found it hard to develop new ideas without having any visualization expertise. Regarding the progress on designing the information system, the participants had the impression that no decisions were made yet because the visualization prototypes shown individually in the second workshop were not yet embedded in the system.

3.4.6 Refining Workshop

The main goal was to refine the user requirements based on the researchers' workflow, as well as to analyze the prototypes according to the requirements. We went through the four steps of the research workflow, and discussed the most important features that the system should have to support their workflow. First, participants

Figure 3.4: Survey answers for statements about the visualizations presented in the user study and about the co-creation process. The color green represents questions that were also asked in the previous survey (see Figure 3.3) and purple represents new questions.

split into small groups to discuss, starting with the preliminary requirements based on previous workshops and interviews. We then refined the requirements with the whole group, and prioritized the most important ones. The following requirements were raised, in order of relevance:

1. Enable to interactively change the threshold values of analysis variables and visualize the change.
2. Show an overview of the missing data.
3. Provide tools to divide continuous variables into categories.
4. Select and compare indicators over time.
5. Combine and order time events from different indicators.

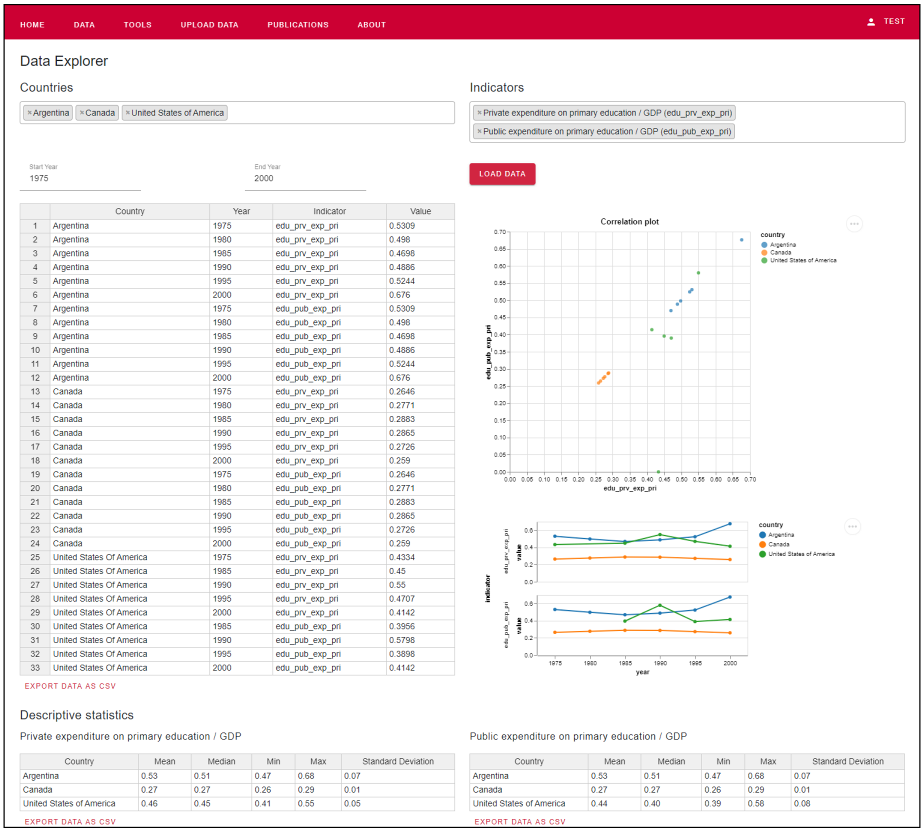


Figure 3.5: This is a screenshot of the Data Explorer, one of the system features evaluated in the user study. We co-designed this page where users can combine data of multiple countries and indicators, including descriptive statistics, to compare different indicators over time.

Afterwards, we presented the software prototypes of the second development iteration. They combined improved versions of the prototypes of the converging workshop with new ideas and previous prototypes that participants were satisfied with. Some of them are presented in Figure 3.2(c). We handed out printed versions and asked researchers to analyze them according to the requirements. This activity resulted in a selection and combination of five prototypes that researchers considered as having the most potential to fulfill the requirements.

Reflective Discussion

We concluded the workshop with a reflective group discussion where we presented the most relevant results of the co-creation survey, to collectively assess the process according to the co-creation principles. We asked participants to reflect on the reasons behind the survey results. Here we describe the discussion based on our posterior analysis of the transcription. The effectiveness for each individual and the commitment to the process were the main topics of discussion. Effectiveness was divided into two categories: the effectiveness of the final visualizations, and the effectiveness of the process for the everyday tasks of each researcher. While the final visualizations may be useful in the future, participating in the process requires a high level of commitment and the short-term impact is unclear, especially for doctoral students eager to advance quickly in their research. Therefore, the ability to create visualizations on their own is more valuable than relying on the success of a long-term process. One participant even suggested that this mindset may be a problem among researchers.

"*...our projects are so specific that we will each probably have some visualization that we actually do that we don't expect [the system] to provide for us because no one else will need it, so its not a public good*" — P3

"*Most researchers I know don't like to rely on others' decisions... it's always hard to commit to such a broad process... it's not a criticism [of the process], it's more of a self-criticism*" — P14

Understanding the computer scientists better and learning about the needs of other social scientists were the main advantages of the process that participants agreed on. Mutual learning helped to better understand how the system could be useful for the researchers. Overall, the participants who had participated in every workshop were much more positive than those who did not. This may indicate a relationship between the degree of participation and their satisfaction with the process.

Reflections

The refined requirements we agreed on reflect not only data exploration, but also analysis. This may be a result of the researchers feeling that exploration has been already tackled, and then moving on to the next steps in their research workflow. Three of the six workshop participants had not participated in the previous workshop. For them, the goal of our process seemed the least clear because they had not experienced the evolution of the ideas. They led the group discussion that focused rather on the overall commitment to the process and its connection to their everyday research tasks.

3.4.7 User study

After working on the next iteration of visualization prototypes and integrating them in the system, we invited the researchers to try them out individually. The resulting web system allowed users to upload social policy indicators and to explore them. It included a descriptive page for each indicator with visualizations such as stacked bar charts and choropleth maps dynamically generated according to the data type. Furthermore, it had a world overview page with an interactive map, country profiles with a summary of the main indicators, and a Data Explorer page that allowed to combine multiple indicators to look for correlations and other patterns. A screenshot of the Data Explorer is presented in Figure 3.5. In the study, the researchers performed three tasks, filled out a questionnaire and were interviewed. This was the first time they saw the visualizations integrated into the information system. This seemed to make a big difference for them, in comparison to seeing them separately before. The questionnaire aimed to learn more about their views on the co-creation process after seeing a further advancement and having the opportunity to interact with the system. In the semi-structured interview, we looked back at the whole process and reflected on how it led to the current result. The details of the user study are included in the supplementary material.

Six participants took part in the user study. Three of them (P1, P2, P3) had attended each of the three workshops, while two participants went once and another one twice. First, we asked participants to perform four exploratory tasks that led them to navigate the main features of the system: (1) Missing data overview, (2) Browsing by topic, (3) Browsing by country, and (4) Data Explorer. In each task, the participants had to reply questions such as "How did the enrollment rate in primary education of the United States change over time since 1880?". The questions were meant to guide them through the system and to interact with the visualizations. We observed them, asked them to think aloud and allowed them to ask for help if necessary. All participants responded the task questions correctly. They were happily surprised because the visualizations were automatically generated based on the indicators, the countries and the time frame they chose. Although a data table was always shown next to the visualizations (see Figure 3.5), they often looked only at the visualization and used the tooltip to check the data values. They particularly liked the possibility of combining any group of indicators — which is not possible on most websites they work with, and the corresponding scatterplots generated. While participants considered the time-based line charts useful to describe the data, they saw the scatterplot as an analysis tool because of regression analyses they often use looking for correlations.

After completing the exploratory tasks, we gathered feedback from the researchers with a questionnaire and an interview. The questionnaire was about the visualizations and the co-creation process (see supplementary material). An overview of the answers is shown in Figure 3.4. According to the results, everyone found the visualizations useful and easy to use. Participants appreciated having the option to download the visualizations, but would have liked to be able to change the labels and colors. In certain features such as the country profiles, participants wished to personalize the content. In contrast to the previous survey (see Figure 3.3), participants considered the workshops more useful to them after interacting with the prototype, and they agreed more on the co-creation of visualizations being helpful for their research. They found the workshops useful to design the visualizations, and they would be open to participating in another co-creation project.

The final interview focused on qualitative feedback about the co-creation process. We discussed the co-creation principles of involvement and ownership, mutual learning, effectiveness, and openness and diversity. Three participants mentioned that co-creation is a new experience for them. Over time, they realized that their role was not just being a consumer asking for a product, but rather being a participant in the process of building the product. P1 noted that this experience contrasted with everyday shopping, where he is used to rather checking the existing options and just choosing one of them. When asked about their influence in the outcome of the project, only P5 could see her personal influence in a particular feature of the system. Most participants had a hard time seeing their personal influence. They rather saw the influence of the group and the result of their collective effort. The system actually represented the requirements and features they agreed on.

What participants liked most of the co-creation process was the interaction with the other researchers. Learning about their data, as well as recognizing the similarities and the differences across their tasks. The workshops were the events in which they most learned about their colleagues' work. The second most positive aspect was to create a tool tailored to their research. According to P2, working in groups made their ideas better than the sum of all individual ideas. He explained this with

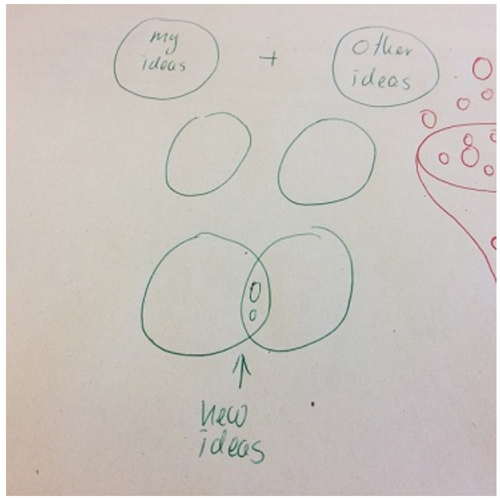


Figure 3.6: Drawing of P2 used to explain how the group work led to new ideas, better than the sum of everyone's ideas.

the drawing shown in Figure 3.6. When asked about what they liked least, P2 and P3 found it problematic that many participants were not present in every workshop. This led to repetition, and they had the impression that those participants could not really follow the progress. P1 also pointed out that it would have been better to have more principal investigators participating. These three participants attended every workshop. In contrast, P5 and P6 mentioned that doctoral students need to focus on publishing papers and participating in the workshops felt like a task that may not necessarily help them with their publications.

Overall, participants found it challenging to start designing the system from scratch. The reasons were two-fold: their feeling of lacking enough technical expertise, and the large size of the design space. Four participants expressed that they were impressed after seeing the prototype. Everyone pointed out that having options such as combining multiple indicators, the descriptive statistics, and the multiple visualizations made this system a better choice than the websites of international organizations they are used to work with.

3.5 Lessons Learned

Here we present our lessons learned for researchers and practitioners alike, who consider using co-creation as a visualization design methodology, based on our co-creation experience. These lessons are the result of our analyses and reflections on co-creating with social science researchers as domain experts. In particular, the lessons L1, L2, and L4 are mainly based on our observations in the workshops and the reflection form. L3, L5, and L6 became clear through the interviews and the user study, and L7 was mostly discussed in the last reflective discussion and the interviews.

L1 *Incorporate an introduction to visualization after defining the problem*. In the first workshop, the introduction called Data Visualization 101 was especially welcomed by the domain experts, who expressed their interest in learning more to further develop their skills. However, this happened too soon in our case, and we had to push the discussion back to the problem space.

L2 *Help participants to feel comfortable as co-creators*. Participants often pointed out that they did not feel confident enough to design visualizations because they were not experts on the topic of data visualization. It is important to emphasize that their domain expertise is valuable and necessary for a successful design process. To help developing creative solutions, previous research on promoting creativity in visualization design offers multiple successful methods (Goodwin et al., 2013; Kerzner et al., 2019).

L3 *Balance openness and commitment*. One of the main challenges of the process was working continuously with a large and diverse group of researchers for a year. We always allowed new participants to join at later sessions, following the co-creation principles of openness and diversity. However, only three out of 14 researchers attended every activity. Before each workshop, we proposed multiple dates to the potential participants, and we limited the workshop length to three hours to maximize the chances of participation. As it was discussed in the last interviews (see Section 3.4.7), the incorporation of new participants slowed the process. Both the goals and our overall progress seemed least clear to participants who attended the least, and the most committed participants considered that this situation was the main disadvantage of the process.

L4 *Don't forget that co-creating is agreeing on the diversity of ideas.* Most visualization designers are trained to figure out individually what the users need. In a co-creation process, it is important to remember that co-creating the outcome is everyone's mission and the ideas of the domain experts should not be easily discarded. Furthermore, agreement on the system requirements may be challenging due to the diversity of data and tasks across participants, as in our case. Although the goal is usually delivering a product that helps everyone, focusing only on features that everyone agrees on may not be good enough for each individual. Our participants recognized their influence on the outcome as a group, but did not recognize it as individuals. This may indicate that we did not correctly adapt to their diversity.

L5 *Participation is not always perceived as an advantage.* We believed that applying co-creation as a design method would satisfy the domain experts because it was an invitation to be involved. Following the co-creation principles of mutual learning, involvement and ownership, we organized multiple workshop activities to help participants expressing their needs and wishes. However, participants struggled with the method because they rather expected to get multiple finished solutions to choose from. Although some researchers later appreciated the method to collaboratively design the system, others did not.

L6 *Mutual learning is not only about learning from the domain experts, but also about the experts learning from each other.* In the interviews, participants agreed on that working together with their colleagues was the main benefit of the co-creation process. The workshops helped them to not only learn from each other, but also to refine their collaboration inside and outside of the workshops. One participant affirmed to not only have learned how to better present his own research, but also to correctly define what he truly needs to reach his research goals.

L7 *In the academic context, early career researchers tend to focus on their individual goals.* As social science researchers, the main goal of our domain experts was to publish their research. In the workshops, participants often expressed their focus on publishing and were mostly interested in how our collaboration could help them in that task. The topic of the doctoral students advancing on their dissertations came up often in the discussions about how the developed visualizations could help the researchers. Although every researcher knew that co-creating a visual information system for social science researchers was one of the goals of the project, senior researchers were more interested in co-creating the visualizations than junior researchers.

3.6 CONCLUSIONS

We presented a co-creation case study with 14 social science researchers to design a visual information system that supports their research tasks. We described the co-creation process together with the design requirements, and a first evaluation of the design method. We documented our experience and reflected on it to present the lessons we learned through this design study.

Participants felt listened to, and found the process useful to analyze their needs. They most valued learning about the topic of data visualization, and working together with their peers to learn from each other. However, participants had a hard time accepting the role of a co-creator, i.e. taking responsibility in the design process, instead of simply receiving a product. This led us to realize that participation and involvement are not necessarily perceived as benefits by the users. As co-creation and other participatory methods become more common in visualization design, it is important to reflect on the influence of the choice of method in the participants and in the outcome. Co-creation led us to a well-received product which participants perceived as the result of their collaborative design efforts. However, we struggled with participant commitment and a formal comparison with other methods is necessary to determine whether the time and effort needed in such an intensive design process is decisive for the user satisfaction and the long-term success of the system. Our work provides insights into the benefits and limitations of using co-creation as a visualization design methodology. Given that our sample size is small, however, the reliability of our findings should be validated with further experiments in future work. Furthermore, we worked in an academic environment and the findings may not necessarily translate to working in other environments or to collaborating with experts from other domains.

4

Mobile and Multimodal? A Comparative Evaluation of Interactive Workplaces for Visual Data Exploration

This chapter is based on the second full paper of the dissertation, published at EuroVis 2022 (Molina León et al., 2022a) and available via open access. The published paper and the chapter are identical.

4.1 Introduction

Nowadays people are using mobile devices more often than desktop computers to access the web (BroadbandSearch.net, 2021). Mobile devices support new interaction scenarios and more input modalities, which have the potential to change the way we interact with data. They enable direct manipulation through touch, are lightweight, and portable. However, they also come with challenges, such as a small screen size and less precision due to the “fat finger” problem. Therefore, designing visualization systems for mobile devices has become an increasingly im-

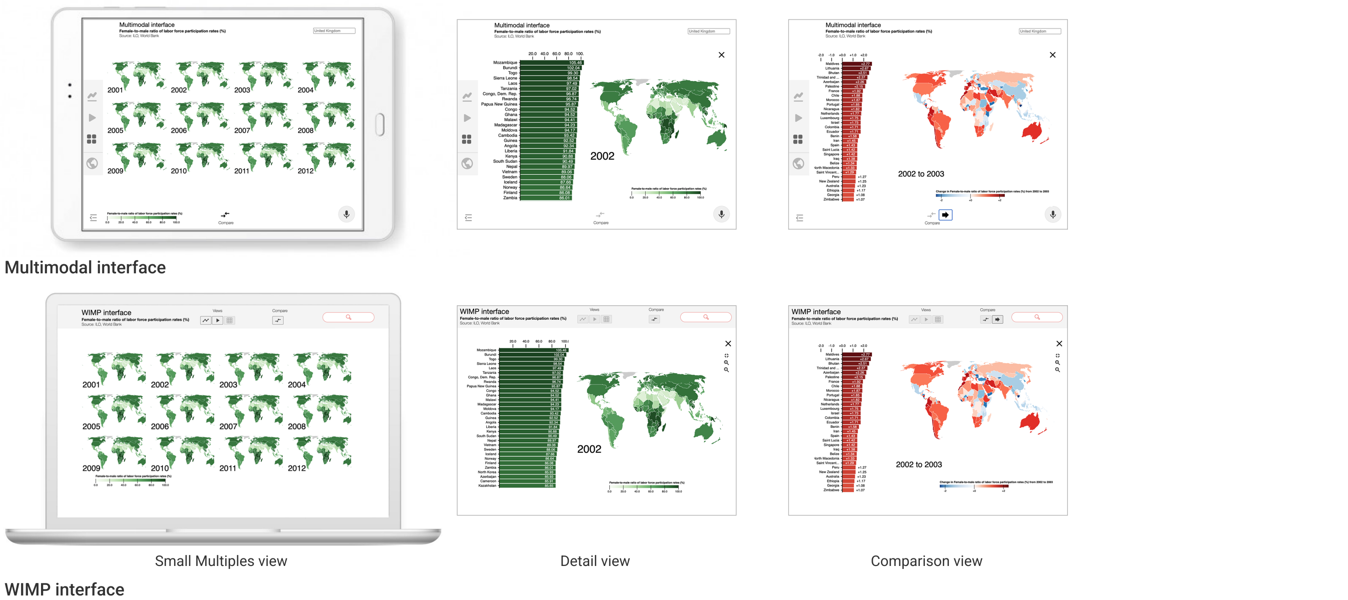


Figure 4.1: The small multiples view, detail view and comparison view of the multimodal and WIMP interfaces compared in our study.

portant research goal (Lee et al., 2012). Particularly tablets are a promising medium for visual data exploration. They have a comparable performance to compete with desktop computers and are increasingly used at work (Jesdanun, 2016). Accordingly, standard visualization techniques, such as bar charts (Drucker et al., 2013), scatterplots (Sadana and Stasko, 2014), and stacked graphs (Baur et al., 2012), have been adapted to tablets with touch interaction. Furthermore, Drucker et al. (2013) found that touch interaction can lead to better performance and user experience on tablets than interactions based on the standard WIMP (window, icon, menu, pointer) metaphor.

Mobile interfaces are called post-WIMP as they are designed differently according to the screen size and input modalities available. Possible modalities include pen (Lee et al., 2015), touch (Baur et al., 2012), and speech (Srinivasan et al., 2020). Hinckley et al. (2010) found that combining pen and touch is both powerful and perceived as more natural. Recent work has combined more modalities to explore data in a "more fluid interaction experience" but most of these systems were evaluated with students or software company workers (Lee et al., 2015; Saktheeswaran et al., 2020; Srinivasan et al., 2020) who would not use them regularly and did not frequently use a tablet. At the workplace, tablets are often used during meetings and could thus become valuable for data exploration. Although mobile visual applications like Tableau Mobile (Tableau Software, 2022a) are already on the market, these are mainly designed for touch and do not leverage multimodal interaction. We do not yet know enough about how tablet-based multimodal visualizations could be used in a work setting, and how they differ from their desktop WIMP counterparts regarding performance and user experience. How would domain experts analyze data on a tablet? How would their performance vary compared to working on the standard desktop setup? How would the experts make use of multimodal interaction? Would they approach their tasks differently?

To compare these workplace setups, we created a visualization system with two interfaces. Following previous work, we designed one interface for tablets supporting multimodal interaction and another for desktop computers supporting WIMP interaction. We consider each interface to be the representation of a setup as its design is based on the corresponding device and input modalities. We compare how domain experts perform exploratory data analysis in those setups by focusing on two research questions:

RQ1 How do the devices and the interaction modalities (mouse and keyboard vs. touch, pen, and speech) affect the performance of the domain experts, in terms of accuracy and response time?

RQ2 How do the devices and the interaction modalities affect the user experience?

We investigate these questions by conducting a within-subjects user study with experts from the social science domain, who explore data as part of their everyday work life. Data exploration plays an important role in the social sciences, in which the increasing availability of open data about government policies and development trends has motivated a growing interest in interactive data visualizations. These data are often spatio-temporal by nature: each data point represents the value of an indicator such as life expectancy or unemployment rate, associated to a country and to a time step. As detailed in section 4.3, we worked together with social scientists from the field of comparative politics to design a system for exploring development indicators, and created two different interfaces for tablets and desktop computers. Our goal was to investigate how the domain experts could work with multimodal visualizations on a tablet, and how their experience differs from conducting the same tasks on a more familiar WIMP environment. According to the interests of the experts, we focused on supporting the exploratory analysis of spatio-temporal data. We followed the task typology of Andrienko et al. (2003a) and Andrienko and Andrienko (2006) for defining exploratory tasks that fit their workflow.

Research on multimodal visualizations has focused on qualitatively evaluating the designed systems so far (Saktheeswaran et al., 2020; Srinivasan et al., 2020). We complement their work by looking also at quantitative metrics such as completion time and accuracy, in comparison with WIMP-based visualizations. We combine these metrics with the analysis of interaction logs and qualitative feedback to compare the performance and experience of the experts in both conditions. Accordingly, we conducted a semi-remote user study with 16 social scientists. Participants were significantly faster and made less errors with the WIMP interface, but were slightly more accurate solving synoptic tasks on the multimodal interface. We found that participants interacted significantly more on the tablet, and pen interaction was particularly appreciated and beneficial. The interaction analysis revealed that the smaller screen size of the tablet did not lead the experts to zoom more often but rather to approach the tasks differently. The experts had different strategies across devices and usually chose specific input modalities for individual actions. Ten participants preferred the multimodal interface, and 15 could imagine using it at work. According to our results, social scientists are interested in working with multimodal visualizations on a tablet, and they could perform as well or better than on a desktop computer after getting familiar with the input modalities.

With this paper, we contribute our quantitative and qualitative findings on how domain experts explore data differently with multimodal visualizations on a tablet, in contrast to using their desktop WIMP counterparts. We identify the different interactions patterns and strategies, and accordingly, provide recommendations for the interaction design of multimodal visualizations for tablets.

4.2 Related Work

In this section, we present previous work on the topics of exploratory analysis of spatio-temporal data and post-wimp interaction for data visualization, and how our research connects to it.

4.2.1 Exploratory Analysis of Spatio-Temporal Data

Andrienko et al. (2003a) devised a typology of exploratory tasks for spatio-temporal data focused on time identification and comparison, at the elementary search level (individual time steps) or at the general level (intervals). The latest classification considers tasks either *elementary* (about individual elements) or *synoptic* (set of elements) (Andrienko and Andrienko, 2006). The authors describe exploratory data analysis as discovering properties of the dataset as a whole, mainly through synoptic tasks. They recommend using different visualization techniques depending on the task. We defined the visualizations of our system based on their recommendations. Furthermore, we designed a series of elementary tasks and synoptic tasks for our experiment based on the work tasks of the domain experts.

Boyandin et al. (2012) conducted a qualitative study on exploring temporal changes with flow maps through animation and small multiples. With animations, participants made more findings related to geographically local events and changes between subsequent years. With small multiples, they made more findings about long time periods. Thus, the authors suggest using both techniques to increase the number and diversity of the findings. Brehmer et al. (2020) confirmed this recommendation on mobile phones. We designed the views of our system following their suggestions.

4.2.2 Post-WIMP Interactions for Data Visualization

In the last decade, visualization researchers have been increasingly investigating the design of visualization "beyond the desktop" (Lee et al., 2012). One of the most well-known studies on this topic is the work by Drucker et al. (2013) on designing and comparing two interfaces of a bar chart application for tablets. The researchers designed one interface based on WIMP elements and another interface focused on using touch gestures. In their study, participants were significantly faster and preferred the gesture-based interface. Inspired by their work, we make a similar comparison but take both tablets and desktop PCs into account, adding pen and speech input.

Sadana and Stasko (2016) designed a multiple coordinated views application combining WIMP elements and touch gestures. One of the main challenges was to define a set of consistent gestures across views. We designed our system with similar design principles (see section 4.4). We have multiple views, and we aimed at having consistent interaction techniques over specialized ones.

Oviatt et al. (1997) investigated first how multimodal interaction could support map-based tasks. They found that participants preferred pen interaction to draw symbols and would write with the pen before using speech, in combined interactions. Similarly, we combine pen and speech input, and supplement it with touch, given its relevance for tablets. More recently, Jo et al. (2017) surveyed 13 studies on leveraging pen and touch interaction. They found that the five most common touch gestures in tablet-based studies were drag, tap, pinch, long press, double tap, and lasso selection. Accordingly, we limited the touch gestures of our multimodal interface to those. On the WIMP interface, we use standard interactions such as click, double click, and drag.

Much of the visualization research on leveraging multiple input modalities has focused on the combination of touch, pen, and speech. Srinivasan and Stasko (2018) created a system for exploring networks with speech and touch. Saktheeswaran et al. (2020) compared said interface with its unimodal counterparts. The participants preferred multimodal input due to having more freedom of expression, and the option to combine modalities. Srinivasan et al. (2021) created a system for a large vertical display combining the three modalities and found that they complemented each other well in complex operations. Srinivasan et al. (2020) surveyed 18 visual systems to collect interaction techniques using touch, pen, or speech input. Based on that survey, they proposed a set of multimodal interactions for visualizations on tablets. We used their results as a base for choosing the multimodal interaction techniques.

4.3 Social Science Domain

As part of a multidisciplinary collaboration, we worked with social scientists to support them in the exploratory analysis of their spatio-temporal data. The scientists are members of a research project which explores the evolution and diffusion of social policy across the globe from 1850 until today. The policies are measured by indicators, such as health care expenditure or unemployment rate, and help the experts assess the development of nations (Castles et al., 2010). Accordingly, the experts want to answer questions such as “how does health expenditure vary across world regions?”, or “how did the unemployment rate in every country change over time?”. The social scientists were already working with data visualizations provided

Table 4.1: Question examples classified according to the typology of Andrienko and Andrienko (2006).

Task question	**Type**	**Sub-type**
How high was the child mortality in Peru in 2001?	Elementary	Direct lookup
Which country had the highest female-to-male labor ratio in 2009?	Elementary	Inverse lookup
Is the child mortality in Myanmar lower, higher or equal to the one in Cambodia in 2009?	Elementary	Direct comparison
How was the female-to-male ratio in Western Europe in 2004?	Synoptic	Behavior characterization (space)
How did the child mortality develop in Northern Africa until 2007?	Synoptic	Behavior characterization (space over time)
In which African country did the female-to-male ratio increase most in the first five years?	Synoptic	Pattern search (time)
In which continent did the child mortality decrease most over the whole time period?	Synoptic	Pattern search (space over time)
How did the 2003 female-to-male ratio in South America compare to the one in Southern Africa?	Synoptic	Direct comparison (space)

by international organizations such as Group (2022), but wished for custom visual tools that would facilitate the recognition and the comparison of spatio-temporal patterns in the data.

We conducted a series of co-creation workshops and contextual interviews with the social scientists to explore visualization opportunities. We chose co-creation as a design methodology to empower the domain experts to actively shape the tools they wished for and to continuously validate the design (Landstorfer et al., 2014b; Molina León and Breiter, 2020). In the workshops, we found that the main interest of the experts was facilitating the first steps of their exploratory analysis, where they would look for countries and time spans of interest to focus on. Given that the indicators they work with often have varying temporal coverage and country samples, they were looking for options to explore the spatio-temporal coverage of the data, to recognize relevant patterns, and to compare data points over space and time. During the two-year collaboration, we co-designed web-based visual tools to explore their data (e.g., Molina León et al., 2020). Furthermore, we observed that many of the experts owned and used a tablet at work to take notes and draw diagrams. Motivated by this observation and previous work on ubiquitous visual analytics (Badam et al., 2019; Srinivasan et al., 2020), we decided to provide a tablet-based interface for one of the tools. We conducted the user study more than a year after the last workshop. Some of the experts participated in both. The study was where they saw the evaluated system for the first time.

4.3.1 Data

The spatio-temporal data the experts work with is often relative data. Such a dataset typically consists of a set of triples of country, year, and value, e.g., (Chad, 2001, 11.3%). For the study, we chose datasets relevant to the social scientists published by Group (2022), one of their primary data sources. On each version, we visualized one of two datasets covering a 12-years time span (2001–2012): child mortality rate per 1000 live births (Roser et al., 2013), and female-to-male ratio of labor force participation rates (Ortiz-Ospina et al., 2018). We counterbalanced the interface order and dataset assignment following a Latin square design.

4.3.2 Tasks

We defined the study tasks based on the exploration tasks that the social scientists described in the workshops, and on examples from the related study of Duncan et al. (2020). We needed clearly defined tasks to compare the performance and user experience of the social scientists across setups, and therefore, decided against an open exploration. According to the typology of Andrienko and Andrienko (2006), we cover the following task types for the exploratory analysis of spatio-temporal data: (1) Direct lookup (elementary), (2) Inverse lookup (elementary), (3) Direct comparison (elementary), (4) Behavior characterization (synoptic), (5) Pattern search (synoptic), and (6) Direct behavior comparison (synoptic).

For the synoptic tasks, multiple variants were possible according to the reference sets: space, time, or space over time. Different combinations of the sets would lead participants to approach each task differently. In the study, we presented 13 tasks per dataset and interface: five elementary and eight synoptic tasks. We selected more synoptic tasks because those were predominant among the examples given by the experts. Given that interactivity played no significant role on the effectiveness of solving elementary tasks in previous work (Duncan et al., 2020), the predominance of synoptic tasks suggested that this case study was suitable for comparing input modalities. Examples for each of the main question types are shown in Table 4.1.

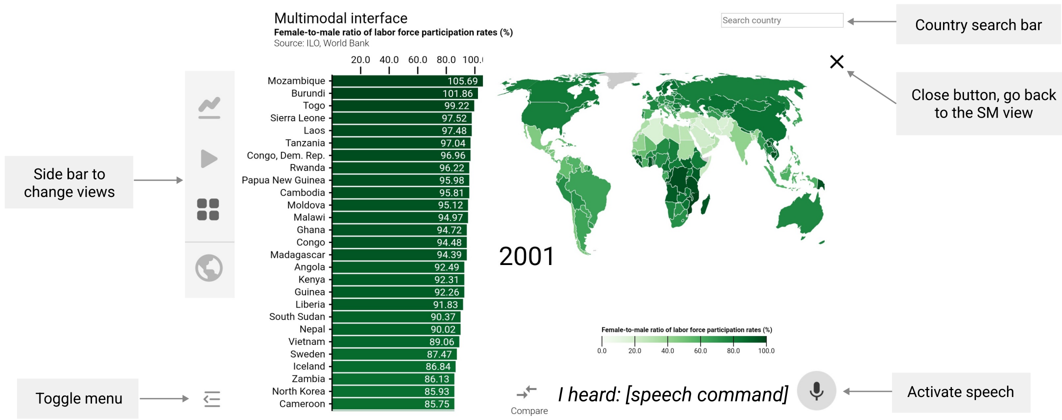


Figure 4.2: Detail view of the multimodal interface. It presents the country values of a chosen year. On the left, the bar chart shows the 2001 values of each country, sorted in descending order by default. On the right, the choropleth map shows the data with its geographical location.

The full lists of questions are included in the supplementary material. Following the conceptual framework of Peuquet (1994) for spatio-temporal dynamics, we formulated the exploratory tasks based on variations of its triad elements: *what*, *where* and *when*. In particular, we focus on varying the *where* and *when*.

4.4 Visualization and Interaction Design

In the following sections, we describe our design principles, the design of our multimodal system, and its WIMP counterpart. We aimed to create two interfaces with equivalent functionality and standard interaction techniques to make a fair comparison.

We present our design principles below, based on design reflections and findings on strategies to visualize spatio-temporal data (Boyandin et al., 2012; Brehmer et al., 2020; Peña-Araya et al., 2020; Robertson et al., 2008) as well as on the interaction design of tablet-based visualizations (Drucker et al., 2013; Jo et al., 2017; Sadana and Stasko, 2016).

DP1 *Leverage standard interaction techniques of multimodal systems*. Srinivasan et al. (2020) surveyed work on multimodal visualizations to determine what the standard interaction techniques are. To make a fair comparison, we mapped our interaction techniques according to those. We surveyed relevant examples to define the interaction techniques of choropleth maps because they were not included in the survey.

DP2 *Leverage standard interaction techniques of WIMP interfaces*. Given that line charts, bar charts and choropleth maps are commonly used visualizations, there are multiple well-known tools and examples that offer similar interaction techniques. We surveyed them to define the techniques of our system.

DP3 *Use standard touch gestures*. Familiar gestures are easier to remember and are usually preferred on touch-based visualizations (Drucker et al., 2013; Jo et al., 2017; Sadana and Stasko, 2016). We avoid complex gestures because discoverability is an issue on touch interfaces (Baur et al., 2012; Drucker et al., 2013) and such gestures can be hard to remember.

DP4 *Achieve interaction consistency*. Users expect that a gesture triggers similar results on different features of a system. Previous work on multiple coordinated views has emphasized the need for consistent gestures across views (Sadana and Stasko, 2016). Accordingly, we put together a set of consistent interactions through all views.

DP5 *Introduce WIMP elements when necessary*. On the multimodal interface, we added redundant WIMP elements to ensure a good experience, following the findings of Drucker et al. (2013). For example, we enabled speech commands to switch views, but also included a side menu to do the same, to make sure that critical interactions could not be limited by speech recognition errors.

We created a first version of the system based on the requirements we elicited in the workshops. After finishing that version, we conducted an expert review with two HCI researchers who own and use a tablet regularly. After improving the system according to their feedback, we conducted an exploratory study with seven participants to further refine the system. In the following sections, we present the final design that resulted from those iterations.

4.4.1 Views and Visualization Techniques

We follow the suggestions of Andrienko et al. (2003b) on the visualization techniques to solve the elementary and synoptic tasks that we tackled. Overall, the system includes the following views:

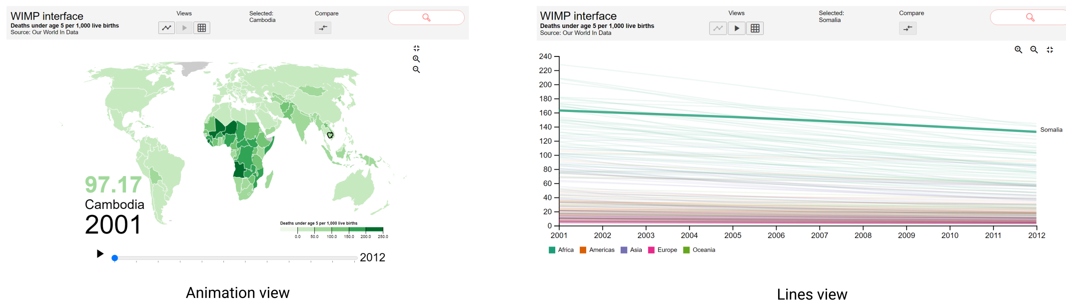


Figure 4.3: Animation view (left) and Lines view (right) of the WIMP interface, showing the child mortality dataset.

1. *Lines view*. We included a multi-line chart to support comparison and behavior characterization tasks, focused on time (Andrienko and Andrienko, 2006). This view gives an overview of the whole dataset and shows time trends (see Figure 4.3).

2. *Animation view*. We included animated choropleth maps for visualizing the temporal behavior of the spatial behavior (Andrienko and Andrienko, 2006) (see Figure 4.3). We chose choropleth maps because many datasets used by the social scientists represent *relative* values.

3. *Small multiples view*. Small multiples provide an overview of the data and allow to visually compare it at different time points (see Figure 4.1). For propagation tasks, small multiples of choropleth maps perform better than other alternatives (Peña-Araya et al., 2020).

4. *Detail year-based view*. We created the detail view for exploring the data distribution for a specific year. It includes a choropleth map to recognize the spatial distribution, and a sorted bar chart that helps identifying the countries that perform best or worst. The detail view of the multimodal interface is shown in Figure 4.2.

5. *Comparison view*. This view facilitates the comparison of two time steps. A choropleth map and a bar chart show the data values derived from calculating the difference between the two steps. The comparison view is depicted in Figure 4.1.

4.4.2 Interaction Techniques

Table 4.2: System actions with their corresponding interactions on each interface. Taps can be with either a finger or the pen. If applicable, the interaction includes the references to previous research or existing systems that it is based on.

Action	WIMP interaction	Multimodal interaction
Go to a main view (Lines, Animation, SM)	Click a view button.	Speech command, e.g., "Go to line chart" (Srinivasan et al., 2020). Tap on a view button.
Go to detail view	Click on a small multiple (SM) (Elzen and Wijk, 2013).	Tap on a SM.
Go to comparison view	Click on Compare button. Then, click on two SMs. Then, click on View comparison button.	Tap on Compare button. Then, tap on two SMs. Then, tap on View comparison button.
Select a country	Click on map (Tableau Software, 2022b). Click on bar (Tableau Software, 2022b). Type country name in search bar (Setlur et al., 2016).	Tap on map (Srinivasan et al., 2020). Tap on bar (Drucker et al., 2013; Srinivasan et al., 2020). Speech command, e.g. "Select Italy" (Saktheeswaran et al., 2020; Srinivasan and Stasko, 2018). Draw lasso (Saktheeswaran et al., 2020; Srinivasan et al., 2020).
Deselect a country	Click on selected country on map (Tableau Software, 2022b). Click on selected bar on bar chart (Tableau Software, 2022b).	Tap on selected country on map. Tap on selected bar on bar chart (Jo et al., 2017; Tableau Software, 2022b).
Select a year	Click on SM map.	Tap on SM map.
Deselect a year	Click on selected SM map.	Tap on selected SM map.
Deselect all	Click on empty space (Tableau Software, 2022b).	Speech command "Deselect all" (Srinivasan and Stasko, 2018). Tap on empty space (Srinivasan et al., 2021; Tableau Software, 2022b).
Get country value	Hover on the map to show tooltip (Datawrapper GmbH, 2021; Tableau Software, 2022b). Hover on a country line (Datawrapper GmbH, 2021; Tableau Software, 2022b).	See select (only possible for selected countries) (Tableau Software, 2022b). Drag pen over the x-axis of the line chart (Srinivasan et al., 2020).
Zoom	Click on any of the zoom buttons (Datawrapper GmbH, 2021; Tableau Software, 2022b). Move mouse wheel (Google, 2021; Tableau Software, 2022b).	Pinch gesture to zoom in or out (Datawrapper GmbH, 2021; Srinivasan et al., 2020).
Pan	Drag the view with the mouse (Google, 2021; Tableau Software, 2022b).	Drag with one finger (Google, 2021; Srinivasan et al., 2020).
Sort	Double click on x-axis of bar chart.	Swipe left or right on x-axis of bar chart (Drucker et al., 2013).

To decide on the interaction techniques of the maps, we surveyed interactive mapping tools such as Google Maps (Google, 2021) and Apple Maps (Apple Inc., 2021). Additionally, we inspected tools like Datawrapper (Datawrapper GmbH, 2021) that support the authoring of the visualization techniques we included. We defined our initial set of interaction techniques following the common features we recognized across tools, taking into account our design principles. We restricted our touch gestures to the standard set: tap, double tap, drag, swipe, and pinch.

The tablet interface was designed for touch, pen, and speech interaction, while the PC version for mouse and keyboard interaction. An overview of the interaction techniques is presented in Table 4.2. We show both interfaces with a few exemplary interaction techniques in the supplementary video. For navigating between views, we included buttons on both interfaces to make the interactions succinct and easy to discover (**DP5**). Both versions support brushing and linking. Overall, the WIMP actions are based on the tools we surveyed (**DP2**), and the multimodal actions follow the recommendations of Srinivasan et al. (2020) (**DP1**).

On the tablet, tapping is possible with either the pen or a finger. In the study of Srinivasan et al. (2020), participants had a hard time differentiating pen and touch. Although the authors attributed this to a lack of experience with pen input, several of the domain experts we co-designed the system with were regular tablet users. They also coincided on enabling actions that could be possible with either the pen or a finger without limitations. We decided to allow for both to compare the setups according to how the experts would actually use them in their everyday life.

4.4.3 Implementation

We implemented both versions of the system as a web prototype with D3.js (Bostock et al., 2011) working with a Samsung Galaxy Tab S3 and an EIZO 23.8-inch desktop monitor. We used the standard HTML5 web speech recognition API (Mozilla and contributors, 2019). The prototype is available at: https://cocreation.uni-bremen.de/workplaces.

4.5 User Study

Our goal was to compare the performance and the user experience of the domain experts with each interface. We wanted to gain a better understanding of how the interaction design influences the data exploration. Based on previous work and our research questions, we established the following hypotheses for the experiment:

H1 *The experts will need more time on the multimodal interface*. Natural User Interfaces (NUIs) are believed to be more engaging than WIMP interfaces, enabling a “more natural” interaction (Srinivasan et al., 2020) that encourages people to explore. Furthermore, multimodal interfaces are still rare and even if touch interaction with a smartphone is nowadays common, the combination of touch, pen, and speech is still new to many. This is why, we believe participants will need more time on the tablet.

H2 *The experts will make fewer errors on the WIMP interface*. We expect participants to be more accurate on the WIMP interface because it is the type of interface they already use at work. On the tablet, direct manipulation may lead to difficulties with precision, as it happens with the “fat-finger problem” (Drucker et al., 2013). Using a pen may compensate for this limitation of touch input because pens have proven to convey precise spatial information on map-based tasks (Oviatt, 1997). Still, we believe that familiarity with WIMP interfaces will lead to better spatial accuracy overall. As Duncan et al. (2020) found that interactivity has a larger impact on accuracy for synoptic tasks on cartograms, we expect the accuracy difference to be large for the synoptic tasks.

H3 *Participants will prefer the multimodal interface*. In previous studies, participants preferred multimodal over unimodal interaction. However, these comparisons included only pairs of modalities: pen and speech vs. pen-only and speech-only (Oviatt et al., 1997), and touch and speech vs. touch-only and speech-only (Saktheeswaran et al., 2020). Recent work has shown that multimodal interaction can enhance the user experience and improve usability (Srinivasan and Stasko, 2018). We continue the research by considering pen, touch, and speech together, and expect multimodal interaction to be preferred due to a more engaging user experience.

4.5.1 Experimental Design

We applied a within-subjects design, where each expert interacted with both interfaces to explore a dataset. We used different datasets for each interface and all combinations of interface and dataset were in counterbalanced order, as described in subsection 4.3.1.

To measure performance and user experience, we prepared 13 tasks for the experts to solve on each setup (see subsection 4.3.2). Each task consisted of a question about the given dataset with three possible answers, similar to previous studies (Brehmer et al., 2020; Duncan et al., 2020). Only one answer was correct. We formulated the questions in a way so that only the years, and the countries or regions, changed between datasets. We aimed to mention all continents equally often, to avoid focusing on a region that participants may be familiar with.

We created an online survey in which we measured the response time as the time between arriving at the task page and clicking on the "Next" button. We measured accuracy as the error rate. Each wrong answer was counted as one error. The study had four parts: (1) Consent and demographics, (2) Introduction and tasks with the first interface and dataset, (3) Introduction and tasks with the other interface and the other dataset, and (4) Comparison survey. The tasks, surveys, and data are included in the supplementary material.

First, we explained the motivation of the study to the participants and asked for their written consent to record the session. They proceeded to answer a series of demographic questions and to perform a color vision test to make sure that they could correctly distinguish colors, similar to Duncan et al. (2020). Then, we shared a web link for accessing the system and provided them with a slide deck that we had prepared for the corresponding interface (see supplementary material). The slides described the views and the interaction techniques. We asked participants to read them and to perform all interactions on the system while reading. We did this to make sure that every participant received the same information. Then, we gave them five minutes to interact freely. If there were any questions, we discussed them. When participants confirmed that they felt confident enough, we proceeded with the tasks. We asked them to think aloud while solving the tasks to better understand their interactions. They did it on both interfaces so the performances were comparable. Afterwards, we asked participants to rate their satisfaction with the System Usability Scale (SUS) questionnaire (Brooke et al., 1996). Subsequently, we asked what they liked, what they disliked, and what they missed about the corresponding setup. At the end of the study, we asked participants what interface they preferred, and to mention one to three reasons for their choice.

Due to the COVID-19 pandemic, we conducted the study semi-remotely. We met participants shortly before and after the session to provide and collect the tablets. During the session, we communicated through a video conference tool. They answered all questions on their office computers, and we recorded their interactions via a tablet app and the video conference tool. The computers were provided by their employer and included a 23.8-inch monitor.

4.6 Results

We recruited 16 participants (eight female), their average age was 32 years. All participants were social scientists from diverse disciplines, mainly political science and sociology. They worked on topics such as international trade and welfare policies. Fourteen of them were researchers, and all had a Master's degree.

Eight participants reported that they work with data visualizations weekly. Ten participants interacted with touch devices daily, while only one interacted with pen-based and speech-based systems daily. For five participants, this was the first time using a pen as an input device. For seven, it was the first time using speech input. All participants spoke English fluently but none was a native speaker. Nine of them owned a tablet. Since this was more than half of the participants, we adjusted our experiment design to conduct between-subjects comparisons of tablet owners and non-owners.

4.6.1 Performance

Participants took longer to solve the tasks with the multimodal interface. Their accuracy was slightly better on the WIMP interface. We detail these results in the following sections. In the statistical tests, we considered a difference significant when the p-value was below 0.05, and we report Pearson's correlation coefficient r as the effect size to provide a measure of the importance of the effect.

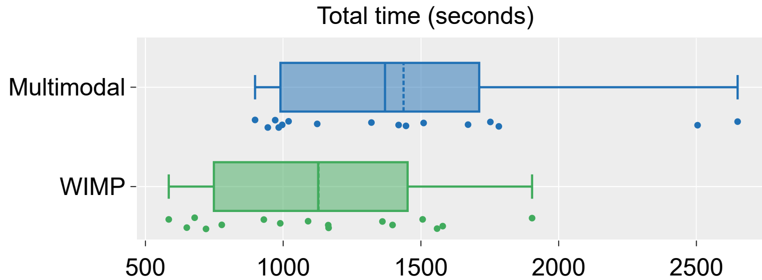


Figure 4.4: Total response time per interface.

Response Time

Figure 4.4 shows the total response time for each session per interface. The mean response time per task was 86.8 seconds with the WIMP interface and 110.54 seconds with the multimodal interface. After confirming that the difference between the response times was normally distributed, we ran a one-tailed t-test. The result showed that the participants were significantly faster on the WIMP interface with a medium to large effect ($t(15) = 1.83, p = 0.043, r = 0.43$). Therefore, **H1** is supported. Looking at individual tasks, the time difference was larger on T4 and T11, which were synoptic tasks. Based on these observations, we compared the time of elementary tasks with the time of synoptic tasks. There was no significant difference between the response times across task types.

In addition, we compared tablet owners and non-owners to find out whether owning a tablet had an impact on their response time with the multimodal interface. Tablet owners took longer ($M = 24.88$ minutes, $SD = 9.05$) than non-owners ($M = 22.75$ minutes, $SD = 9.39$), but not significantly ($W = 25, p = 0.54, r = -0.21$).

Accuracy

Participants made less errors with the WIMP interface than with the multimodal interface. They solved 86.54% of the tasks correctly with the former and 85.1% with the latter. The error distribution is shown in Figure 4.5. Given that the distribution was not normal, we ran a Wilcoxon signed-rank test to investigate the differences. Participants were not significantly more accurate on the WIMP interface ($W = 33.5, p = 0.39$), so **H2** is not supported by our experiment. We tested whether the interface order had an effect and found no significant differ-

Figure 4.5: Distribution of the errors per interface.

ence ($t = -0.31, p = 0.76$). There was also no significant effect of the dataset ($t = 1.32, p = 0.21$).

We also tested whether owning a tablet had an impact on the accuracy of the participants with the multimodal interface. Tablet owners actually made slightly more errors ($M = 2.22$) than non-owners ($M = 1.57$) but the difference was not significant ($W = 35.5, p = 0.70, r = 0.13$). Additionally, we tested whether the accuracy of the participants was significantly better on the WIMP interface for each task type, to investigate whether our results fit the results of Duncan et al. (2020). On average, participants made more errors on the synoptic tasks ($M = 1.28$) than on the elementary tasks ($M = 0.56$), which corresponds to their difficulty level. For the elementary tasks, participants made less errors on the WIMP interface, but not significantly. For the synoptic tasks, participants made less errors on the multimodal interface, but the difference was not significant ($U = 128.0, p = 0.49, r = -0.05$).

4.6.2 Interactions

We recorded the interactions of the participants with screen video and audio recording. Then, we logged and coded the interactions per session. For each interaction, we documented the participant ID, the device, the dataset, the task, the view, the input modality, the action (see Table 4.2), the command or gesture, and the outcome. We successfully logged the interactions of 14 participants in 28 videos (two videos per person). For two participants, there were technical issues that did not allow to record their interactions properly.

We logged 4087 individual interactions. Each interaction corresponds to an attempt to perform an action, e.g. pinch to zoom. We classified its outcome as either *successful* (i.e. the system reacted to the action as it should have), *erroneous* (i.e. the

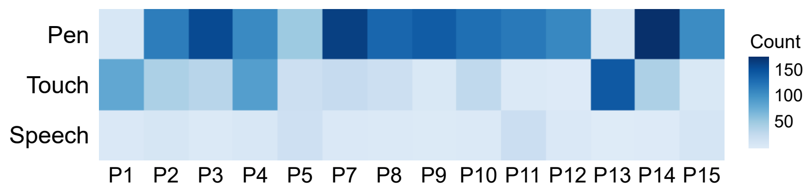


Figure 4.6: Total of interactions per input modality and per participant on the multimodal interface.

system did not react as it should have), *invalid* (i.e. the interaction is not valid in the current view or state) or *unsupported* (i.e. the interaction is not one of those the system recognizes). Overall, the success rate of the interactions was 0.94. The erroneous rate was 0.03, and the rates for unsupported and invalid were 0.02 and 0.01. Most errors happened during selection on the tablet due to speech recognition errors. A common issue was that participants were unsure about the English pronunciation of country names, and this led to errors.

We logged 2197 interactions on the multimodal interface and 1890 on the WIMP interface. On the WIMP interface, participants performed 1697 (89.8%) mouse interactions and 193 (10.2%) keyboard interactions. On the multimodal interface, the pen prevailed: 1544 (70.3%) interactions were pen-based, 554 (25.2%) were touch-based and 99 (4.5%) were speech-based. The dominance of the pen over touch is surprising because 10 participants used pen interaction rarely or never, but most used touch daily. According to the qualitative feedback, participants liked the pen because of its high precision, and the ability to select by drawing.

We show the total of interactions per modality and participant on the tablet in Figure 4.6. Every participant tried each input modality at least once. However, most had one dominant modality. Of these 14 participants, 11 mostly used the pen to interact, two mostly touch, and one almost equally used both pen and touch. Among the 11 participants who mostly used the pen, four had never interacted with pen-based systems, two interacted with it less than once per month, and five at least monthly. This suggests that the tendency to use the pen was independent of the frequency of pen use in their everyday life. Of the 11 participants, seven owned a tablet.

Table 4.3: Interactions per interface and input modality of the most common actions.

Interface	Multimodal			WIMP	
Action	**Pen**	**Touch**	**Speech**	**Mouse**	**Keyboard**
Go to Animation view	47	1516	0	56	0
Go to Comparison view	247	41	0	156	0
Go to Detail view	101	19	0	109	0
Go to Lines view	39	21	0	72	0
Go to SM view	213	41	0	246	0
Select country	353	87	95	234	193
Deselect country	93	9	0	69	0
Deselect all	89	21	3	56	0
Get country value	9	8	0	286	0
Zoom	0	114	0	125	0
Pan	202	93	0	37	0

Actions and Views per Interface

Previous work suggests that multimodal interaction is more engaging and consequently leads to more interactions (Srinivasan and Stasko, 2018). Thus, we compared the interactions per participant across devices. On average, participants interacted significantly more with the multimodal interface ($M = 156.93, SD = 40.27$) than with the WIMP interface ($M = 135.00, SD = 36.19$), $t(13) = 1.85, p = 0.046, r = 0.45$.

We show what type of actions participants performed with each input modality in Table 4.3. Looking at the actions across devices, participants selected more often on the tablet, and performed the *Get a country value* action more often on the desktop PC. This makes sense given that country values were visible while hovering, but getting values on the tablet required selecting first. *Pan* was the second most often action on the tablet, mostly on the bar charts.

Regarding the views, participants interacted most often on the *Detail* view on the tablet (29.30%), and on the *Small multiples* view on the desktop (26.95%). The larger size of the visualizations of the detail view may have been more important on the smaller screen. Participants interacted more with the *Animation* view on the tablet (21.98%) than on the desktop (13.98%), but used the *Lines* view more on the desktop (19.60% vs. 6.59%). The screen size and the possibility of hovering are the most likely reasons for this.

Interactions per Task

More than half of the participants interacted with the pen on every task, either alone or in combination with speech or touch input. In contrast, only one person used touch on every task. We compared the interactions based on the accuracy of the participants on the tablet. Participants who solved a task correctly interacted more with the interface. Looking at the combination of input modalities used per task, we found that for the tasks solved correctly, the most common interaction was the pen only (35.48%) followed by the combination of pen and touch (30.97%). On tasks answered wrongly, the most common combination was pen and touch (33.33%). Interacting with the pen alone is thus associated with better accuracy.

By analyzing how participants successfully solved each task, we found the following patterns across devices:

1. *On the tablet, participants used most views with larger maps. Pen selection on the map, panning on the bars*. For tasks about specific countries, participants often went to the Detail or Animation view for selecting them with the pen, based on where they thought the country was. They would sometimes combine this with zooming on the map to make a more precise selection, and with panning on the bar chart to get an idea of the relation of that country to others. This was the most common pattern on the tablet. For tasks where two time steps were involved, a variation of this pattern would take place on the Comparison view.

2. *For time intervals, most used the line chart on the PC, and the comparison view on the tablet*. When comparing the views used for solving temporal development tasks across, it was noticeable that people preferred to use the *Comparison* view on the tablet and the *Lines* view on the PC. This reveals that participants had different ways to solve the same task across devices.

3. *Hovering was key to solve most tasks on the PC*. On the WIMP interface, participants selected less often and compensated with hovering. Some participants solved most tasks with the *Lines* view and mainly interacted by using the search bar to select a country, and then hovering to inspect its values.

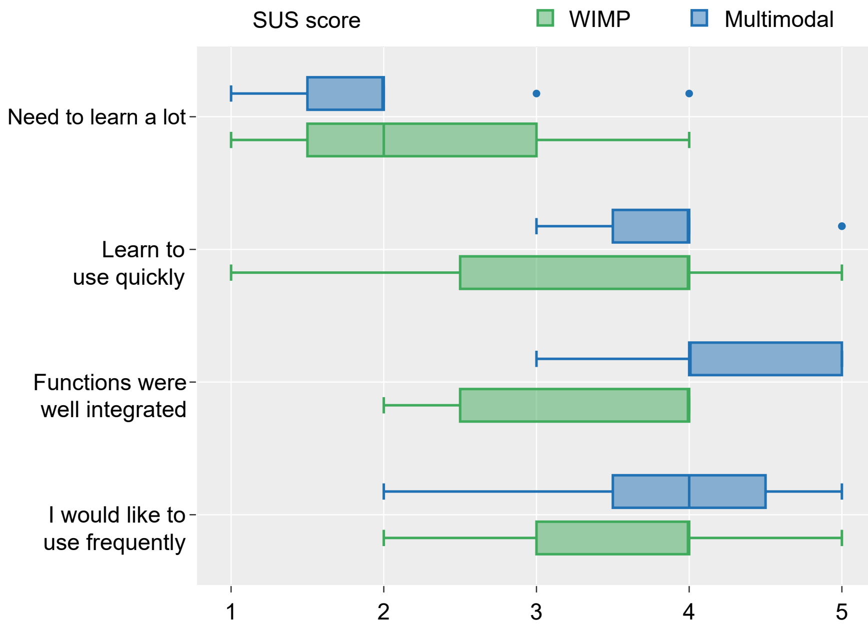


Figure 4.7: Score distribution of the four SUS statements for which the answers differed most between the two interfaces. The score range was from 1 (strongly disagree) to 5 (strongly agree).

4.6.3 User Experience

We asked participants to rate their experience based on the standard SUS questionnaire. Afterward, we asked them a few questions about their impression of the system. As mentioned above for **H3**, we expected multimodal interaction to provide a better experience. Participants considered that the system features were better integrated on the multimodal interface, which is noteworthy considering the multiple input modalities. Furthermore, participants scored the multimodal interface as quicker to learn. We show the scores for the questions with largest difference across interfaces in Figure 4.7.

After interacting with each version of the system, we asked participants what they liked and disliked about it. On the multimodal interface, four participants particularly liked the ability to select with the pen due to its high precision, and the option to draw for selecting. This corresponds to the pen being the most used input modality on the tablet (see subsection 4.6.2). Multiple participants mentioned that the speech input was important for searching for countries whose geographical location they did not know. When asked what they disliked, participants mentioned the speech recognition errors, and selecting countries with their fingers by accident while zooming. On the WIMP interface, participants especially appreciated the possibility to hover to get the tooltip, and the ease of searching for countries with the keyboard. The most common issue was that hovering on the line chart led to highlighting the closest line and this line would sometimes overlap with the line of interest.

In addition, we asked participants whether they could imagine using the system at work. Fifteen answered positively. They highlighted the advantage of performing "quick checks" and "fast comparisons" of their data to answer questions such as "how has the child mortality changed due to the civil war in Syria?". Fifteen participants would use the multimodal interface at work and 14 would use the WIMP interface.

4.6.4 Preferences

Ten of 16 participants preferred the multimodal interface. Therefore, our results support **H3**. Participants focused on different input modalities to justify their choice. Three participants argued that zooming with touch gestures felt easier than with the buttons of the WIMP interface. Three participants especially liked the lasso selection with the pen, and for P4, it felt faster than the mouse. According to P3, the speech input made the multimodal interface "really superior." For three participants, the multimodal interface felt "more intuitive", and one felt more confident with it. For P5, touch interaction was "much more fun than just keyboard and mouse."

Six participants preferred the WIMP interface. Their main reasons were the familiarity of working with mouse and keyboard and the larger screen of the PC. P2 did not own a smartphone, and therefore, felt that he was not skilled enough with touchscreens to perform well on the tablet. P7 preferred the multimodal interface but pointed out that she would rather use the WIMP version at work because tablets with good performance are scarce at the workplace.

4.7 Discussion

Our results indicate that the different combinations of devices and interaction modalities affect the performance (**RQ1**) and user experience (**RQ2**) of the domain experts during exploratory data analysis. The social scientists were significantly faster solving tasks in the familiar WIMP context, but they were similarly accurate on the tablet, and were even better solving synoptic tasks on it. At first sight, it may seem that the smaller screen size of the tablet caused participants to be slower because they may have spent more time zooming and panning. However, the logged interactions presented in Table 4.3 indicate that participants did not zoom more on the tablet than on the PC (114 times on the tablet vs. 125 on the PC). Moreover, the interaction patterns show that participants had different strategies to solve the tasks across conditions. For example, they used the *Lines* view more often on the PC because they could easily hover on the lines. Therefore, we conclude that the interaction modalities were the most decisive factor on the interaction choices of the experts. This is confirmed by their qualitative feedback where they justified their preferences based on the modalities available.

Although most participants had interacted rarely or never with a pen, they used it for most interactions and were successful with it. Pen and speech interaction were especially helpful to select with precision and more comfortable when the country location was unknown. That suggests that each input modality fits best to specific actions, and its benefits depend on the task at hand. In visualizations of large datasets, the precision of the pen may be most valuable to interact with each data point. Furthermore, most participants already used a tablet in the office, but rather for simple tasks such as taking notes. Our results suggest that if multimodal tools are given, domain experts would consider including them into their workflow.

Our results are not as positive for the tablet interface as the ones of Drucker et al. (2013), but given the success of the participants with pen interaction and their comparable accuracy overall, we believe that data exploration on the tablet may become more beneficial and preferred at work, after getting familiar with it. We consider the lack of a significant difference on accuracy as a positive result because pen and speech interaction are still not as common as touch, which makes the accuracy and user experience results promising. Moreover, Drucker et al. compared both interfaces on a tablet while our WIMP condition included a PC because we wanted our results to reflect the real-world experiences of the experts. For the same reason, we used a PC and a tablet with different screen sizes. While the interaction analysis and the qualitative feedback suggest that the modalities were the most decisive factor on the participant choices, the interaction patterns also reveal that participants tended to use the views with larger visualizations on the smaller display. Thus, our findings do not compare the interaction modalities only, but rather the combination of devices and modalities. They describe how the experience is shaped by both factors.

4.7.1 Recommendations for Interaction Design

Our findings suggest three main recommendations for the design of multimodal visualization systems for tablets:

1. Pen interaction was dominant regardless of previous experiences. Thus, the pen should be able to perform most interactions, and all critical interactions should be possible with it.

2. Participants described and appreciated each input modality based on the actions they preferred to perform with it. The pen was notably helpful to select small countries, confirming the findings of Oviatt (1997). Thus, we recommend pen interaction for selecting in map-based visualizations. Given the better performance of touch with bar charts (Drucker et al., 2013), we conclude that performance depends on the modality that suits better the corresponding mix of visualization and interaction techniques.

3. According to the qualitative feedback, speech input was very appealing despite its problems. This is consistent with the findings of Saktheeswaran et al. (2020) on multimodal interaction being less error-prone than speech-only. Leveraging speech interaction may lead to a more engaging experience, but other modalities should support the same actions to guarantee usability.

4.7.2 Limitations

We defined the experiment tasks according to the work of the experts we collaborated with. A study with open exploration would help verify whether interacting significantly more on the tablet is associated with being more interested in exploring multimodally in general. Furthermore, we defined our interaction techniques based on previous work, but adding more complex techniques that combine the three input modalities sequentially or simultaneously would help to learn more about how people interact multimodally.

Due to the COVID-19 pandemic, we conducted the study semi-remotely. We provided the tablets, but the experts used their office computers. Although the monitors were of the same model, this means that the experiment was not fully controlled, and further investigation is needed to confirm our findings. We also asked participants to think aloud. We acknowledge that this may have influenced the results, yet without any clear observable bias in one direction. Furthermore, having a larger sample, and using products provided by a third party, would help to test the reliability of our results.

4.8 Conclusions

We investigated how devices and input modalities affect the performance and user experience of domain experts while solving exploratory tasks on spatio-temporal data. Participants used pen interaction for high precision tasks without having much experience, used touch for zooming, and speech for selecting countries. Our work suggests that combining touch, pen, and speech, is a promising option for visual data exploration, and that different modalities fit better to specific tasks and lead to different interaction patterns.

Although we designed the system for domain experts, we think that exploring such data is also relevant for the general public. Leveraging multimodal interaction may make the exploration process more engaging, but the data literacy of the audience should be taken into account. Furthermore, exploring data with different input modalities may be an opportunity to make visualizations more accessible. The relevance for accessibility should be studied further.

5
Eliciting Multimodal and Collaborative Interactions for Data Exploration on Large Vertical Displays

This chapter is based on the third full paper of the dissertation, accepted for publication at IEEE Transactions on Visualization and Computer Graphics (Molina León et al., 2024b) and available via open access. The published paper and the chapter are identical.

5.1 Introduction

The standard mouse and keyboard devices used to interact with desktop computers are not as well suited to interact with large vertical displays due to the larger screen size, potentially changing distances between the users and the screen (Jakobsen et al., 2013), and collaborative work scenarios that require awareness of each others' actions (Isenberg et al., 2009). By *large vertical displays*, we refer to displays fixed in their vertical position and significantly larger than desktop displays. In our

(a) (b)

Figure 5.1: The two most common collaborative interactions. (a) One participant selects a view element via touch first. Then, the other participant indicates the annotation text via speech. (b) Two participants use the pens simultaneously to annotate.

work, we set out to explore alternative interaction modalities: touch, pen, speech, and mid-air gestures. We were specifically interested in multimodal interaction: interaction where these four types of input can be used in combination to perform certain actions. Combining interaction modalities can have various benefits, such as allowing to support user input from different distances, offering multiple degrees of freedom, and providing better support for particular tasks (Badam et al., 2016; Hinckley et al., 2010; Williams et al., 2020b). The best possible combinations of these interaction modalities for large vertical displays, however, are not immediately obvious: touch and pen interaction require standing close to the screen, and speech and mid-air gestural interaction are often not easily discoverable (Hincapié-Ramos et al., 2014). Also, we do not yet know how specific interaction modalities should be combined for different tasks and to support collaborative work. The preference for specific modality combinations and the order in which they are used may change depending on the modalities and tasks (Oviatt et al., 1997). As large vertical displays provide more interaction space, group work is an important scenario to consider. Combining multiple modalities and users leads to a more complex scenario. As such, many questions are still open in the space of multimodal interaction for large vertical displays. Here, we focus on the following two research questions:

RQ1 What interaction modalities are preferred for exploring data visually on a large vertical display?

RQ2 How can groups benefit from using multimodal interaction for collaborative data exploration?

In order to explore these questions, we conducted and analyzed an interaction elicitation study with 20 participants — in groups of two — in which we asked them to come up with interaction proposals for 15 tasks, giving the option of using touch, pen, speech, and mid-air gestures. With this methodology, we examine what end-users propose intuitively and assess the elicited interactions.

We found that people preferred unimodal interactions with either speech, touch, or pen to perform the exploration-focused tasks we gave them, which differs from previous findings about multimodality being preferable (Oviatt et al., 1997; Saktheeswaran et al., 2020). When acting with more than one modality, participants opted for using touch first and speech later. When collaborating, participants worked closely together and acted either simultaneously using the same modality or in a sequence with two different modalities. With our work, we provide design knowledge on user preferences for multimodal and collaborative interactions with large vertical displays. We contribute the elicited gesture set, the top proposals, the interaction patterns, and our analysis on what interaction modalities were chosen in specific scenarios.

5.2 Background: Elicitation Studies

Past work has proposed multimodal interactions based on intuition and related work (e.g., Badam et al. (2016) and Srinivasan et al. (2021)). Here, we tackle the subject with a different methodology called the *elicitation study*. The elicitation study methodology was first proposed by Wobbrock et al. (2009). It is an interaction design methodology in which end-users are presented with the effect of an action on a computing system and are asked to propose the action to trigger the effect. The effect of the interaction is known as the *referent* and the proposed command or gesture is known as a *symbol* (Wobbrock et al., 2009). After the elicitation, symbols are classified into clusters of *signs* based on their similarity (Vatavu and Wobbrock, 2022). Elicitation studies are mainly used to inform the design of interactions for a system (Tsandilas and Dragicevic, 2016). The main outcome of elicitation studies is the *consensus set* which is the set of interaction proposals that reached the highest agreement per referent (Villarreal-Narvaez et al., 2020). Usually, it is called consensus *gesture* set because standard elicitation studies tend to be about mid-air gestures. In this article, we refer to it as *consensus set* because our study involves multiple modalities. Recently, Villarreal-Narvaez et al. (2020) conducted a literature review on elicitation studies. Based on their findings, they suggest future studies explore other modalities besides mid-air gestures and elicit more than one symbol per referent, to investigate further the design space for interacting with smart environments.

The first elicitation study on multimodal interaction was conducted by Morris (2012), without data visualizations. She elicited voice and mid-air gesture commands for interacting with a web browser on a living room TV and found that gestures had more commonalities among participants than speech. The results suggest that specific modalities fit better for certain referents. Willett et al. (2014) conducted the first elicitation study for post-WIMP interaction with data visualizations. The researchers elicited multi-touch gestures for selection in four types of data charts. They found that participants strongly preferred simple, one-handed selection gestures, mainly using only one finger. According to Lee et al. (2021), an open research direction for post-WIMP interaction with data visualizations is exploring creative adaptations from broader human-computer interaction (HCI) research. As the authors suggested, we take this successful HCI method — the elicitation study — to investigate multimodal and collaborative interactions for data visualizations.

5.2.1 Benefits of Elicitation Studies

Researchers have cited multiple benefits of elicitation studies. Elicitation studies allow us to understand user proclivities and preferences for interactive technologies (Vatavu and Wobbrock, 2022). They serve not only to define a set of preferred interactions but also to characterize the diversity of the proposed interactions, aiming to understand better how people associate (or not) some types of interactions with specific tasks. Elicitation studies are considered a type of participatory design (Morris, 2012), as they allow end-users to get closely involved in the design process of interactive systems. Although designing with end-users may be more complex and time-consuming than the alternative, it leads to developing more usable and satisfying designs (Abras et al., 2004). User-defined gestures tend to be preferred and more memorable than gestures predefined by a professional designer (Wobbrock et al., 2005). In the experiment of Nacenta et al. (2013), participants considered the user-defined gestures less effortful and less time-consuming. In the field of HCI, more generally, the design of novel interactive systems is often based on elicitation studies (Nebeling et al., 2014), as eliciting interactions without the technical limitations of a gesture recognizer facilitates the exploration of the design space.

5.2.2 Challenges of Elicitation Studies

One of the main challenges of elicitation studies is legacy bias. This type of bias describes the tendency for users to propose commands they know from previous interaction experiences. Morris et al. (2014) recommend three techniques to reduce legacy bias: *production*, *priming*, and *partners*. We follow their recommendation by applying these three techniques in our study, as explained in Section 5.4.6. However, this bias is not always seen as a disadvantage (Williams and Ortega, 2021). Legacy interactions can be more discoverable and therefore lead to a consensus set that feels intuitive to the users.

Tsandilas and Dragicevic (2016) critiqued that the standard formulas used for agreement calculation in elicitation studies (e.g., Vatavu and Wobbrock, 2015; Wobbrock et al., 2005) do not consider *chance agreement*. Chance agreement is the likelihood that two or more participants propose the same type of interaction by chance. While Tsandilas (2018) proposed agreement indices that take chance agreement into account, these indices do not consider our scenario where participants make more than one proposal per referent. More recently, Vatavu and Wobbrock (2022) argued that chance agreement should not affect agreement but also focuses only on studies with single proposals.

Other challenges may arise when applying the findings of an elicitation study to a real-world system. For example, technical limitations may prevent the detection of the elicited interactions, or these interactions may conflict with other existing interactions in the system. Those are issues that the researchers do not necessarily encounter in the study, as the methodology does not consider implementation details. Instead, the results of the elicitation study are meant to serve as a basis for navigating the design space of interaction techniques for new systems. Moreover, having no technical limitations allows end-users to be more creative and, accordingly, to propose innovative ways of interaction.

5.3 Related Work

In this section, we present related work on interacting with large vertical displays, and collaborative data exploration.

5.3.1 Interaction Design for Large Vertical Displays

Working on large vertical displays has various benefits and challenges. While the display size and resolution facilitate sensemaking (Andrews et al., 2010) and collaborative work (Jakobsen and Hornbæk, 2014), the extreme viewing angles up close can impact perception accuracy for certain data encodings (Bezerianos and Isenberg, 2012), and users may have difficulty reaching some display areas (Riehmann et al., 2020). Consequently, researchers have investigated multiple ways of interacting with these displays: direct manipulation through touch (Riehmann et al., 2020) or pen (Guimbretière et al., 2001), gaze (Herholz et al., 2008), using mobile devices, such as smartwatches (Horak et al., 2018), tablets (Kister et al., 2017), and augmented reality displays (Reipschlager et al., 2021), through mid-air gestures (Nancel et al., 2011) and body movements (Badam et al., 2016). Other researchers, like Baudisch et al. (2001), have proposed to apply focus and context techniques to visualize information at different resolution levels without the need for additional actions. This diversity is reasonable as users often physically move in front of the screen (Ball et al., 2007): they tend to stand far from it to get an overview and move closer to access the details (Horak et al., 2018). As such, supporting interaction modalities that allow both close-up and distant interaction is crucial. We explore four possible modalities: the use of speech and mid-air gestures from afar as they have proved helpful in other contexts (e.g., Badam et al., 2016; Srinivasan et al., 2020) and do not require additional screens—and pen and touch for close-up interaction.

We are not the first to propose using touch and pen interaction for visualization. Lee et al. (2015) proposed leveraging touch and pen interaction for authoring and annotating visualizations. While touch and pen are often used interchangeably, touch is more pervasive thanks to the popularity of smartphones, but the pen is more precise, as it does not have the fat-finger problem (Drucker et al., 2013). Walny et al. (2012) found that although touch is preferred to move objects, both pen and touch were used for selecting menu items. Badam et al. (2016) proposed using mid-air gestures and proxemics for visual exploration with interactive lenses. They found that people preferred using proxemics for navigation and mid-air gestures for "direct" actions, such as terminating a lens composition. Pointing from a distance, resembling a laser pointer, was considered a mid-air gesture that involved extending the hand using a special glove. However, these findings come from comparing pairs of interaction modalities (e.g., pen with touch, mid-air gestures with proxemics). We extend their work by considering four modalities and their combinations.

Previous work has already proposed ways of interacting with the visualization techniques we included in the study. For example, Drucker et al. (2013) suggested sorting data items in a bar chart by dragging the finger along the corresponding axis, while Srinivasan et al. (2020) recommended using speech commands for filtering. Nevertheless, we wanted to discover whether the study participants would propose similar actions or go in a different direction, given that they were free to choose among and combine multiple modalities. We also included the symbol map for which there are no multimodal interaction proposals yet.

Badam et al. (2017) suggested mapping modalities to specific interaction techniques based on their affordances in immersive environments. Inspired by their work, we seek to identify the preferred modality combinations for visual exploration tasks. Srinivasan et al. (2021) proposed to interact with unit visualizations on a vertical display, mixing touch, pen, and speech interaction. The authors recommend using direct manipulation to interact with single items and natural language to interact with item groups. However, more recently, experts preferred the pen over touch and speech for exploring data on tablets (Molina León et al., 2022a). On a large display, we investigate user preferences about these three modalities combined with mid-air gestures, including more visualization techniques. While other modalities like gaze and proxemics are also worth investigating, we limit the scope of our research to four modalities, as it is already complex to consider them in combination with collaborative work.

5.3.2 Collaborative Data Exploration

Isenberg et al. (2011a) define *collaborative visualization* as the "shared use of computer-supported, interactive visualizations by more than one person to perform joint information processing activities." Collaboration can be co-located or distributed and synchronous or asynchronous. It can go from loosely coupled to closely coupled depending on how much information participants share and how much they interact with each other (Isenberg et al., 2010). Collaborative systems should support not only *taskwork* (actions to complete the task) but also *teamwork* (actions to complete the task as a group) (Lam et al., 2012). In this paper, we focus on the co-located scenario and study participant choices regarding timing, collaboration style, and the use of different interaction modalities for achieving taskwork and teamwork.

Interaction challenges on large displays during collaboration have been subject of research. While comparing horizontal and vertical displays, Rogers and Lindley (2004) found that vertical ones make it easier to show content to an audience. However, sharing devices is harder than on a horizontal display because it requires moving closer to the screen or a table to put the device down and give the opportunity to someone else to pick it up afterward. That may represent an added challenge for interactions that require an additional device, such as pen input. Based on their findings, Rogers and Lindley suggest providing the option of adding annotations and performing calculations directly on the vertical display to facilitate collaboration. The need for coordination (Prouzeau et al., 2017) and privacy (Brudy et al., 2014), for example, can impact interaction while sharing the screen space. In the experiment of Prouzeau et al. (2017), pairs consistently divided space while working on a wall-sized display, even if the task was not spatially divisible. In the study of Isenberg et al. (2009), participants solved interaction conflicts (e.g., two users trying to drag the same element) by talking or establishing rules. Moreover, participants asked for dedicated features to help group members be aware of what the others were doing. Adding annotations is a helpful awareness feature that we included in our study. Dostal et al. (2014) proposed measuring user attention via gaze tracking to adapt the visualizations according to the status of each collaborator.

When working next to each other, participants can closely collaborate through cooperative gestures, i.e., gestures by multiple users that contribute to a single joint command (Morris et al., 2006). Liu et al. (2017) found that these gestures can reduce the physical effort required to manipulate data items on a large display. In the tabletop system Cambiera, Isenberg and Fisher (2009) proposed an interaction technique called *collaborative brushing and linking* that helped each user to be aware of the interactions of others in their personal views of the data. We elicit collaborative interactions to learn when end-users favor multi-user interactions and how they coordinate their work when interacting multimodally.

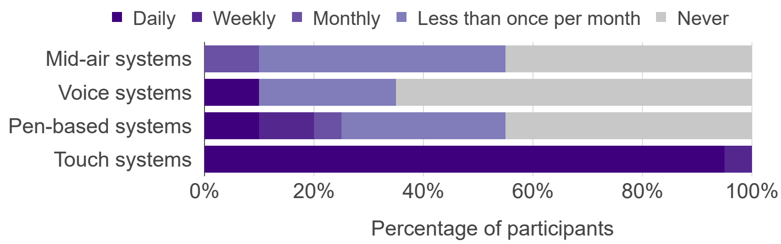


Figure 5.2: Participants' reported prior experience with each of the tested interaction modalities, expressed by frequency of use.

5.4 Study Design

To conduct our elicitation study, we recruited researchers who explore data in their everyday work life. We asked them to participate in pairs and to brainstorm together about diverse ways to interact with data visualizations through touch, speech, pen, and/or mid-air gestures. While they could discuss all proposals together, the final proposals were individual and did not need to overlap. Working together and making multiple proposals per referent were two strategies to combat legacy bias (Morris et al., 2014) (see details in Sect. 5.4.6). We recruited participants who already worked together to add ecological validity to our study (Morris, 2012).

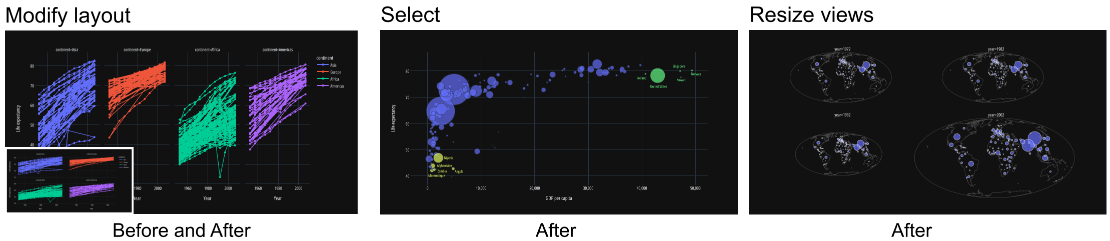
We conducted a pilot study with an additional pair of experts (P1 and P2) that helped to adjust the prompts and the minimum number of proposals required. Then, we proceeded to conduct the main study with 10 pairs of participants. We followed common practice in elicitation studies and recruited 20 participants (Villarreal-Narvaez et al., 2020). The study took around 90 minutes for each pair.

5.4.1 Apparatus

We conducted the study with an 86-inch Promethean ActivePanel display of 4K resolution in a meeting room of 28 m^2. We created the visualizations in Python with the Plotly Express library (Plotly, 2022). The experiment was video recorded with the informed consent of the participants.

5.4.2 Participants

In total, we recruited 20 researchers and research assistants (eight female, aged 20–52) through university mailing lists. They had already worked with spatio-temporal



(a) (b)

Figure 5.3: (a) Prompts showing the data visualizations before and after the interaction for the *Modify layout* referent. (b) After prompts of the *Select* and *Resize views* referents.

data in diverse scientific domains, mainly in the social sciences. The main domains of expertise were political science and geography. There were 11 doctoral students, six postdoctoral researchers, two bachelor students, and one professor.

All participants reported that they interacted with data visualizations and explored spatio-temporal data as part of their job. Twelve of 20 participants worked with visualizations at least once per week. All pairs of experts were either working together or had collaborated in the past.

When asked about how frequently they had interacted with the proposed modalities, participants had most experience with touch and least experience with speech, as shown in Figure 5.2. While everyone had experience with touch interaction, 45% had never interacted with mid-air gestures, another 45% had never used a pen as an input device, and 65% had never interacted with speech commands. When asked about their experience interacting with large vertical displays, 75% of the participants reported to have worked with them before. Everyone was right-handed except for two ambidextrous people.

5.4.3 Dataset

The data we used to create the visualizations was a set of development indicators published by the Gapminder Foundation (2022). They included the life expectancy, GDP per capita, and population of 142 countries from 1952 to 2007. Given that the experts worked in different fields within scientific research, we decided to use a real-world dataset that everyone would understand to ensure ecological validity.

5.4.4 Referents

The referents in our study were low-level data interaction tasks. We chose low-level tasks because every more complex exploration task is composed of these low-level tasks and requires combinations of interactions to be completed. Specifically, we examine 15 low-level tasks relevant to the exploration of spatio-temporal data inspired by the typology of Andrienko and Andrienko (2006). We focused on tasks relevant for working collaboratively with a large vertical display based on the interaction taxonomy of Yi et al. (2007), the task typology of Brehmer and Munzner (2013), and the interactions with multiple coordinated views investigated by Langner et al. (2019). Six of our 15 referents were associated with managing multiple coordinated views (MCV): activating and deactivating brushing & linking (B & L), merging views, splitting a view, resizing views, and rearranging views. We added multi-selection and annotation authoring as referents to support visual awareness across users (Isenberg and Fisher, 2009). Moreover, we differentiate between data-centric and view-driven filtering based on the different interactions proposed by Sadana and Stasko (2014).

We present the final list of referents and their classification according to related work in Table 5.1. We excluded zooming and panning from the list because previous work has consistently found successful interactions to perform them using touch (Sadana and Stasko, 2016), mid-air gestures (Badam et al., 2016), and proxemics (Jakobsen et al., 2013). The referent images we used in the experiment can be found in the supplemental material. The supplemental material is publicly available on OSF at `https://osf.io/m8zuh`.

Table 5.1: Our 15 Referents and Their Classification. Citations Refer to the Corresponding Task Typologies and Taxonomies.

Referent	Task type
Show item details	Direct lookup (Andrienko and Andrienko, 2006), Abstract/Elaborate (Yi et al., 2007)
Select	Inverse lookup (Andrienko and Andrienko, 2006), Select (Brehmer and Munzner, 2013; Yi et al., 2007)
Deselect	Select (Brehmer and Munzner, 2013; Yi et al., 2007)
Activate B & L	Select (Brehmer and Munzner, 2013), Connect (Yi et al., 2007)
Deactivate B & L	Select (Brehmer and Munzner, 2013), Connect (Yi et al., 2007)
Data-centric filter	Filter (Yi et al., 2007), Behavior characterization (Andrienko and Andrienko, 2006)
View-driven filter	Filter (Yi et al., 2007), Pattern search (Andrienko and Andrienko, 2006)
Sort	Reconfigure (Yi et al., 2007), Arrange (Brehmer and Munzner, 2013)
Change encoding	Encode (Brehmer and Munzner, 2013; Yi et al., 2007)
Merge views	Aggregate (Brehmer and Munzner, 2013), Direct comparison (Andrienko and Andrienko, 2006)
Split view	Aggregate (Brehmer and Munzner, 2013), Direct comparison (Andrienko and Andrienko, 2006)
Resize views	Abstract/Elaborate (Yi et al., 2007)
Modify layout	Reconfigure (Yi et al., 2007), Arrange (Brehmer and Munzner, 2013)
Show regression line	Connection discovery (Andrienko and Andrienko, 2006)
Adding annotation	Annotate (Brehmer and Munzner, 2013), Reconfigure graphics (Badam et al., 2017)

5.4.5 Visualization Techniques

We chose the visualization techniques based on the recommendations of Andrienko et al. (2003a) for exploring spatio-temporal data. Accordingly, we included the following techniques: line charts, bar charts, scatterplots, bubble charts, and symbol maps. We selected these charts for the variety of visual channels they use to encode data. Additionally, eleven referents involved multiple views to display spatial entities or temporal steps via a *small multiples* technique.

5.4.6 Study Procedure

The elicitation study was composed of four parts. First, we explained its structure to the participants and asked for their informed consent to record their interaction proposals through video and audio. Then, they filled out a demographics questionnaire that included questions about their previous experience with each of the interaction modalities.

To start the elicitation, we asked participants to picture themselves in a scenario where they wished to explore a new dataset together, and they needed to perform a series of actions as part of the exploration process. As our goal was to investigate how multimodal interaction can benefit group work (**RQ2**), it was important that the participants would see themselves as a team. We asked the experts to propose individually at least three interactions for each of the 15 referents, including at least one collaborative proposal and at least one that was multimodal. One of the three proposals being both collaborative and multimodal was also sufficient. We presented each referent graphically through a pair of images showing the visualization before and after the interaction (see examples in Figure 5.3). We considered using animated prompts that would show transitions but decided against them to avoid biasing the participants by implicitly suggesting specific ways of interaction (e.g., resizing a view could start by dragging one corner if the animation showed the view getting enlarged in a specific direction first). For each referent, the experimenter read a question out loud presented above the pair of images of the form *How would you...?*, such as "How would you merge two views into one?", before switching to the first image on full screen to start eliciting. We show how the elicitation took place in the supplemental video.

Participants were free to come up with any proposal that they felt was best suited without restricting themselves to any set of "allowed" interactions. There was no time limit. We set the order of the tasks according to how they complemented each other, e.g., deselect after select. We allowed participants to use any of the four interaction modalities: touch, pen, speech, and mid-air gestures. As we aimed to investigate what modalities were preferred (**RQ1**), participants could propose using any modality alone or combined with others. Participants were also free to add interface elements to the screen if they wished to have them, so we could observe and analyze what they preferred. For each referent, we encouraged participants to consult with each other and to show their ideas by performing the corresponding actions. Each participant had to make their own proposals that could be similar to or different than those of their partner. For each referent, each person described their final proposals on paper and picked a favorite among them. Most participants started each task from a table around three meters away from the display where they had written down their proposals for the previous task. The experimenter then asked the participants to move back towards the display for each task but did not prescribe a specific starting distance to take on. During the study, participants could stand where they wanted and relocate freely.

We applied the Wizard of Oz technique for changing between the referent images (the *before* and *after* images) when participants made an interaction proposal to demonstrate the effect of the interaction (Perera et al., 2021). We made clear that the technical interaction recognition of the system would hypothetically work perfectly. This was done to avoid that participants would not propose interaction techniques out of fear that they might not be technically realizable. The study concluded with a short questionnaire asking participants to rate the perceived effectiveness of each modality with a five-point Likert scale, as in the study of Morris (2012).

To reduce legacy bias, we applied the *priming*, *production*, and *partners* techniques, as recommended by Morris et al. (2014). After answering the demographics questions, we primed participants by asking them to report three life situations where they had behaved creatively in the past, as suggested by Sassenberg and Moskowitz (2005) and successfully tested in previous elicitation studies (Ali et al., 2021). During the elicitation, we applied *production* by asking participants to produce at least three proposals for each referent (Williams et al., 2020a). Moreover, we asked them to make at least one multimodal proposal and at least one collaborative proposal, given the small number of multimodal interactions elicited in the study of Morris (2012). We applied the *partners* technique by inviting the experts to brainstorm and interact in pairs. We asked them to come with someone they already knew or worked with.

5.4.7 Data Analysis

We first extracted the interaction proposals from the list that each participant wrote down during the experiment. Then, we completed or corrected the details of each proposal based on the video recordings that were analyzed by two researchers separately. For each proposal, we documented the referent, the participant, the sequence of steps and their modality, whether it was performed by one person or two (if two, whether the steps happened in parallel), and whether it was a favorite. Afterward, we grouped the interaction proposals into *signs* based on their similarity according to the modality, the data attributes, and the target involved.

For analyzing the signs, we calculated the metrics *max-consensus* and *consensus-distinct ratio* proposed by Morris (2012) for each referent, and we report the consensus set based on the most popular proposal per referent, according to frequency. The max-consensus indicates the percentage of participants that proposed the most common interaction for a given referent. The max-consensus is 100% if all participants recommended the most common proposal. The consensus-distinct ratio indicates the proportion of distinct interactions proposed by a minimum number of participants (the *consensus threshold*). The consensus-distinct ratio is 1.0 when every interaction proposed for the referent is over the threshold. Although most researchers calculate an agreement score or rate for elicited interactions (Villarreal-Narvaez et al., 2020), our study included multiple proposals per participant and was conducted in pairs. Thus, it required different measures (Vatavu and Wobbrock, 2022). Vatavu (2019) proposed other metrics for this type of studies, such as consensus and growth rate, but said calculations are based on spatio-temporal coordinates of body gestures and do not consider multiple interaction modalities.

5.5 Results

We elicited a total of 1015 interaction proposals for the 15 referents. Each proposal included at least one step (action) completed with one modality. The multimodal and collaborative proposals included at least two steps taking place either sequentially or simultaneously. We present a summary of the proposals per referent and participant in Table 5.2. In elicitation studies involving speech input, the speech pro-

Table 5.2: Descriptive Statistics of the Proposals per Referent and Participant.

Metric	All	Multimodal (M)	Collaborative (C)	M & C
Min	3.00	1.00	1.00	0.00
Median	3.00	1.00	1.00	1.00
Mean	3.38	1.41	1.08	0.85
Std	0.61	0.60	0.28	0.43
Max	6.00	4.00	3.00	2.00

posals whose text overlaps with the referent name are sometimes excluded because the participants tend to use those words first. We did not remove those commands, given that, in referents like *sort*, ignoring commands using that verb would radically limit the possibilities of appropriate terms (Williams et al., 2020b). Each participant successfully produced at least three interaction proposals per referent, with one person (P9) even proposing six ways for selecting. Participants made more multimodal proposals than required (424 instead of 300), leading to 42% of the elicited interactions being multimodal. They also proposed slightly more collaborative interactions than required (324 instead of 300), resulting in 32% of the interactions being collaborative.

After grouping the proposals into clusters based on their similarity (e.g., grouping speech commands such as "Deselect group 1" and "Deselect yellow group"), we identified 360 *distinct* interactions or *signs* among the 1015 proposals. Of the 360 signs, 215 were multimodal, and 161 were collaborative. More specifically, 127 proposals were both multimodal and collaborative (35.28%), 111 were unimodal performed by a single person (30.83%), 88 were multimodal and performed by a single person (24.44%), and 34 were collaborative and unimodal (9.44%).

Table 5.3: Consensus Set and Metrics per Referent for the Four Most Common Modality Combinations. MC Stands for Max-Consensus and CDR Stands for Consensus-Distinct Ratio. The Highest Value for Each Metric per Modality Combination Is in a Blue Cell.

		All		Speech only		Touch only		Touch-Speech		Pen only	
Referent	Most Common Interaction	MC	CDR	MC	CDR	MC	CDR	MC	CDR	MC	CDR
Show details	Tap on mark	80.0%	0.45	55.6%	0.67	94.1%	0.50	44.4%	0.75	75.0%	0.50
Select	Lasso around marks	40.0%	0.29	35.7%	0.29	50.0%	0.60	100.0%	0.50	66.7%	0.67
Deselect	“Deselect group 1”	70.0%	0.11	77.8%	0.25	15.4%	0.00	28.6%	0.00	42.9%	0.33
Activate B & L	“Extend yellow to all years”	60.0%	0.11	100.0%	1.00	30.0%	0.17	37.5%	0.17	50.0%	0.00
Deactivate B & L	“Deselect group 1 except on 1992”	55.0%	0.09	68.8%	0.67	22.2%	0.00	25.0%	0.14	100.0%	0.00
Data-centric filter	“Show only Asia”	75.0%	0.20	78.9%	0.67	87.5%	0.33	33.3%	0.29	-	-
View-driven filter	“Show me the outliers”	50.0%	0.17	62.5%	0.25	57.1%	1.00	22.2%	0.00	66.7%	0.25
Sort	“Sort by population”	60.0%	0.38	63.2%	1.00	45.5%	0.50	71.4%	0.33	100.0%	0.00
Change encoding	“Set population size as point size”	65.0%	0.13	81.2%	0.25	15.4%	0.00	36.4%	0.29	50.0%	0.00
Merge views	“Merge graphs”	75.0%	0.27	78.9%	0.67	57.1%	1.00	50.0%	0.00	100.0%	0.00
Split view	“Split by country”	55.0%	0.18	68.8%	1.00	53.8%	0.25	100.0%	0.00	-	-
Resize views	Pinch on top of the view	50.0%	0.19	69.2%	0.20	66.7%	0.50	40.0%	0.00	100.0%	0.00
Modify layout	Drag a view	60.0%	0.18	68.8%	0.20	80.0%	0.67	-	-	-	-
Show regression line	“Add regression line”	95.0%	0.21	95.0%	0.33	55.6%	0.20	38.5%	0.25	100.0%	1.00
Add annotation	Write text with the pen	65.0%	0.35	50.0%	0.25	100.0%	0.00	60.0%	0.67	81.2%	0.50

5.5.1 The Consensus Set Is Unimodal and Personal

On average, participants proposed 27 distinct interactions per referent. We present the *consensus set* of our study in Table 5.3, together with the metrics for the four most common modality combinations among the top proposals. As the standard agreement scores for elicitation studies do not consider the case of multiple proposals per referent (Vatavu and Wobbrock, 2022), we calculated the top proposals based on their frequency and present the agreement metrics *max-consensus* and *consensus-distinct ratio* proposed by Morris (2012) for this case. As mentioned in Sect. 5.4.7, the max-consensus indicates the percentage of participants that proposed the most common interaction for a given referent. A high consensus suggests that the interaction was considered the most intuitive for the task. Consistently, participants agreed most on interactions involving only one modality (speech, touch, and pen) and performed by a single person. Despite previous evidence suggesting that multimodal interaction may lead to a more fluid experience (Srinivasan et al., 2020), participants preferred to explore the data with simple unimodal interactions. Yet, a system would require to support speech, touch, and pen input to include the most commonly proposed interactions. Accordingly, it has to enable both distant interaction and direct manipulation. Only speech was common for interacting from a distance and should likely be supported at a minimum, together with touch or pen for close interaction. Participants proposed more mid-air gestures than pen interactions, but only pen interactions made it to the consensus set.

For 10 of the 15 referents, participants preferred speech interaction. Overall, we noticed that when given the freedom to propose any interaction with any of the four modalities, participants often came up with a speech command as the first proposal. Accordingly, 591 of the 1015 interactions proposed (58.23%) included speech interaction. Of them, 243 (23.94% of all) were speech commands only. During the study, several participants commented that using speech felt like the easiest option. In contrast, the most common proposals for the referents *show details*, *resize views*, and *modify layout* were standard touch gestures, while the most common proposals for the referents *select* and *add annotation* were with pen interaction. For these five referents, depending on the task, about 20-60% (39% on average) of the first proposals made by the participants became part of the consensus set. Participants preferred direct manipulation for lookup tasks (Andrienko and Andrienko, 2006) and for modifying the position and size of the views. Speech interaction was instead favored for more abstract tasks aimed to find patterns across sets of data items, such as regression, and for synoptic tasks (Andrienko and Andrienko, 2006) related to comparing the sets across views. We did not find any evidence of a relationship between the visualization techniques we used and the interaction modalities of the consensus set.

On average, participants tended to agree most on speech interactions. Speech interactions had a mean max-consensus of 70%, mid-air gestures of 63%, pen interactions of 62%, and touch interactions had a mean max-consensus of 55%. While pen and mid-air interactions had a max consensus higher than touch, participants made no proposals at all using those modalities for three (data-centric filter, split view, and modify layout) and two referents (view-driven filter and add annotation), respectively. Multimodal proposals starting with a touch gesture, followed by a speech command, had a mean max-consensus of 46%. Touch-speech interactions were preferred over interactions using pen-only and mid-air gestures among the top proposals. We present more details about the multimodal proposals in Sect. 5.5.2. Overall, the proposal with the highest max-consensus (95%) was a speech command to apply a regression model.

Now we look at the consensus-distinct ratio, which gives a sense of the spread of the agreement. Unlike Morris (2012), we used a *consensus threshold* of three instead of two because we elicited many more interactions due to applying the *production* principle (Morris et al., 2014), and pairs often agreed on their proposals after brainstorming, so reaching a consensus between two people was common. So the consensus-distinct ratio is 1.0 when every interaction proposed for that referent was proposed by at least three participants. On average, the interactions that made it to the consensus set were proposed by at least half the pairs in agreement (i.e. per referent, 5.6 groups proposed the same interaction twice), in contrast to each person proposing something different than their partner. The referent with the highest ratio was *show details* suggesting that most interaction proposals for invoking a tooltip reached a high agreement among the participants. The lowest ratio was 0.09 for *deactivate brushing & linking (B & L)* which suggests a higher diversity of proposals overall, with less agreement.

The difference between the metrics across the top four modality combinations suggests that participants mapped some referents to specific modalities. For example, the referent *activate B & L* reached the highest consensus among speech proposals, in contrast to a low one with other modalities. Some referents like *data-centric* *filter* and *show regression line* reached a high consensus with touch and pen, respectively. Still, the popularity of speech commands overall determined the top proposal of those referents. Accordingly, the second top proposals for *data-centric filter* and *show regression line* were touch-only and pen-only, respectively (see the list of top three proposals in the supplemental material). Moreover, although touch and pen both serve for direct manipulation, participants favored touch interaction for *resize views* and *modify view layout*.

Multimodal Synonyms

When looking beyond the consensus set, we find more diversity regarding modalities and collaboration among the top proposals per referent (see the list of top three proposals). We detected what Morris (2012) calls *multimodal synonyms* among the top proposals of the referents *show details*, *select*, and *activate B & L*. Multimodal synonyms are equivalent interactions with different modalities that participants propose as alternatives for the same command. For example, participants proposed to perform lasso selection with either the pen or touch. For brushing & linking, they wished to drag and drop a selected group of items via touch or a mid-air gesture. During the experiment, several participants commented that having modalities to choose from made the system more accessible and allowed them to select a modality depending on the situation.

Favorites Matched Consensus

We asked participants to propose at least three interactions for each referent and to select a favorite among them. We were interested in finding out what interactions participants considered best, given that the most common proposal in an elicitation study may not necessarily be considered the most appropriate in practice. However, when comparing the favorites with the consensus set, the most common interaction was also the most commonly named favorite interaction for all referents. The only exceptions were the multiple top favorite interactions for the referents *select* and *merge views*. For *select*, participants had four favorite proposals besides the most common one. The first was the collaborative version of the lasso selection with pen, the second was a query via speech command, the third was the collaborative version of the second, and the fourth was a multimodal and collaborative interaction with speech and touch. To *merge views*, participants favored dragging one view towards

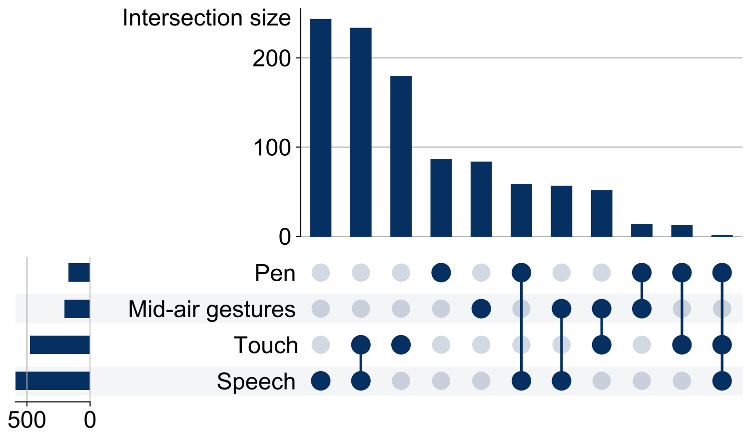


Figure 5.4: UpSet plot showing the frequency of the different combinations of the four interaction modalities across the 1015 proposals.

the other besides using a speech command. Although multimodal and collaborative interactions did not make it to the consensus set, we examine them in the following sections to investigate in which situations and how these interactions can support data exploration on a large vertical display.

5.5.2 Multimodal Interaction: Mainly Touch and Speech

We show the distribution of the modality combinations among the 1015 proposals in Figure 5.4. Speech and touch interaction were the most used interaction modalities, with the total of touch-only proposals even surpassing all the pen interactions and mid-air gestures combined. However, the second largest group of interactions was multimodal, combining touch and speech (23% of all proposals). Such multimodal commands also appeared in nine top proposals.

Participants tended to divide tasks into multiple steps and associate each step with a different modality. 61% of the proposals consisted of a touch gesture followed by a speech command. Using touch followed by speech was most suggested for *add annotation*, *show regression line*, and *change encoding*. Participants used touch first for choosing a view or data items of interest. Then, they specified an action to apply to them orally. For example, someone first tapped on a view element to select it and then added an annotation via voice, or they selected a view with touch first and then asked for the calculation of a regression model via speech. In the inverse

(a) (b)

Figure 5.5: Collaborative synonyms proposed for merging and splitting views. (a) Two participants perform two gestures to merge the views together. (b) One participant performs the gesture alone to split the view back into two.

order (39%), participants first used a speech command to select data items, activate a mode, or invoke a menu. Then, they performed the main action with touch. For example, using speech to activate a rearranging mode and then approaching the screen to drag and drop multiple views until reaching the desired view layout. Using speech followed by touch was most common among the proposals for *resize views*, *modify layout*, and *select*.

Nineteen of 20 participants chose at least one multimodal interaction as their favorite. The single participant (P21) who did not choose any multimodal interaction as his favorite said, during the experiment, that he preferred the simplest interactions. In contrast, his partner (P22), always chose a multimodal interaction as his favorite. Among the favorite multimodal interactions, the most common combination was a touch gesture followed by a speech command (26%) and a speech command followed by a touch gesture (21%) in line with their frequent occurrence among the multimodal proposals.

5.5.3 Collaborative Interactions

During data analysis, collaboration styles can range from being closely coupled to loosely coupled (Isenberg et al., 2010). Most of the collaborative proposals we elicited belong to the closely coupled case as they involve two persons working together to perform a low-level task in a co-located scenario. There were a few exceptions showing loosely coupled collaboration, where the persons stood next to each other but worked in parallel interacting with different data items.

The most common collaborative proposals were two distinct interactions for the *add annotation* referent, shown in Figure 5.1. These interactions demonstrate the patterns we found: the first consisted of a sequence where one person started by tapping on a bar inside a bar chart, and then, their partner used a speech command to attach an annotation to the bar. The second interaction involved two users writing different annotations simultaneously with the pen.

Collaborative Work Was either Unimodal and Simultaneous or Multimodal in Sequence

We distinguish between two interaction types in collaborative proposals: *sequential* and *simultaneous* interactions. In the sequential case, one person performed the first step of the interaction, and their partner waited for that step to be over before proceeding to execute the next one. Although each step may have targeted different objects on screen, their actions were part of a single joint command. In the *simultaneous* case, both persons interacted simultaneously to perform two steps in parallel without conflict. Of the 161 distinct collaborative interactions, 90% were sequential, and 10% were simultaneous.

The sequential and simultaneous types of collaboration were often paired with specific modality combinations. When two participants interacted simultaneously, they mostly interacted using the same modality (63%). For example, to resize views, the third most common proposal was that both users drag a view border to adjust the size. Others performed mid-air gestures in synchrony to merge two views into one, as shown in the supplemental video.

In the sequential case, partners mostly interacted multimodally (83%). Each person became responsible for one modality. Overall, we identified three types of multimodal sequences in collaborative interactions. They demonstrated that groups often performed their actions at different distances: one person stayed close to the screen for direct manipulation, and the partner stood farther away and used speech or mid-air gestures. The most preferred form of collaboration consisted of a two-step sequence where one person performed a touch gesture, and their partner interacted via voice afterward. Such interactions were proposed for all referents but happened most often for compound tasks (Buxton, 1995). Filtering is such a task that may seem to be a single entity, but in reality, it can be divided into two sub-tasks: choosing (or selecting) a set of items and then subtracting items based on the selection (as it works on Tableau YouTube Channel, 2018). Using touch followed by speech was most proposed for calculating a regression model, adding annotations, and data-centric filtering. The most common proposal of this kind was the tap-and-speech sequence to add an annotation, shown in Figure 5.1a.

The second most common combination of multimodal sequences used speech, followed by touch. It was proposed most for tasks associated with managing multiple views. A person looking at the screen from afar spoke to select an element or activate a mode (e.g., modification mode). Then, the partner completed the joint action by dragging view elements with touch. The third combination involved touch and mid-air gestures. A person started the interaction by tapping or tap-and-holding to select a view element. Then, the partner, who stood further away from the screen, air-dragged other elements or performed a dedicated gesture, e.g., extending arms to split a view into two.

Collaborative Synonyms

When comparing interactions of single users and pairs, we identified what we call *collaborative synonyms*. Among the distinct interactions, we often found proposals that were identical except for being done by a single participant or by two people. Sixty-six interactions followed this pattern: 33 described a sequence of steps performed by a single person, and for each one, an equivalent existed, performed by two people. We show two collaborative synonyms illustrating that in Figure 5.5. We found at least one example of this case for every referent, mainly for tasks associated with selection and using the B & L technique. There are two pairs of collaborative synonyms among the top proposals: one to *change encoding* and one to *add annotation*. Most synonyms were multimodal, and the collaborative version meant that the second person would introduce the second modality.

5.5.4 Perceived Effectiveness

After the elicitation, we asked participants to rate each interaction modality according to how they perceived its effectiveness for the given scenario, as Morris (2012)

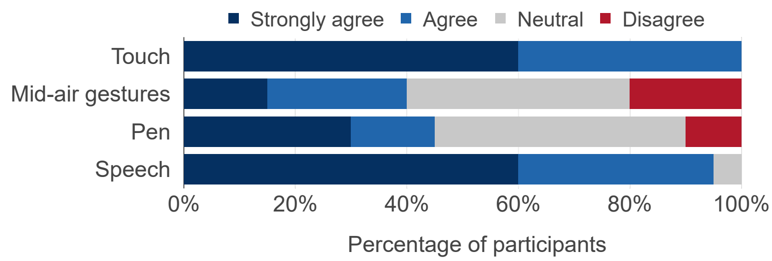


Figure 5.6: Ratings of how participants perceived each modality as an effective way to interact with the visualizations on a large vertical display.

did. Figure 5.6 shows these ratings that reflect the overall assessment of the participants about each of the four modalities for exploring data visually on a large vertical display. Participants considered touch the most effective way to interact with the data visualizations, followed closely by speech. The pen and mid-air gestures were rated neutral by 45% and 40% of the participants respectively. The fact that speech was rated second most effective is surprising, given that 65% of the participants initially reported that they had never used speech interaction before. Although people had more experience with the pen and mid-air gestures, they still perceived speech as more effective. Nevertheless, these ratings of perceived effectiveness roughly fit the appearances of each modality in the top proposals per referent. However, when interpreting these results, it is necessary to bear in mind that the assessment of the interaction modalities comes from the experience of the participants with the 15 referents we chose according to the scenario (see Sect. 5.4.4). Therefore, these findings might be expanded with the study of other referents and scenarios.

5.6 Discussion

In this section, we discuss and interpret our findings.

5.6.1 Is Touch and Speech All We Need?

Based on previous work, we expected participants to associate the exploration tasks with specific interaction modalities. In our study, they chose to focus on speech and touch interactions. Similar to the findings of Mignot et al. (1993), our results suggest that people prefer using speech commands for tasks that have no or only a loose connection to specific screen coordinates (i.e. a location on the screen). The preference of speech to filter also matches the recommendations of Badam et al. (2017). However, they propose using touch interaction to select data items, and in contrast, our participants deemed the pen most suitable for selection and speech for deselection. Thus, we hypothesize that participants considered deselecting different than selecting because it did not require looking for the marks on the display.

Touch interaction was one of the two interaction modalities we offered for close interaction. Although large vertical displays provide a wide surface to interact with, participants sometimes wished to interact with the pen instead of touch. For example, they proposed to select with the pen instead of the finger due to the higher precision. The pen was also a clear favorite for adding annotations. Therefore, both touch and pen might have their place for large display interaction, especially when precision becomes critical. We also find the dominance of speech commands for 10 referents intriguing. In the demographics questionnaire, 65% of the participants indicated that they had never used speech interaction, although most web browsers and smartphones recognize speech commands nowadays. So, how come the scientists preferred speech in the study but had rarely used it before? A potential explanation is that the lack of technical limitations during the elicitation gave participants confidence to brainstorm speech commands. Another reason might be that the prevalence of physical navigation in front of large vertical displays gives priority to speech as a natural way to interact from a distance. Thus, speech interaction becomes more relevant as the display size increases. Compared to mid-air gestures, expressing more complex commands was easier — more so when they did not involve the definition of a spatial component in the data (e.g., adding a regression line). Also, some mid-air gesture proposals required standing closer to the screen (e.g., air-dragging a view). Among the top interactions, the most popular mid-air gestures allowed the user to stand further away. Therefore, mid-air gestures were actually proposed both for interacting from afar and at a close range, depending on the characteristics of the specific gesture, but they were still the least proposed among the four modalities. At the end of the study, one participant indicated that mid-air gestures were best combined with another modality to add precision. For example, one proposal involved pointing to a mark from a distance, such as with a laser pointer, and specifying the data attributes to show via voice. The use of voice commands to provide more details suggests that participants felt they could be more specific with speech. Future work should look into more referents to better understand user preferences when having the possibility to interact through multiple modalities with data visualizations. Investigating diverse tasks is necessary as some may lead to clear tendencies (e.g., participants will most likely prefer using the pen for high-precision tasks).

5.6.2 Should We Offer Multimodal Interaction?

The consensus set of our study is unimodal despite previous evidence of user preferences for multimodal interaction (Saktheeswaran et al., 2020). However, as Oviatt (1999) points out, having the possibility to interact multimodally does not mean that users will take it. Although 42% of the elicited interactions were multimodal, the unimodal alternatives had the highest frequency. Participants appreciated the expressiveness of multimodal interaction but favored the simplicity of single modalities for low-level tasks. Moreover, multimodal interaction is still rare in industry products, and legacy bias may have influenced the preference for single modalities.

Nevertheless, multimodal interactions were among the top proposals for nine referents. Those multimodal interactions were mainly sequences of direct manipulation (touch or pen) followed by an action at a distance (speech or mid-air gesture). That suggests that given the freedom to stand at any distance from the screen, participants preferred to combine modalities that would work at different distances, especially when collaborating. For eight of the nine referents with top multimodal proposals, the preferred proposal that surpassed the multimodal proposal in popularity was a speech command. That suggests that participants opted to express the whole task through speech instead of dividing it into two steps. In a real-world scenario where speech recognition errors are common (Saktheeswaran et al., 2020), multimodal interaction may be more reliable and precise than a speech command. For example, for view-driven filtering, defining a query orally to define a group of data items that the user noticed visually may be more challenging than tapping on the data items and then applying the filter with speech. Thus, supporting multimodal interactions would make the visualization system more robust. Future work should study what factors may influence the choice of the participants to combine specific modalities, such as physical movement and interaction costs.

5.6.3 Should Cooperative Input Be Supported?

Collaborative interactions were the smallest group among the top proposals. For referents like *sort*, two participants expressed that the task was too simple to interact collaboratively. For referents like selecting and splitting views, participants appreciated working collaboratively. That fits the finding of Morris et al. (2006) about cooperative gestures not being performed too often to avoid interrupting their partner. Participants favored collaboration when there were two item groups or two views to interact with. When we examine the collaborative interactions that made it to the top proposals, we mostly find combinations of a modality suitable for direct manipulation and a modality for distant interaction. Those combinations correspond to the findings of Hinrichs and Carpendale (2011) on interaction with tabletops. They found that the actions performed by multiple users were strongly influenced by the social context. When our participants proposed collaborative interactions combining close and distant interaction, they were often already in position: one person was standing close to the display and the other further away. Proposing such multimodal sequences may therefore be a direct consequence of their placement. That suggests that participants may divide not only the screen space between them (Prouzeau et al., 2017) but also the larger area in front of the display. As we did not ask participants to start the task at a specific distance to the screen, future work should analyze how users position themselves in the 3D space in front of the screen, with the help of a motion tracking system (e.g., Dostal et al., 2014), to investigate how the initial position and movement may influence their choices. The visualization design choices should also be considered to investigate whether they influence the interaction distance. Identifying the reasons why people choose to interact collaboratively, taking interaction cost and engagement into account, is also an interesting research question for future work.

5.6.4 Are Elicitation Studies Helpful for Designing Interactive Data Visualizations?

One of the main goals of an elicitation study is to define a consensus set. In ours, there was no conflict among interactions, i.e. participants did not map the same interaction to two different referents. Thus, we could implement a system that would have no problem distinguishing between tasks. That lack of conflict was potentially due to the dominance of speech and natural language being more expressive than other modalities. Implementing the consensus set would require speech recognition combined with touch and pen input. Given that the touch and pen interactions were standard actions (e.g., tap, drag and drop, draw a line), the main technical challenge would be having reliable speech recognition. If the implemented version struggled with speech recognition errors as in previous work (Saktheeswaran et al., 2020), the extended list of top proposals provides alternatives to speech commands. In this respect, the elicitation was a successful methodology for us. In future work, conducting a study with a system enabling the consensus set is necessary to assess how effectively the elicited interactions can support data exploration on a vertical display when put together. The system should offer multiple interaction options per referent (including *multimodal synonyms*) so that participants can choose among different modalities or modality combinations to perform a task. Including *collaborative synonyms* would also give participants the opportunity to choose between personal and collaborative work. Testing the system in different contexts (e.g., meeting room, public space) would help to assess how the circumstances may influence the participant choices.

Participants consistently associated complementary referents with the same interaction modality. For example, activating and deactivating the *brushing & linking* technique were both preferred via speech. *Select* and *deselect* were the only exception. While the pen prevailed for selection, speech interaction was preferred to deselect, with a high max-consensus. The study results gave us insights into how users perceived the tasks and how the interaction techniques could be designed based on the groups by modality and directness we found.

In the study design, we applied all recommended techniques to avoid legacy bias. The dominance of speech commands despite the lack of experience of the participants with it suggests that we mitigated the bias successfully. However, three standard touch gestures made it to the consensus set which might not be problematic as the support of standard operations will be expected by future users of an interactive large display system. Participants often started speech commands with phrases like “Hey Siri”, suggesting an influence of their knowledge about voice assistants but also pointing out that they wished for speech input to be given explicitly rather than having their speech analyzed throughout their work in front of the display. Although Morris et al. (2014) introduced the three principles for reducing legacy bias almost a decade ago, the standard analysis and agreement calculation recommendations still focus on single elicitation (Vatavu and Wobbrock, 2022), i.e. when one person participates alone and makes only one proposal per referent. The study of Morris (2012) is the only known example of group elicitation with multiple modalities. Elicitation research has not yet considered how to incorporate the three principles in the data analysis, and therefore, the options for quantitatively analyzing agreement are still limited.

5.7 Conclusion

In this work, we explored how different interaction modalities can be used to explore data visually on large vertical displays. Our results suggest that unimodal and personal interactions are preferred, but a system should enable touch, pen, and speech interaction to support data exploration with direct manipulation and natural language according to user preferences. Participants favored touch and speech, alone or in combination, to perform low-level exploration tasks with diverse visualizations of spatio-temporal data. However, when taking the top proposals into account, the choices and combinations of modalities are diverse. An evaluation with a real-world system would help assess whether and how the interaction choices of users may match the results of the elicitation study. The interface design also needs to be considered, as it will influence the interaction cost and the positioning of the users. In our study, we used an interface as simple as possible and encouraged participants to suggest interface elements to add if they wished. When working collaboratively, participants either used a single modality in parallel or used two modalities in a sequence, one for direct manipulation and another for distant interaction. We provide the consensus set and our analysis of the interaction proposals elicited in the study, to inform the interaction design of visual systems for collaborative data exploration. Future work should consider other interaction modalities relevant to large vertical displays, such as proxemics, and participant groups from other application scenarios with different data and task types.

6 Talk to the Wall: The Role of Speech Interaction in Collaborative Work

The work presented in this chapter is based in particular on findings derived in Chapter 5. At the time of submission of this thesis, the content has not been published yet, but was to be submitted soon as an independent research paper.

6.1 Introduction

Wall displays come with a set of benefits and challenges for the visualization and analysis of large datasets. They provide more pixels to show large data sets and offer people "space to think" (Andrews et al., 2011). The large physical size also supports co-located collaboration (Isenberg et al., 2011b) and promotes physical navigation for better performance (Ball et al., 2007). On the other hand, large displays require different, possibly unfamiliar, interaction techniques that may be unlike standards known from mouse and keyboard coupled with windows, icons, menus, and pointers (WIMP). On large displays, touch is a classic way to support close-up interaction. Touch, however, cannot be the only interaction modality for

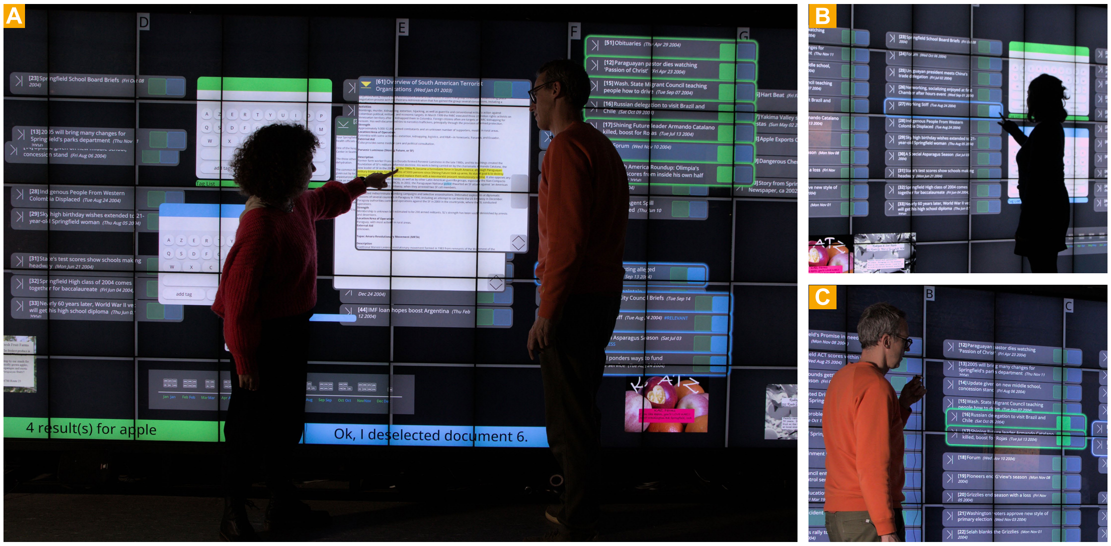


Figure 6.1: Collaborative sensemaking system on the wall-sized display: (a) Participants working together on solving the task. (b) Selecting a document via touch interaction. (c) Searching for a keyword via speech interaction.

wall displays, because people often stand far away from the wall to look at overviews of the shown data. For interaction distant from the wall, researchers have proposed techniques that do not require touching the display such as laser pointers (Myers et al., 2002), mobile phones/smartwatches (Horak et al., 2018), gestural interfaces (Nancel et al., 2011), or speech (Saktheeswaran et al., 2020). Integrating different interaction modalities is a dedicated research challenge in this context, since the seamless interplay between different close-up and distant interaction techniques is important to support sensemaking activities. In addition, it is not only important how interaction modalities integrate for one analyst, but also how they support multiple analysts working together. As soon as multiple people collaborate in front of a wall display, their interactions should not interfere and should not break a harmonious coordination of tasks and activities between team members.

In this paper, we focus on exploring two specific interaction modalities during a collaborative sensemaking task: touch and speech. We chose touch as the most common type of close-up interaction modality for wall displays. Our choice of speech is motivated by participants' strong preference for speech commands on a large vertical display in our recent collaborative elicitation study (Molina León et al., 2024b) when compared to touch, pen, and mid-air gestures. Speech (or natural language) interaction is also a very relevant candidate to study, as it is not yet clear how its use may influence the communication between collaborators, their awareness of each others' activities, and whether personal characteristics may influence the choice of using speech at all. Since little is still known about how speech and touch may be used in practice during collaborative analysis, we conducted an exploratory study, focused on how dyads or pairs solve a sensemaking task. More specifically, we were interested in their interaction choices as well as in how multimodal interaction and collaboration styles intertwine.

To investigate these aspects, we designed a wall-based interactive system supporting touch gestures and speech commands for co-located collaborative work. We conducted the study with 20 participants (10 pairs) asking them to solve a fictitious mystery by interacting with a collection of documents on the wall display. We included interaction techniques supported by both modalities, to allow participants to use any of them. Analyzing their choices, movements, and collaboration styles, we found that participants mostly used speech interaction from a distance as well as for global tasks, such as sorting. Through the examination of personality traits, we found that personal characteristics can predict the use of speech commands: the more agreeable (i.e., cooperative) a person was, the less likely they were to use speech interaction. Participants stood at an average of 1.52 meters of the display when using speech commands, which suggests they effectively used them to interact from a distance.

In terms of collaboration styles, participants spent most of the time in closely coupled collaboration and leveraged speech interaction evenly across close and loose collaboration. They stood nearer to each other while collaborating closely, but they did not step further away from each other to use speech commands. While distracting their partner through speech interaction was a common concern, only three participants reported a negative effect, and 14 claimed to be aware of the actions of their partner through their speech commands.

In summary, we make the following contributions:

1. We present our insights on the interaction choices, movements, and collaboration styles of participants and on the interplay with speech interaction.

2. We present our insights on how personality traits relate to the use of interaction modalities, such as speech commands.

3. We introduce TOUCHTALKINTERACTIVE, an open-source system with touch and speech interaction to support co-located collaborative work of pairs.
4. Based on our findings, we derive a set of design considerations on how to support co-located collaborative work leveraging speech interaction on wall-based interactive systems.

6.2 RELATED WORK

We present relevant work related to our research on the topics of interaction on large vertical displays, collaborative visual analytics, and personality traits.

6.2.1 Interaction with large vertical displays

Large vertical displays provide more pixels and more space to interact from (Belkacem et al., 2022). As the space in front of the display allows physical navigation, which can lead to better performance (Ball et al., 2007), wall-based systems should provide different ways of interacting from varying distances, to make the most of the surface and space. Touch is the standard interaction modality to support direct manipulation, but it comes with the disadvantage that the display areas beyond the arm's reach become inaccessible. Thus, researchers have come up with interaction techniques to virtually move the display content towards the user (Riehmann et al., 2020). Interacting with a pen or stylus is another type of direct manipulation that people leverage to write annotations (Molina León et al., 2024b) and author visualizations (Walny et al., 2012). While touch and pen can be a powerful combination (Hinckley et al., 2010), we decided to leverage touch only, since it is the most common modality supported and since it requires no additional hardware. However, direct manipulation requires participants to stand close to the wall display. Thus, we need other interaction modalities to facilitate interaction from a distance. For instance, Nancel et al. (2011) evaluated and proposed mid-air gestures to navigate through pan and zoom interactions on wall displays. Another possibility is to use additional devices, such as mobile phones (Langner et al., 2019) and smartwatches (Horak et al., 2018) to interact from a distance. However, our work focuses on supporting distant interaction without the need of additional handheld devices.

Previous work has studied the combination of multiple interaction modalities on large vertical displays. Srinivasan and Stasko (2018) proposed combining direct manipulation and speech commands to interact with network visualizations and found that participants chose different modalities across tasks, e.g., speech for filtering and touch for highlighting connections between nodes. DataBreeze (Srinivasan et al., 2021) combines touch, pen, and speech to interact with unit visualizations. While interacting, participants tended to use speech for global tasks and touch-based context menus for local tasks. Accordingly, we incorporated examples of such interactions in our system, to examine whether the same happens in the context of collaborative work. Saktheeswaran et al. (2020) found that, when comparing touch, speech, and touch-and-speech interactions, participants preferred the multimodal interactions. However, Molina León et al. (2024b) found that people preferred speech commands, over touch, pen, and mid-air gestures, alone or in combination, for a series of exploratory tasks. Therefore, we examine the use of touch and speech in a real-world system to study how participants choose to interact.

6.2.2 Collaborative visual analytics

Researchers have proposed different ways of supporting multi-user interaction on large displays. Langner et al. (2019) proposed using touch and touch-enabled mobile devices in order to allow participants to interact from different distances from the display. Badam et al. (2016) designed interaction techniques combining proxemics with mid-air gestures to interact with lenses on wall displays. James et al. (2023) proposed to leverage augmented reality to extend the interaction possibilities in the space in front of the display. We designed a system to leverage touch and speech interaction between two users in parallel if they want.

When people do not only interact with a system, but also with each other, social aspects need to be considered, such as collaborative distance, awareness, and communication among team members (Lee et al., 2012). We examine how the use of speech may influence these aspects, e.g., encouraging people to move away from each other before using a speech command. We studied how speech interaction may influence awareness and its interplay with human communication. Territoriality and privacy are other aspects to consider. For example, Reipschlager et al. (2021) proposed using augmented reality to extend the display surface by creating a personal virtual space for each collaborator. While these aspects are out of the scope of our work, we designed our system to provide a set of interface elements (e.g., tag list, virtual keyboard) that would allow participants to work alone if desired.

Regarding the use of multiple interaction modalities for collaborative work, Tse et al. (2008) explored how combining speech and hand gestures affected collaborative work on tabletops. The authors identified the use of speech in parallel commands as a design issue, i.e., if speech commands are often used for tasks that people need to perform in parallel, one person may decide to work sequentially to avoid voice overlapping. Their findings inspired us to study the effects of speech commands on wall displays. Furthermore, the way people work together can qualitatively vary depending on how closely or loosely they collaborate. Tang et al. (2006) proposed a set of codes to classify collaboration styles according to how strong the relationship between the activities of each person was. Isenberg et al. (2010) extended the code set for analyzing collaborative work on tabletops to take multiple data views into account. Other researchers have extended or adapted the collaboration styles proposed by Tang et al. and Isenberg et al. for scenarios where more than one device is involved (Brudy et al., 2018) or for hybrid collaborations (Neumayr et al., 2018). However, those characteristics do not apply to our non-hybrid scenario with a single display only. Therefore, we use the set of codes of Isenberg et al. to analyze the collaborative work of our participants. We selected the dataset and designed our experiment based on their work and on follow-up studies on collaborative work (Andrews et al., 2010; Jakobsen and Hornbæk, 2014) to facilitate the comparison and validation of our findings.

6.2.3 Personality traits

Designing one-size-fits-all interfaces for visualization systems is problematic, because the individual differences among people can affect how a person uses a visualization (Liu et al., 2020). Thus, researchers have looked at specific personality traits to analyze whether and how the treats may affect the experience of using visualizations. The most known personality research model is the Five-Factor Model. As its name suggests, it refers to five personality traits that describe individual differences: extraversion, openness, agreeableness, conscientiousness, and neuroticism. We focus on the first three traits, as we consider them relevant for speech interaction and collaborative work.

Extraversion is the first of those traits. Extroverted people are outgoing and sociable (Völkel et al., 2020). In a study on task performance, Green and Fisher (2010) found that extroverts solved search tasks faster than introverts. However, introverts gained more insights. Ziemkiewicz et al. (2013) found that introverts took more time to analyze the task and, thus, solved more tasks accurately than introverts. Given the characteristics of speech interaction, we hypothesize that extraverted participants use speech interaction more often than introverts.

Openness to experience is another personality trait that could influence the choice of using speech interaction, as it refers to curiosity and the personal wish to seek new experiences (DeYoung et al., 2012; Völkel et al., 2020). According to Liu et al. (2020), openness is a personality trait that is underexplored in data visualization research. We expect their willingness to experiment with less common interaction modalities (such as speech) may influence their choices.

Agreeableness is the third trait we take into account, as it is considered a positive trait for people to succeed in collaborative work (Liu et al., 2020). It refers to being cooperative (Völkel et al., 2020). We hypothesize that the use of speech interaction correlates with low agreeableness. In other words, we suspect that they will be reluctant to use speech so as not to disturb their colleagues.

We use the Five-Factor Model due to its clear dominance in personality research. However, as with any model, it has limitations. For example, previous work argues that this model cannot properly describe personalities from all cultures. While there are dozens of personality assessment instruments, many of them are copyrighted and users need to pay for using them. Therefore, we chose the freely available IPIP-NEO-60 of Maples-Keller et al. (2019), which suggests 12 questions per trait, and it is, thus, more practical than others available with over 300 questions.

6.3 Experiment Design

To explore the synergy between speech and touch interaction, we designed an in-person experiment where pairs of participants had to solve a sensemaking task collaboratively. Our goal was to investigate the following two research questions:

RQ1 How do partners use touch and speech individually while working as part of a team?

RQ2 How are touch and speech used during collaboration? When is each modality chosen and how does it relate to in-team communication and coordination?

We asked each pair to solve a fictitious mystery described in a collection of text documents that could be interacted with on a touch and speech-enabled wall display. Participants were free to use any touch gestures and speech commands supported by the system. To understand participants' choice of modalities, we ensured that all commands were available in both speech and touch; see Section 6.4 for details on the interaction techniques. We video-recorded the sessions, tracked the position of the participants, and logged their interactions with the system.

Before the main experiment, we conducted two pilot studies. A first pilot study with two participants helped us to finalize the list of interaction techniques the system should support to allow participants to solve the task. A second pilot study with another pair of participants helped us to identify usability issues that were then solved before the main study and to refine the protocol and the setup in terms of audio capture when multiple people were talking simultaneously.

6.3.1 Participants

We recruited 20 participants organized into 10 pairs. We recruited the pairs from within a research organization, as the goal was to recruit people who interact with data as part of their work. People did not need to know each other beforehand. Participants were required to be at least 18 years old and to be fluent English speakers. The study was approved by the corresponding IRB (Inria COERLE, opinion 2023-39).

6.3.2 Dataset and Task

We created an adapted version of the dataset used in the "Stegosaurus" scenario from the interactive session of the 2006 VAST challenge (Grinstein et al., 2006). This dataset has been used in previous work to analyze how groups can solve a sensemaking task collaboratively (Andrews et al., 2010; Isenberg et al., 2010; Jakobsen and Hornbæk, 2014). Our dataset contained 61 text files and four images that described a fictitious crime. Based on that dataset, we asked participants to generate hypotheses for a police investigation about the fictitious crime that the dataset described. The text files were news articles from a fictitious town called Springfield and miscellaneous other documents, such as a list of diseases and a list of

We modified the town name (previously named Alderwood) and other proper nouns in the dataset to ensure the speech engine would understand them.

terrorist organizations. From the 61 text files, 10 news articles contained relevant information about what happened, seven provided background information, and the miscellaneous documents gave some important clues about the evidence (e.g., the characteristics of the chemical involved). Similar to previous work (Isenberg et al., 2010; Jakobsen and Hornbæk, 2014), we expected participants to explore most documents and images to then come up with their hypotheses together.

6.3.3 Personality traits

As discussed in Section 6.2.3, we measured three personality traits for each participant: extraversion, openness to experience, and agreeableness. We used the IPIP-NEO-60 representation by Maples-Keller et al. (2019) of the NEO PI-R domains from the instrument of Costa and McCrae (1992). We used the data to calculate a score per participant that we could operationalize as a description of a personality trait and compare to the likelihood of using speech interaction.

6.3.4 Procedure

We started the experiment by explaining our research motivation to the participants and handing out a consent form for them to read and sign. If both participants gave consent, we proceeded to give each of them a pre-questionnaire to fill out. The questionnaire included five questions about their previous experience with interactive technologies, two questions about their collaboration experience with their partner, two demographic questions, and 36 questions from the personality traits measure of Maples-Keller et al. (2019). Afterward, we proceeded to a 10-minute training session. The experimenter introduced the system and then asked participants to go through a list of all supported interactions using both speech and touch gestures, asking them to use every interaction at least once. We proceeded to the main phase of the study once the participants felt confident enough to interact with the system independently. We also provided participants with a cheat sheet (one page), listing all the interaction techniques supported by the system and describing the necessary actions for using them with touch and speech.

The main phase of the study took 45 minutes. This phase started with the participants reading a one-page background document that introduced the task scenario. That document told the story of a mysterious incident in the fictional city of Springfield that the police wished to investigate further. To generate hypotheses, explore the dataset, and solve the mystery, participants could interact with the system as they wished. They could ask questions to the experimenter about the system, but they did not receive additional help to solve the mystery. Once the 45 minutes were over, we asked participants to fill out a post-questionnaire about their experience in terms of collaboration and interaction modalities. Afterwards, we conducted a short interview where we asked them to elaborate on what they found out about the mystery and their impressions of the interaction modalities, as well as of the collaborative work with their partner. We did not control the initial position of the participants; they were free to move in the space in front of the wall display as they wished. We also did not measure their performance, because our goal was to generate holistic, qualitative, observational data in order to understand how they interacted with each other and the system—and not to conduct a statistical analysis.

6.3.5 Data collection and analysis

To answer **RQ1**, we logged the interactions of the participants to assess how often they used each interaction modality and for what tasks. To take the distance of the participants to the screen into account, we used a Vicon motion tracking system to document the movements of the participants in front of the wall display.

We video-recorded the main phase of the study with a camera located at the back of the room and a microphone above the wall display. We analyzed the videos qualitatively to identify the interaction and collaboration styles of the participants in order to be able to answer **RQ2**. We also analyzed the videos to log collaboration breakdowns (e.g., when participants talked over each other). We applied the codes associated with collaboration styles proposed by Isenberg et al. (2010). One author of this paper coded the videos in two passes. The first pass was meant to validate and complete the interaction logs and the second was focused on coding the collaboration styles.

The post-questionnaire helped to obtain subjective feedback of the participants about the collaborative experience and the interactions. We were especially interested in whether speech created any issues in the collaboration.

6.4 System and interaction design

We designed and implemented an interactive system called TouchTalkInteractive to support co-located collaborative sensemaking on wall-sized displays. This visual analytics system is inspired by previous research-oriented systems created to support collaborative tasks (Andrews et al., 2010; Isenberg and Fisher, 2009; Jakobsen and Hornbæk, 2014) and by studies on interaction with large vertical displays (Molina León et al., 2024b; Saktheeswaran et al., 2020). We designed the system to enable the use of speech interaction in a collaborative scenario. While previous systems supported speech and other interaction modalities on large vertical displays (Saktheeswaran et al., 2020; Srinivasan et al., 2021), TouchTalkInteractive is the first one to support speech interaction for collaborative work on wall displays. It is designed to assist two people in exploring and making sense of a large collection of documents through touch gestures and speech commands. Each document is visually represented as a unit that can be opened or extended to read the content. The system includes a timeline to interact with the documents according to the temporal dimension, and it also supports images. By default, the images are positioned at the bottom of the screen, as previous work suggests that the lower area of wall displays should not be used for data representations (Bezerianos and Isenberg, 2012). Each person has the option to directly interact with one or multiple documents at a time, and both persons can interact simultaneously. Each person has a designated color (in our experiment, blue or green) and a set of interface elements in their own color, e.g., a keyboard to search for keywords. That way, participants are free to work individually if they wish to. Figure 6.2 shows the system interface.

6.4.1 Interaction techniques

Our goal was to support a comprehensive list of interactions with the shown documents, both through touch gestures and speech commands so that each of the two participants could interact in parallel while choosing any of the two modalities. Although single multimodal interactions would have been possible (i.e., combining speech and touch in one command), we kept the modalities separate to compare our results to the elicitation study of our own related work (Molina León et al., 2024b) and to examine when and in what contexts participants would choose either modality. We present the 20 interaction techniques that the system supports in Table 6.1.

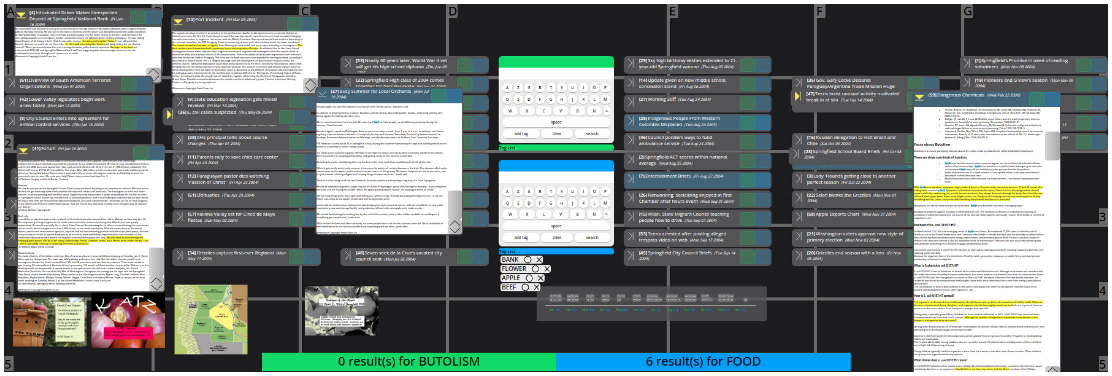


Figure 6.2: Screenshot of the system at the end of one session. The blue and green colors are used to identify the interface elements associated with each participant. The bars on the bottom of the screen give feedback to the user regarding the outcome of their last speech command.

We carefully chose the techniques to facilitate co-located and synchronous work, based on previous work on multimodal and collaborative interaction by Isenberg et al. (2010), Jakobsen and Hornbæk (2014), Molina León et al. (2024b), and Saktheeswaran et al. (2020).

6.4.2 Setup

Our wall display was 5.91 × 1.96 meters in size, composed of 75 liquid-crystal displays (LCDs) with a total resolution of 14400 × 4800 pixels (at 60 ppi), and controlled by a cluster of 10 computers. The wall was placed in a room of 31.5 m^2, and it recognizes touch input via infrared light. For recording the position of the participants, each of them wore a 3D tracker, secured using a headband. Each participant had a wireless mouse and a microphone to activate the speech recognition and therefore issued commands without the need for a wakeword.

TouchTalkInteractive is a web-based application, implemented in Typescript and Svelte. The documents are provided to the system as JSON files. The speech recognition was implemented on Rhino, a deep learning speech-to-intent engine by Picovoice Inc. (2023). Through the engine, we trained an English-based model according to the vocabulary used in the dataset documents. The open-source code is available at https://gitlab.inria.fr/aviz/TouchTalkInteractive.

Table 6.1: Interaction techniques supported by the system. Each interaction can be performed with touch gestures and speech commands.

Task	Touch interaction	Speech command
Open document	Tap on the left corner of header (toggle)	*“Open document `12`”*
Close document	Tap on the left corner of header (toggle)	*“Close document `12`”*
Select document	Tap on radio box on the document header	*“Select document `12`”*
Select documents per month	Tap on the month name on the timeline	*“Select documents of the month `May`”*
Select documents per tag	On the tag list, tap on the tag name	*“Select documents with the tag `money`”*
Deselect document	Tap on the header of the selected document	*“Deselect `12`”*
Deselect all	Long press on canvas to open the context menu. Tap “Deselect”.	*“Deselect all”*
Sort	Long press on canvas to open the context menu. Tap “Sort”. Tap on sort parameter.	*“Sort by `date`”*
Resize document	Drag bottom-right corner of an opened document.	*“Make document `12` smaller”*
Move a document	Drag document by its header.	*“Move document `12` to column `A`, row `3`”*
Move multiple documents	Drag selected documents by dragging one of them.	*“Move selected to column `A`, row `3`”*
Navigate in document	Drag scrolling bar inside document	*“Scroll `down` in document `12`”*
Mark document	Tap on sentence (toggle)	*“Mark [sentence `1` in] document `12`”* (first by default)
Clear document	Tap on sentence (toggle)	*“Clear sentence `1` in document `12`”*
Search	Tap phrase on keyboard. Then, tap on “search”.	*“Search for `flowers`”*
Clear search	Tap on “clear” on keyboard.	*“Clear search”*
Add tag	Type tag name on keyboard. Tap on “add tag.”	*“Add the tag `flowers`”*
Assign tag to document	Select document. Tap on O icon next to the tag	*“Assign tag `flowers` to the selected documents”*
Remove tag from document	Select tagged document. Tap on O icon next to the tag	*“Remove tag `flowers` from the selected documents”*
Delete tag	On the tag list, tap on X icon next to the tag	*“Delete the tag `flowers`”*

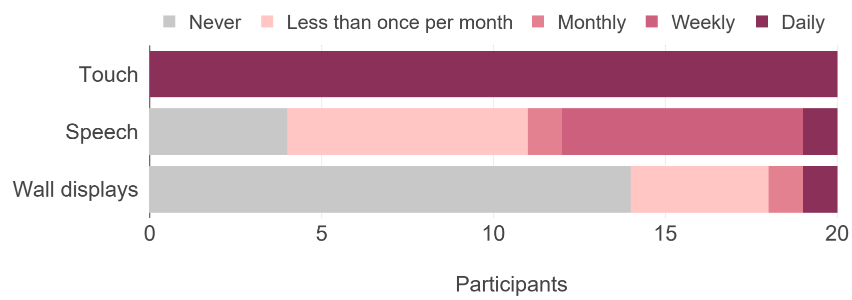


Figure 6.3: Previous experience reported by the participants in the pre-questionnaire.

6.5 Results

We recruited 20 participants (five identified as women, and fifteen as men). The average age was 27 years. All participants reported interacting with touch-based systems daily. Nine participants used speech interaction at least monthly, while four had never used it before. For 14 participants, this was the first time interacting with a wall display (see Figure 6.3). We did not require the participants to know each other before the experiment, in order to foster getting a diverse sample of collaborators. Sixteen participants reported to work closely with others at least once a week. Six pairs were familiar with each other and four pairs were not familiar with each other before participating.

6.5.1 Personality traits

Based on their answers to the personality questions in the pre-questionnaire, we divided participants into three groups per trait: those with low, those with high, and those with average scores, in comparison to the sample, according to the mean and standard deviation, as suggested in the literature Community and Services (2023) and Ziemkiewicz et al. (2013). Seven participants scored high on extraversion, six were average, and seven scored low. Seven participants scored high on openness to new experiences, eight were average, and five were low. Finally, nine participants scored high on agreeableness, three were average, and eight scored low.

We compared the scores of the three personality traits against the number of times each participant attempted to use speech interaction (regardless of the task and outcome) in order to examine whether those traits could relate to the tendency of using speech interaction. We visualize the relations in Figure 6.4.

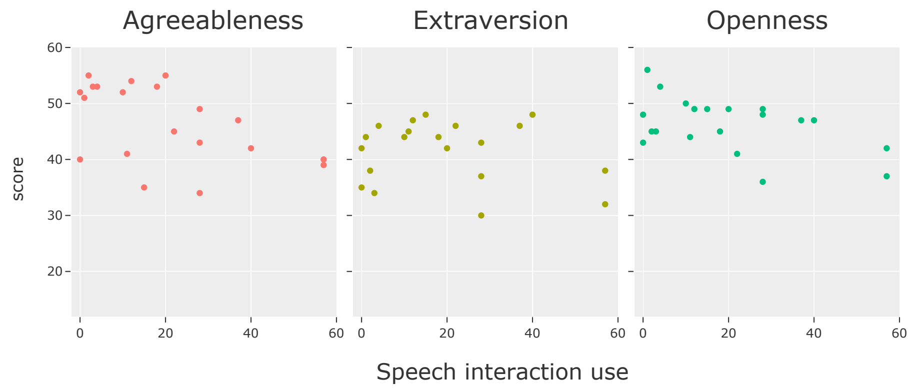


Figure 6.4: Correlation of agreeableness scores and total count of speech interactions per participant.

The Pearson's correlation between agreeableness and speech use was significant ($r(20) = -0.502, p = .02$), suggesting that there is a strong relationship between how agreeable a participant was and how often they used speech interaction. The more cooperative a person was, the fewer speech commands they used, potentially to avoid interrupting their partner. In contrast, the correlation between extraversion and speech use was not significant ($r(20) = -0.178, p = .45$). Furthermore, contrary to our expectations, we found a negative correlation between openness to experience and speech use ($r(20) = -0.490, p = .028$). The coefficient suggests that the less open to new experiences the person was, the more speech commands they used.

6.5.2 Interaction modalities per task

In total, we registered 4020 interactions from 20 participants. Of those, 3633 were touch gestures, and 387 were speech commands. Given that some actions tend to happen more often than others (e.g., moving documents), we analyzed the interactions per task. The system supported 20 tasks. However, we excluded the interactions associated with the task of *select documents per month* from the data analysis, given that we had technical problems with that feature in the first two sessions.

Participants were free to choose how to interact to solve the sensemaking task, after having tried all the interaction techniques with both modalities in the training session. Overall, participants made use of all the interaction techniques provided. The most used ones were: searching for keywords, opening, closing, and

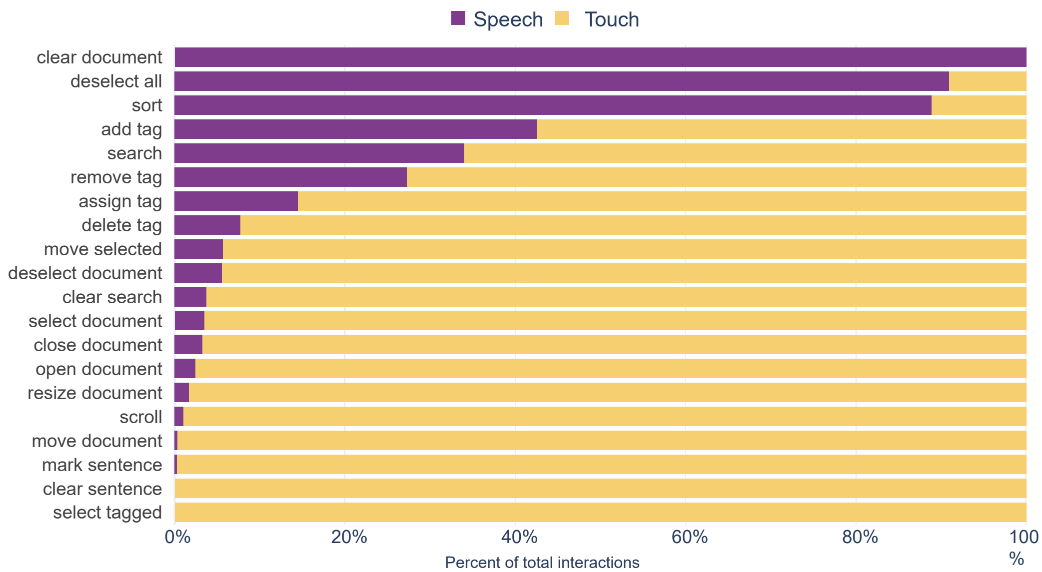


Figure 6.5: Distribution of the interactions across tasks and modalities.

moving documents, as well as highlighting and clearing sentences of interest in the documents. Most groups created and assigned tags to documents, but only a few removed or deleted those tags. Participants selected documents one by one more often than per tag or month. Similarly, they deselected individually more often than *deselecting all*.

In Figure 6.5, we present the recorded interactions distributed across tasks, with the proportion of modalities used. We show only the successful interactions (e.g., without speech recognition errors) where the system interpreted the task correctly. Overall, participants mainly used touch interaction, but there were exceptions among the tasks. Participants cleared sentences, sorted the documents, and used the *deselect all* command more often with speech interaction. To sort, nine participants used speech only, and two participants used touch (once in both cases). To deselect all documents, seven participants used only speech and only one participant used touch. To search for keywords, 18 participants used the touch-based keyboard, while six participants used speech commands more often than touch. For opening and closing documents, all except one participant used touch gestures more often than speech. For selecting single documents, one person used only speech while everyone else used touch gestures more often (or only). The diversity of the modality choices suggests that participants were able to use both modalities, but made different personal choices. For instance, while one participant (P12) interacted with speech commands throughout the whole session, another participant (P20) only used speech once.

In Figure 6.6, we present the modality preferences of the participants after interacting with the system. Most participants preferred speech interaction to sort the documents. For keyword search, more than half of the participants preferred speech or favored touch and speech equally. To clear search results, half of the participants preferred speech or both. For the remaining tasks, touch was the preferred modality. To select documents with a given tag, less than half of the participants preferred touch, but six of them wished to do it with either modality. Some participants made remarks about their reasoning when they chose the *I prefer them equally* answer. For example, P2 pointed out that choosing between touch and speech for closing a document depended on their distance from the document. Similarly, for moving documents, the choice would depend on the distance between the current and the desired position of the document on the wall display. Another participant also referred to the distance to the wall and to the document as a decisive factor in choosing how to open a document. For two participants, speech would be the first choice to search for a keyword, and if there would be a recognition error, they would switch to the touch-based keyboard. Another participant considered that searching by touch took too much time.

6.5.3 Speech interaction from a distance

We compared the frequency of speech commands with the distance of the participants from the wall display, based on the interaction logs and the tracking data we collected during the study.

While touch interaction required standing at a close distance, participants could interact with speech from any position. The data we collected about the participants' movements suggest that they tended to stay away from the wall while using speech commands. On average, participants stood at 1.52 meters from the display when interacting with speech. Furthermore, when a participant used a speech command, the average distance to their partner was 1.88 meters (std=0.86) and the median was 1.80 m. In contrast, when someone interacted with touch gestures, the distance between partners varied more, with an average of 1.94 meters (std=1.25) and a

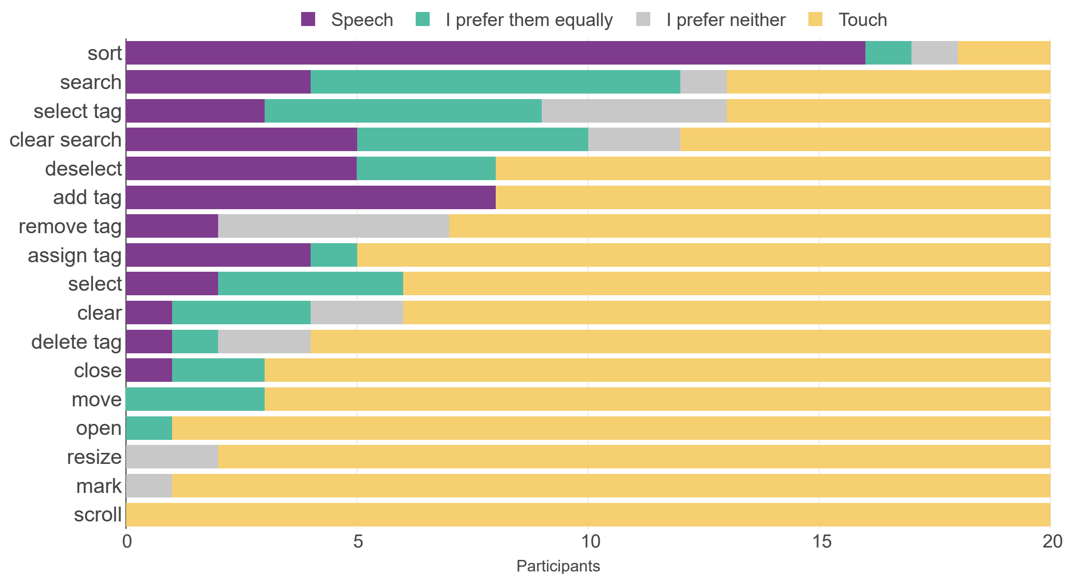


Figure 6.6: Modality preferences of the participants per task.

median of 1.52. While the larger variation of the distances associated with touch may be due to the higher frequency of touch gestures overall, the distances indicate that participants did not intentionally go away from their partner to use speech commands.

6.5.4 Collaboration styles

We qualitatively coded the collaboration styles of the participants throughout the sessions according to the codes proposed by Isenberg et al. (2010). Overall, participants spent more time in closely coupled collaboration than in loosely coupled collaboration. On average, the groups spent 54% of the session collaborating closely and 36% loosely. The remaining 10% of the time corresponds to periods of no clear collaboration, e.g., when participants talked to the experimenter.

All groups reported to have divided tasks among the two participants while solving the mystery. In the post-questionnaire, we asked each participant to estimate the proportion of time they spent working together with their partner, alone researching a shared question, and alone researching their own question. On average, participants reported to have spent 31% of the time working with their partner ($std = 21.05$), 31% alone researching a shared question ($std = 19.43$), and 38%

researching their own question ($std = 22.85$). When asked about their effectiveness as a team, nine out of the 10 groups considered that they worked together effectively or very effectively to solve the mystery.

Figure 6.7 (right) shows the average distance to the wall of each participant per collaboration style. Most participants stood further away from the wall while collaborating closely with their partners than while collaborating loosely. However, the difference was small, given that, on average, participants tended to stand 1.02 meters away from the wall display in close collaboration, and 0.98 meters away while collaborating loosely. Most groups discussed and worked on the same specific problems while standing slightly further from the screen. The most common type of close collaboration was *active discussion*, which refers to the situations where participants discussed their findings and hypotheses about the mystery they were trying to solve. Meanwhile, the most common form of loose collaboration was working on different problems. Participants worked independently while standing closer to the screen, probably reading a document, while their partner interacted with another document related to a different problem.

In the post-questionnaire, we asked participants whether they were afraid to use speech commands to annoy their partner. The answers suggest that it was a common concern, although not shared by the majority of the participants: Eight participants were concerned, while eleven disagreed and one was undecided. When asked about being annoyed by the speech commands of their partner, fourteen participants disagreed, and only three were annoyed by the commands of their partner during the study. Furthermore, 14 participants reported having learned about what their partner was doing from their speech commands, suggesting that the use of speech interaction helped them to be aware of the actions of their partner. When we asked about the different ways they learned about what their partner was doing, participants learned by hearing the speech commands 5% of the time on average. They learned by noticing what the partner was touching on the screen 8% of the time. The two most common answers were that the partner explicitly said it (24%) and that they knew about the action because they performed it together (25%). These answers corresponded to the groups spending most of the time in close collaboration, actively discussing.

We also examined how the distance between the two participants changed across collaboration styles. Figure 6.7 (left) compares the average distance between closely and loosely coupled collaboration per pair. We found that the teams tended to be

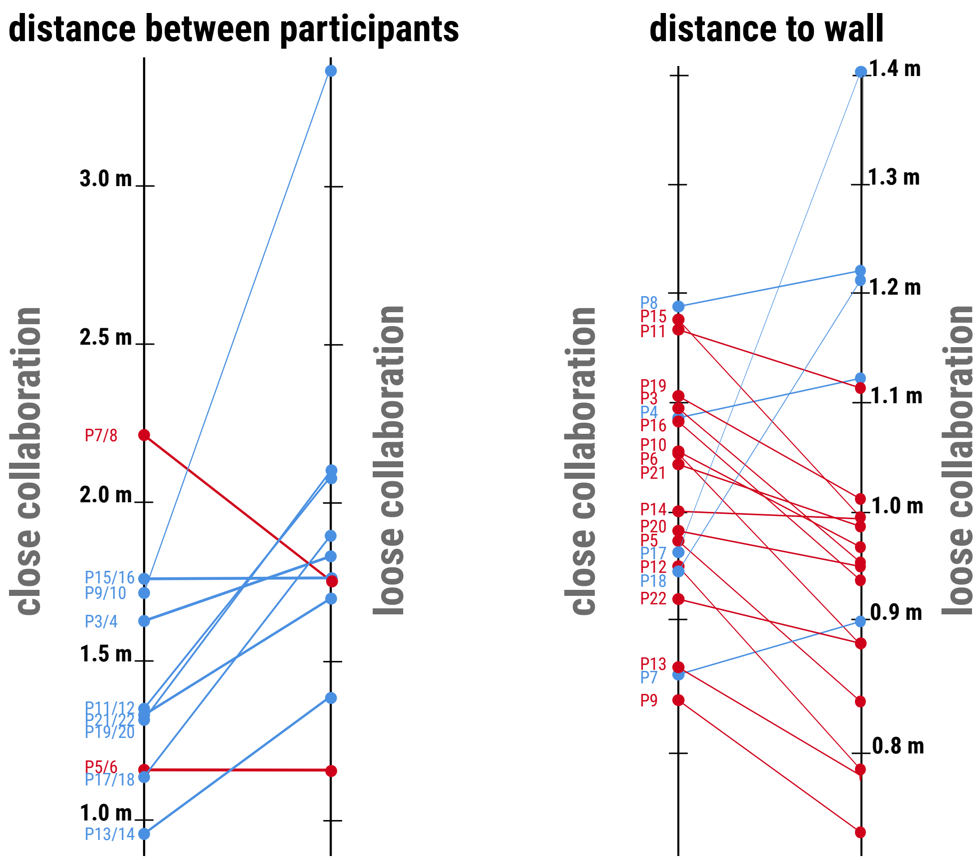


Figure 6.7: Slope charts comparing: LEFT: the distance between participants of each pair across collaboration styles. RIGHT: the distance to the wall of each participant across collaboration styles.

closer to each other during closely coupled collaboration than during loosely coupled collaboration. That difference was clearer and larger than in the case of the distance to the wall display. On average, participants stood at 1.45 meters from each other while collaborating closely and at 1.90 meters from each other while collaborating loosely. Therefore, they came closer or farther by half a meter. The largest difference corresponds to the pair of participants P5 and P6. Overall, they collaborated closely most of the time, but when they collaborated loosely — mainly at the end of the session — they tended to go to opposite ends of the display.

Modality choices during collaboration

We examined the relationship between interaction modalities and collaboration styles to find out whether participants tended to choose a specific modality while collaborating. Surprisingly, participants interacted very similarly across collaboration styles. From the interactions that overlapped with a collaboration style, 53% of the touch gestures happened during close collaboration and 47% during loose collaboration. Regarding speech commands, it was 49% and 51% respectively. Although speech interactions constituted less than 10% of all interactions, they were evenly performed across collaboration styles, suggesting that speech commands were deemed appropriate in both loose and close collaboration.

Awareness

All but one participant claimed to be aware of what their partner was doing, and 18 thought their partner was aware of their actions. Six groups reported that both persons were often working on the same question during the study, while one participant said that never happened (their partner answered "sometimes"). Nine participants reported to have looked across the wall often or very often, to find out what their partner was doing. However, six participants from six different groups had the impression that their partner often or very often found important information that would have helped them, but they only found out later. Given that speech interaction helped participants to be aware of what their partner was doing, using more speech commands may lead to more awareness.

6.5.5 Design Considerations

Based on our quantitative and qualitative findings, we draw the following design considerations for collaborative systems:

1. **Speech interaction enables distant interaction, while working individually and collaboratively**. Speech was preferred for tasks that influenced the complete dataset, such as sorting. It was not an obstacle to collaborate closely, and participants did not walk away from each other to use speech commands. Moreover, participants used half of their speech commands during loose collaboration and the other half during close collaboration, suggesting that speech was considered appropriate in both cases.

2. **Speech interaction contributes to the awareness of the partner**. In our exploratory study, 14 of 20 participants became aware of the actions of their partners through speech commands. While some participants were concerned about annoying their partner by using speech, only three of them reported being annoyed by the speech commands. Therefore, including speech interaction may have a positive effect on collaborative work, while carefully considering what interaction techniques to associate it with, to minimize the chances of distraction.

3. **Personal characteristics influence interaction choices**. From the three personality traits we examined, we found a significant relation between two of them and the frequency of speech commands. Thus, designers should consider that each person using an interactive system may choose differently how often to make use of speech commands. Furthermore, two of our participants pointed out that being non-native speakers made them hesitant to use speech, which suggests that, in a sample of native speakers only, participants may be more likely to leverage speech.

6.6 Discussion

Speech interaction is considered a promising interaction modality, because it allows people to express what they wish for, instead of taking the time to translate it into formulas or widget input. Moreover, it has the potential to make interactive visualizations more accessible (Hoque et al., 2018). However, it often suffers from a lack of trust due to recognition errors (Baughan et al., 2023).

Our findings show that speech commands can support distant interaction during collaboration without major issues, but they tend to only be used for specific tasks and still suffer from technological challenges. We used up-to-date tools powered by deep learning technology to enable speech interaction, but recognition errors were still an issue, and a few participants seemed to give up on speech after some errors. Compared with the dominance of speech commands in terms of how people *wished to* interact on large displays (Molina León et al., 2024b), it seems that the speech recognition technology still needs to catch up. In our study, a common concern among participants was that not being an English native speaker may have been the cause of the recognition errors. Another challenge we faced during the system implementation was that multiple words were not recognized as English words by the speech engine. Thus, we had to modify some terms from the original VAST dataset.

Despite the challenges, participants clearly favored speech interaction for global commands such as sorting and deselecting. Moreover, search commands via speech were common and were even part of the conversation within some teams. The sensemaking task of our study promoted collaborative work but relied heavily on browsing and reading multiple documents. Some participants argued that using a speech command to open or move a document was cumbersome, because it required first identifying the document by its number. That suggests that speech commands are more suitable for group tasks or set-related tasks, such as in previous work about data exploration by individuals (Srinivasan and Stasko, 2018). To explore which other tasks would benefit from leveraging speech, future work should systematically explore interaction taxonomies in order to determine when is speech interaction suitable and when not.

Regarding the characteristics of collaborative work, our study focused on examining how speech interaction may affect group communication, the distance between team members, and awareness. However, there are other social aspects relevant to collaborative work. Future work should investigate how speech interaction may affect privacy and territoriality (Azad et al., 2012). While our participants did not report any inconveniences, working separately and keeping interactions private may be more relevant in other collaborative scenarios. Moreover, if the team has more than two members, it would be possible for two people to talk while a third person interacts via speech commands. It remains to be studied whether speech interaction become a significant obstacle to communication then.

Finally, we did not focus on performance, as our study was rather exploratory. An important future research direction to assess speech interaction would be to investigate how the use of speech commands may influence performance in the context of collaborative work. With our findings, we hope to encourage interaction designers to leverage speech interaction for interactive systems not only for individuals but also for collaborative scenarios.

6.7 Conclusion

We conducted an in-depth exploratory study on combining touch gestures and speech commands for collaborative sensemaking to extend our understanding of whether and how speech interaction can be leveraged for co-located collaborative work. Our findings provide evidence that speech commands are a viable option to support distant interaction. The differences between modality preferences and actual use suggest that, in concordance with previous work (Molina León et al., 2024b), people are interested in using speech interaction, but other aspects, such as speech recognition errors, may influence their interaction choices on implemented systems. Furthermore, we found the tendency to use speech commands correlated with personality traits. Future work should investigate whether the choice to leverage other interaction modalities may be influenced by such personal characteristics.

Regarding the communication between team members, we did not find enough evidence that speech interaction was an obstacle to collaboration. On the contrary, 14 of 20 participants reported that speech commands contributed to becoming aware of the actions of their partner. Based on our findings, we introduced design considerations for collaborative systems that can serve interaction designers to incorporate speech commands in their designs if appropriate. Finally, we presented TouchTalkInteractive, a collaborative visual analytics system that facilitates touch and speech interaction for co-located collaborative sensemaking by pairs, openly available for future work to build upon.

7 Conclusion

This dissertation provides an in-depth perspective of how multimodal and collaborative interaction can be leveraged to support data exploration and sensemaking. In terms of visualization design, it contributes an evaluation of co-creation as a design methodology as well as data and task abstractions derived from a collaboration with experts from the social policy domain. Regarding interaction design for data visualization, it contributes to the characterization of interaction patterns and strategies on multimodal systems across a series of interactive surfaces. It also contributes to the understanding of interaction preferences regarding touch, pen, and speech interaction by analyzing interaction choices and proposals in the context of individual and collaborative work. Based on the quantitative and qualitative findings from three empirical studies, it provides design considerations for facilitating multimodal and collaborative interaction on interactive surfaces.

My goal was to investigate how different interaction modalities and devices can support data experts to visually explore and make sense of data, individually and collaboratively. I have approached this challenge within the following scope: In the context of social policy research as the main *application scenario*, I have focused on visual data exploration and sensemaking as the *processes* to support the data experts on. I have examined tablets, large vertical displays, and wall displays as target *devices* and evaluated touch, pen, speech, and mid-air gestures as the *interaction modalities* of interest. For the *number of users*, I have considered single-user and group scenarios, using pairs as the fundamental example of a group. In this final chapter, I recap the gained insights, bring together the final design considerations, reflect on the limitations of my work, and describe directions for future research.

7.1 Research Questions and Contributions

I structured the dissertation around four research questions (RQ). In this section, I will revisit them to discuss the corresponding research contributions.

The starting point was an interdisciplinary collaboration with experts from the social policy research domain to empower them through the co-creation of visual tools for their data work. We wanted to investigate interaction modalities and devices in a real-world scenario. Thus, we started by asking what data and tasks are relevant to support real-world experts in their data-driven work (**RQ1**), following the analysis framework of Munzner (2014). We answered this question by conducting a visualization design study comprised of a series of co-creation workshops, including surveys, interviews, and a user study, to elicit the design requirements of the experts. We applied the co-creation methodology because we wished to design tools *for* them but also *with* them. To assess the application of that methodology to visualization design, we conducted a formative and summative evaluation during the design process.

The design study led to the first major contribution of the dissertation, which can be divided into two: First, we characterized the **data and task abstractions** based on the workflow of our collaborators to later leverage appropriate visualization and interaction techniques. The experts focused on spatio-temporal data, and their goal was to visually explore said data, as they were collecting more than a thousand new datasets on the global history of social policies and needed computational support to comprehend them. Second, while previous work on visualization design had involved co-creation, there was no methodological evaluation. Thus, we contribute **the first evaluation of the co-creation methodology** for visualization design.

After defining the data and tasks of interest, the rest of the dissertation focused on how to support the experts in performing their work on different devices with multiple interaction modalities. First, we examined the single-user scenario and conducted a comparative evaluation with two interactive workplaces for visual data exploration: a desktop-based one with a WIMP interface and another tablet-based one with touch, pen, and speech interaction. The goal was to investigate how the devices and modalities affect performance and user experience (**RQ2**). To achieve that, **we designed and developed a visualization system, the *Modality Explorer***, which facilitates elementary and synoptic exploratory tasks based on the task typology of Andrienko and Andrienko (2006), and related work (Duncan et al., 2020). Participants were significantly faster on the desktop and similarly accurate on the tablet, with whom they reported a better user experience. The interaction analysis led us to identify **interaction patterns** that revealed the experts had different **strategies** to solve the tasks across workplaces.

Afterward, we shifted our focus to collaborative work on large vertical displays. As the standard mouse and keyboard are not suitable for interacting on such displays, we conducted an elicitation study to investigate **user preferences** regarding multimodal and collaborative interaction with spatio-temporal data on large vertical displays (**RQ3**). As a result, we contributed the **consensus set and the analysis of the top proposals**. Overall, participants favored unimodal and personal interactions, using either speech, touch, or pen interaction. The dominant modality was speech interaction due to the preference of speech commands for 10 of 15 tasks. When interacting multimodally, they mainly preferred combining touch and speech (distant interaction), using touch gestures for executing the first step of the tasks (i.e., starting with close interaction). When collaborating, participants either interacted unimodally in parallel or multimodally in a sequence.

After eliciting interaction preferences, we investigated whether the preferences transferred to interaction choices and how the elicited interactions affect collaborative work (**RQ4**). In a sensemaking study with the ***TouchTalkInteractive* system**, participants interacted more often with touch than speech, in contrast to the elicited preferences. However, speech was still used most on tasks without a relation to display coordinates (e.g., global tasks). Despite speech recognition errors, participants tended to use speech commands from a distance, and the speech had no significant impact on the communication with their partner. We also found that speech commands contributed to awareness and their use correlated with personality traits.

7.2 Design Considerations For Multimodal and Collaborative Interactions

In Chapter 4, we provided design recommendations for multimodal visualization systems on tablets. In Chapter 5, we discussed whether visualization systems should support multimodality and collaborative interactions based on the elicited preferences. After analyzing the interaction choices for collaborative sensemaking in Chapter 6, we provided another set of design considerations. Now, looking back at the recommendations and reflections of the three studies together, I define the overall design considerations for multimodal and collaborative interactions for visualization systems, derived from the chapters.

DC1 On large displays, leverage speech commands for distant interactions, especially for global commands with little or no relation to display coordinates. Participants appreciated speech interaction on all the interactive surfaces, especially on the large displays. In Chapter 5, we hypothesized that speech becomes relevant as the display size increases. The results of the last study (Chapter 6) suggest this is true, not because of the display size but rather because of the distance between the person and the display or the person and the visual elements of interest. Through their qualitative feedback, some participants indicated that they would choose speech to reach far-away elements, even when standing close to the display. The consensus set of Chapter 5 also suggested that people prefer speech commands to interact when the task has no or only a loose connection to specific screen coordinates. That also corresponds to the interaction choices on the wall display, where speech was the most used modality for sorting and deselecting all, closely followed by the search features.

DC2 Provide multimodal interactions that combine direct manipulation with distant interaction. Touch and pen interaction, the modalities for direct manipulation, prevailed in the studies with both systems. While touch is the most common modality on interactive surfaces, the comparative evaluation of Chapter 4 suggests that the pen should be prioritized to support primary tasks due to its higher precision. Contrasting these results with those of Drucker et al. (2013) on tablets, we conclude that choosing between touch and pen interaction depends on the visualization technique. Regarding large displays, participants favored touch over the pen but still preferred the latter for high-precision tasks, such as selecting small visual elements. In the final collaborative study, participants tended to use speech commands during active discussions with their partner, at a distance from the wall, without major obstacles to team communication. Thus, speech is a potential modality to leverage for enabling distant interaction.

DC3 **Consider collaborative synonyms to enable collaborative interactions**. As defined in Chapter 5, *collaborative synonyms* are interactions (in this case, we focus on the multimodal ones) from which there are two identical, except for the number of people who perform them. Collaborative synonyms would help to ensure that multimodal interactions are supported consistently regardless of the number of people involved, and that would facilitate getting familiar with the interactions, similar to using consistent interactions across multiple views (Sadana and Stasko, 2016). As participants switch between loosely and closely coupled collaboration, collaborative synonyms can facilitate the transition by leveraging familiarity.

DC4 **Consider speech interaction to support collaborative work; it may become part of the team communication**. We conducted the exploratory study on sensemaking of Chapter 6 to investigate how speech interaction may influence the communication and work of co-located teams. Our results matched those of the elicitation study in Chapter 5 regarding the wish to incorporate speech commands in team conversations. While some participant discussions led to the decision to use a speech command, some participants of the elicitation study even wished for the system to listen to their dialogue to react accordingly. While distracting their partner through speech commands was a common concern, most participants acknowledged the interactions of their partner without objections.

7.3 Remaining Challenges

The research scope of this dissertation only covered part of the design space regarding interaction modalities, techniques, devices, and collaborative work. In this section, I reflect on the remaining challenges based on the limitations of the conducted research.

7.3.1 Range of Examined Interaction Modalities

Throughout the dissertation, I considered four interaction modalities: touch, speech, pen, and mid-air gestures. However, other modalities, such as gaze interaction and proxemics, are also relevant to collaborative work. For example, Dostal et al. (2014) used gaze tracking to adjust the appearance of visual elements, while Badam et al. (2016) found that techniques combining proxemics and mid-air gestures can support lens-related operations. Moreover, while the elicitation study considered all possible types of modality combinations (in parallel or sequentially, independent or combined) in the classification space of Gourdol et al. (1992), the implemented systems mainly support independent interactions (using one modality) that can be used both in parallel or in sequence with other modalities, optionally involving someone else. While those design choices reflect the user preferences we elicited in Chapter 5, combined interactions such as those incorporated in *DataBreeze* (Srinivasan et al., 2021) may be convenient in other collaboration scenarios.

7.3.2 Range of Examined Interaction Techniques

In terms of interaction techniques, the dissertation focuses on those enabling exploratory tasks for spatio-temporal data. While those were the tasks relevant to the main application domain, there are many other ways to interact with visualizations. Previous work has examined techniques related to visualization authoring, such as binding an attribute to a visual encoding on tablets (Srinivasan et al., 2020), or associated with specific visualization techniques, such as networks (Kister et al., 2015). Still, there are other actions (e.g., semantic zooming, clustering) that could support groups in exploring large datasets and may benefit from using multiple modalities.

7.3.3 Range of Examined Devices and Scenarios

We conducted studies with desktop computers, tablets, large vertical displays, and wall-sized displays because they are present in the workplace or suitable for collaborative work. Still, other devices such as smartwatches, smartphones, and head-mounted displays can support multiple interaction modalities. They come with their own set of challenges, such as small display size and spatial awareness, but they are increasingly commonplace, so we should study how to support specific data-driven work tasks on them. Furthermore, we investigated the use of tablets in offices, and while that demonstrated the real-world use of tablets at the workplace, it would be interesting to look at their use *on the go*.

7.3.4 Range of Co-located Collaborators

I considered groups of only two people to keep the research scope manageable, but having more than two group members would entail new aspects to consider. For example, Azad et al. (2012) studied co-located groups of up to four people around public displays and examined group formations within a group and between sub-groups. If sub-groups are possible, would each group take responsibility for one modality, or would each group assign different roles to its members? Moreover, there was no collaborative interaction among the most common proposals in the elicitation study. Future work should look into other processes and task types where collaborative interactions may feel more pertinent.

7.4 Future Directions

While my dissertation research has focused on supporting experts in exploring spatio-temporal data, there are other audiences and scenarios where leveraging multiple interaction modalities may be beneficial. In the following, I highlight three research directions to further examine, building upon my work.

7.4.1 Supporting Other Collaboration Scenarios

I have focused on co-located collaborations, but since the COVID-19 pandemic, distributed collaborations have become much more common. Such collaborations involve more than one device in contrast to the scenarios I examined with a single shared display. When collaborating in a virtual or mixed reality context, there are more factors to consider for finding modalities that fit the task, such as the lack of a physical display. Would collaborators still prefer combining direct manipulation with speech interaction? What about hybrid scenarios where participants use and share co-located and distributed setups? And what about asynchronous teamwork? In one of the related workshop papers, we discussed the challenges and opportunities of using speech interaction in these different scenarios (Molina León et al., 2023). For example, overlapping voices of collaborators would not be an issue in asynchronous work.

7.4.2 Towards Personal and Accessible Interactive Experiences

We only started exploring how modality choices may relate to personal characteristics. As cognitive factors such as spatial abilities and working memory vary across individuals, personality factors vary as well (Ziemkiewicz et al., 2012). In Chapter 6, we found that the agreeableness personality score correlated with the likelihood of using speech commands. The more agreeable or cooperative a person was, the less likely they were to use speech commands. This relation suggests that personal characteristics could help predict whether someone is likely to use a modality.

Data literacy is another variable to consider. For instance, people with low data literacy may find speech interaction easier to use for performing a data query than WIMP elements. What about using natural language interaction to facilitate collaborative work among people with different levels of data literacy?

Moreover, we cannot assume that everyone can operate any modality. Using touch gestures or a pen may be challenging for someone with visual or motor impairments. Thus, we should reflect on how multimodal interaction can ensure an accessible experience. If we think of the vision of Bolt (1980) to "Put that there", combining speech and mid-air gestures to interact may be quick and powerful, but relying on such interactions may become frustrating, as there are more possible errors that can stop the action. Therefore, multimodal interaction has the potential to either improve or make difficult the user experience.

7.4.3 Leveraging New Technologies

Since I started working on my dissertation, there have been plenty of relevant technological advancements. The launch of ChatGPT in November 2022 introduced and promoted significant improvements in natural language interaction. While the first version of the chatbot was only text-based, a new version with speech interaction has been available since a few months ago (OpenAI, 2023). Moreover, visualization commercial platforms, such as Tableau, have started to leverage generative artificial intelligence in new features such as *Einstein Copilot* and *Tableau Pulse* to provide a conversation-based experience to their clients (Nichols and Wang, 2023). While speech interaction suffers from recognition errors, these developments suggest that is likely going to improve, potentially increasing the use of conversational assistants.

Speech is not the only modality that has benefited from recent industry releases. The Apple Vision Pro headset released in 2024 provides mixed-reality experiences supporting interaction through mid-air gestures, gaze tracking, and speech. Such releases show that when the technology is mature enough, new possibilities open up. However, reviews about the new headset suggest that the current technology is not yet ready to provide the devices users wish for (e.g., lightweight augmented reality glasses). In that context, elicitation studies like the one presented in Chapter 5 can help us explore the design space for when technology is ready to deliver. The participant proposals may also evolve as the experience with the latest commercial products may influence them.

7.5 Critical Reflections

After presenting the final results, I reflect on the limitations of this dissertation. The research work I presented studies visualization and interaction design for diverse tasks and devices. Regarding end-users, it focuses on data experts, starting with the perspective of social policy researchers and then extending the sample to experts who work with data in their everyday lives. Thus, the generalization of the findings of the studies needs to be evaluated with other audiences, including the general public, to extend our understanding of how people make interaction choices and approach visualization systems in these settings. A higher diversity of data literacy may reveal other interaction choices and strategies for the given tasks.

As it is common practice in interaction design research, I investigated the proposed research questions through empirical studies with custom systems created for that purpose, partly due to the lack of support for multimodal and collaborative interaction in commercial products. However, conducting experiments with such systems may lead to *experimental demand characteristic effects*, which refer to participants being likely to give positive feedback because they collaborate with the researchers (Sedlmair et al., 2012). Future work should aim to replicate our findings with other systems, and ideally, commercial products.

In terms of collaboration, I focused on pairs and recruited participants who worked in the same institution, often with previous collaborative experience. However, collaborative work can take place over weeks or longer, and controlled experiments in a laboratory cannot account for the evolution of collaborative work over time. Future research should take collaboration duration into account and examine how participants collaborate after interacting more than once with a given system.

7.6 Concluding Remarks

Through a series of empirical studies and the development of visual multimodal systems, we extended our understanding of how people interact and wish to interact to solve exploratory and sensemaking tasks on different devices, and we provided technical solutions to support those tasks. We have made methodological, artifact, and empirical research contributions to the HCI, VIS, and CSCW research fields. Considering the technological advancements in modalities and collaborative tools of the last few years, we hope that the findings of this dissertation will serve to design multimodal and collaborative tools that facilitate advanced interactive experiences.

References

[1] Chadia Abras, Diane Maloney-Krichmar, Jenny Preece, and William Bainbridge. "Encyclopedia of human-computer interaction". In: *Thousand Oaks: Sage Publications* 37 (2004), pp. 445–456.

[2] Abdullah Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. "'I Am Iron Man': Priming Improves the Learnability and Memorability of User-Elicited Gestures". In: *Proc. CHI*. Yokohama, Japan: ACM, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445758.

[3] Thomas Alsop. *Number of tablet users worldwide from 2013 to 2021 (in billions)*. 2022. URL: https://www.statista.com/statistics/377977/tablet-users-worldwide-forecast/.

[4] Christopher Andrews, Alex Endert, and Chris North. "Space to Think: Large High-Resolution Displays for Sensemaking". In: *Proc. CHI*. Atlanta: ACM, 2010, pp. 55–64. ISBN: 9781605589299. DOI: 10.1145/1753326.1753336.

[5] Christopher Andrews, Alex Endert, Beth Yost, and Chris North. "Information visualization on large, high-resolution displays: Issues, challenges, and opportunities". In: *Information Visualization* 10.4 (2011), pp. 341–355. DOI: 10.1177/1473871611415997.

[6] Natalia Andrienko and Gennady Andrienko. *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media, 2006.

[7] Natalia Andrienko, Gennady Andrienko, and Peter Gatalsky. "Exploratory spatio-temporal visualization: an analytical review". In: *Journal of Visual Languages & Computing* 14.6 (2003). Visual Data Mining, pp. 503–541. ISSN: 1045-926X. DOI: 10.1016/S1045-926X(03)00046-6.

[8] Natalia Andrienko, Gennady Andrienko, and Peter Gatalsky. “Tools for visual comparison of spatial development scenarios”. In: *Proceedings on Seventh International Conference on Information Visualization, 2003. IV 2003.* 2003, pp. 237–244. DOI: 10.1109/IV.2003.1217985.

[9] Apple Inc. *Apple Maps*. https://apps.apple.com/us/app/maps/id915056765. Accessed: 07/2020. 2021. URL: https://apps.apple.com/us/app/maps/id915056765.

[10] Alec Azad, Jaime Ruiz, Daniel Vogel, Mark Hancock, and Edward Lank. “Territoriality and behaviour on and around large vertical publicly-shared displays”. In: *Proceedings of the Designing Interactive Systems Conference*. Newcastle Upon Tyne, United Kingdom: Association for Computing Machinery, 2012, pp. 468–477. ISBN: 9781450312103. DOI: 10.1145/2317956.2318025.

[11] Sriram Karthik Badam, Fereshteh Amini, Niklas Elmqvist, and Pourang Irani. “Supporting Visual Exploration for Multiple Users in Large Display Environments”. In: *IEEE Conference on Visual Analytics Science and Technology (VAST)*. 2016, pp. 1–10. DOI: 10.1109/VAST.2016.7883506.

[12] Sriram Karthik Badam, Andreas Mathisen, Roman Rädle, Clemens N. Klokmose, and Niklas Elmqvist. “Vistrates: A Component Model for Ubiquitous Analytics”. In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 586–596. DOI: 10.1109/TVCG.2018.2865144.

[13] Sriram Karthik Badam, Arjun Srinivasan, Niklas Elmqvist, and John Stasko. “Affordances of input modalities for visual data exploration in immersive environments”. In: *2nd Workshop on Immersive Analytics*. 2017.

[14] Robert Ball, Chris North, and Doug A. Bowman. “Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays”. In: *Proc. CHI*. San Jose: ACM, 2007, pp. 191–200. ISBN: 9781595935939. DOI: 10.1145/1240624.1240656.

[15] Patrick Baudisch, Nathaniel Good, and Paul Stewart. “Focus plus Context Screens: Combining Display Technology with Visualization Techniques”. In: *Proc. Annu. ACM Symp. on User Interface Software and Technol.* Orlando, Florida: ACM, 2001, pp. 31–40. ISBN: 158113438X. DOI: 10.1145/502348.502354.

[16] Amanda Baughan, Xuezhi Wang, Ariel Liu, Allison Mercurio, Jilin Chen, and Xiao Ma. "A Mixed-Methods Approach to Understanding User Trust after Voice Assistant Failures". In: *Proc. CHI*. Hamburg, Germany: ACM, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3581152.

[17] Dominikus Baur, Bongshin Lee, and Sheelagh Carpendale. "TouchWave: Kinetic Multi-Touch Manipulation for Hierarchical Stacked Graphs". In: *Proc. ITS*. Cambridge: ACM, 2012, pp. 255–264. ISBN: 9781450312097. DOI: 10.1145/2396636.2396675.

[18] Ilyasse Belkacem, Christian Tominski, Nicolas Médoc, Søren Knudsen, Raimund Dachselt, and Mohammad Ghoniem. *Interactive Visualization on Large High-Resolution Displays: A Survey*. 2022. arXiv: 2212.04346 [cs.HC].

[19] Anastasia Bezerianos and Petra Isenberg. "Perception of Visual Variables on Tiled Wall-Sized Displays for Information Visualization Applications". In: *IEEE Trans. Visual Comput. Graphics* 18.12 (Dec. 2012), pp. 2516–2525. DOI: 10.1109/TVCG.2012.251.

[20] Tanja Blascheck, Lonni Besançon, Anastasia Bezerianos, Bongshin Lee, and Petra Isenberg. "Glanceable Visualization: Studies of Data Comparison Performance on Smartwatches". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 630–640. DOI: 10.1109/TVCG.2018.2865142.

[21] Richard A. Bolt. "Put-that-there: Voice and gesture at the graphics interface". In: *SIGGRAPH Comput. Graph.* 14.3 (July 1980), pp. 262–270. ISSN: 0097-8930. DOI: 10.1145/965105.807503.

[22] Claus Bossen, Christian Dindler, and Ole Sejer Iversen. "Evaluation in Participatory Design: A Literature Survey". In: *Proceedings of the 14th Participatory Design Conference: Full Papers - Volume 1*. PDC 16. Aarhus, Denmark: Association for Computing Machinery, 2016, pp. 151–160. ISBN: 9781450340465. DOI: 10.1145/2940299.2940303.

[23] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. "Dş Data-Driven Documents". In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2301–2309. ISSN: 1941-0506. DOI: 10.1109/TVCG.2011.185.

[24] Ilya Boyandin, Enrico Bertini, and Denis Lalanne. "A Qualitative Study on the Exploration of Temporal Changes in Flow Maps with Animation and Small-Multiples". In: *Computer Graphics Forum* 31.3pt2 (2012), pp. 1005–1014. DOI: 10.1111/j.1467-8659.2012.03093.x.

[25] Matthew Brehmer, Bongshin Lee, Petra Isenberg, and Eun Kyoung Choe. "A Comparative Evaluation of Animation and Small Multiples for Trend Visualization on Mobile Phones". In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), pp. 364–374. ISSN: 2160-9306. DOI: 10.1109/TVCG.2019.2934397.

[26] Matthew Brehmer and Tamara Munzner. "A Multi-Level Typology of Abstract Visualization Tasks". In: *IEEE Trans. Visual Comput. Graphics* 19.12 (2013), pp. 2376–2385. DOI: 10.1109/TVCG.2013.124.

[27] BroadbandSearch.net. *Mobile Vs. Desktop Internet Usage (Latest 2021 Data)*. 2021. URL: https://www.broadbandsearch.net/blog/mobile-desktop-internet-usage-statistics.

[28] John Brooke et al. "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

[29] Frederik Brudy, Joshua Kevin Budiman, Steven Houben, and Nicolai Marquardt. "Investigating the Role of an Overview Device in Multi-Device Collaboration". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–13. ISBN: 9781450356206. DOI: 10.1145/3173574.3173874.

[30] Frederik Brudy, David Ledo, Saul Greenberg, and Andreas Butz. "Is Anyone Looking? Mitigating Shoulder Surfing on Public Displays through Awareness and Protection". In: *Proceedings of The International Symposium on Pervasive Displays*. Copenhagen: ACM, 2014, pp. 1–6. ISBN: 9781450329521. DOI: 10.1145/2611009.2611028.

[31] Christopher N. Bull, Will Simm, Bran Knowles, Oliver Bates, Nigel Davies, Anindita Banerjee, Lucas Introna, and Niall Hayes. "Mobile Age: Open Data Mobile Apps to Support Independent Living". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI

EA 17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 2410–2415. ISBN: 9781450346566. DOI: 10.1145/3027063.3053244.

[32] William Buxton. "Chunking and phrasing and the design of human-computer dialogues". In: *Readings in Human–Computer Interaction*. Elsevier, 1995, pp. 494–499.

[33] Stuart Card, Jock Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision To Think*. Morgan Kaufmann, Jan. 1999. ISBN: 978-1-55860-533-6.

[34] Francis G. Castles, Stephan Leibfried, Jane Lewis, Herbert Obinger, and Christopher Pierson, eds. *The Oxford Handbook of the Welfare State*. Oxford: Oxford University Press, 2010. ISBN: 978-0-19-957939-6. DOI: 10.1093/oxfordhb/9780199579396.001.0001.

[35] Mohammad Chegini, Shao Lin, Dirk Joachim Lehmann, Keith Andrews, and Tobias Schreck. "Interaction Concepts for Collaborative Visual Analysis of Scatterplots on Large Vertically-Mounted High-Resolution Multi-Touch Displays." In: *Forum Media Technology*. 2017, pp. 90–96.

[36] Citational Justice Collective, Amy Ogan, Frederick van Amstel, Gabriela Molina León, Juan Fernando Maestre, Kristin Williams, Nicola J Bidwell, Pedro Reynolds-Cuéllar, Saiph Savage, Sushil Oswal, and Vishal Sharma. "Why Do We Need to Learn about Citational Practices? Recognizing Knowledge Production from the Global Souths and Beyond". In: *XRDS* 29.3 (Apr. 2023), pp. 12–17. ISSN: 1528-4972. DOI: 10.1145/3589256.

[37] Darius Coelho, Ivan Chase, and Klaus Mueller. "PeckVis: A Visual Analytics Tool to Analyze Dominance Hierarchies in Small Groups". In: *IEEE Transactions on Visualization and Computer Graphics* 26.4 (2020), pp. 1650–1660. DOI: 10.1109/TVCG.2020.2969056.

[38] Philip R. Cohen, Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen, and Josh Clow. "QuickSet: multimodal interaction for distributed applications". In: *Proceedings of the Fifth ACM International Conference on Multimedia*. MULTIMEDIA '97. Seattle, Washington, USA: Association for Computing Machinery, 1997, pp. 31–40. ISBN: 0897919912. DOI: 10.1145/266180.266328.

[39] IPIP Project: ORI Community and Evaluation Services. *Interpreting Individual IPIP Scale Scores*. 2023. URL: https://ipip.ori.org/InterpretingIndividualIPIPScaleScores.htm.

[40] Kristin A Cook and James J Thomas. “Illuminating the Path: The Research and Development Agenda for Visual Analytics”. In: (May 2005). URL: https://www.osti.gov/biblio/912515.

[41] International Data Corporation. *Tablet Shipments Returns to Pre-Pandemic Levels, According to IDC*. 2023. URL: https://www.idc.com/getdoc.jsp?containerId=prUS50644623.

[42] Paul T Costa and Robert R McCrae. “Manual for the revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI)”. In: *Psychological Assessment Resources* 396 (1992).

[43] Global Dynamics of Social Policy CRC 1342. *Global Welfare State Information System*. https://wesis.org/. Accessed: 2023-12-22. 2023.

[44] Datawrapper GmbH. *Datawrapper*. https://www.datawrapper.de. Accessed: 07/2020. 2021. URL: https://www.datawrapper.de.

[45] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. “The Participatory Turn in AI Design: Theoretical Foundations and the Current State of Practice”. In: *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. Boston: Association for Computing Machinery, 2023. ISBN: 9798400703812. DOI: 10.1145/3617694.3623261.

[46] Design Council. *The double diamond design process model*. Accessed: 12/2019. 2015. URL: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond.

[47] Colin G. DeYoung, Rachael G. Grazioplene, and Jordan B. Peterson. “From madness to genius: The Openness/Intellect trait domain as a paradoxical simplex”. In: *Journal of Research in Personality* 46.1 (2012), pp. 63–78. ISSN: 0092-6566. DOI: 10.1016/j.jrp.2011.12.003.

[48] Evanthia Dimara and Charles Perin. “What is Interaction for Data Visualization?” In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), pp. 119–129. DOI: 10.1109/TVCG.2019.2934283.

[49] Mollie Dollinger and Jason M. Lodge. "Co-Creation Strategies for Learning Analytics". In: *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. LAK 18. Sydney, New South Wales, Australia: Association for Computing Machinery, 2018, pp. 97–101. ISBN: 9781450364003. DOI: 10.1145/3170358.3170372.

[50] Marian Dörk and David Monteyne. "Urban co-creation: envisioning new digital tools for activism and experimentation in the city". In: *Proceedings of the CHI Conference*. 2011, pp. 7–12.

[51] Jakub Dostal, Uta Hinrichs, Per Ola Kristensson, and Aaron Quigley. "SpiderEyes: Designing Attention- and Proximity-Aware Collaborative Interfaces for Wall-Sized Displays". In: *Proc. IUI*. Haifa, Israel: ACM, 2014, pp. 143–152. ISBN: 9781450321846. DOI: 10.1145/2557500.2557541.

[52] Steven P. Dow, Alana Glassco, Jonathan Kass, Melissa Schwarz, Daniel L. Schwartz, and Scott R. Klemmer. "Parallel Prototyping Leads to Better Design Results, More Divergence, and Increased Self-Efficacy". In: *ACM Trans. Comput.-Hum. Interact.* 17.4 (Dec. 2011). ISSN: 1073-0516. DOI: 10.1145/1879831.1879836.

[53] Steven M. Drucker, Danyel Fisher, Ramik Sadana, Jessica Herron, and m.c. schraefel m.c. "TouchViz: A Case Study Comparing Two Interfaces for Data Analytics on Tablets". In: *Proc. CHI*. Paris, France: ACM, 2013, pp. 2301–2310. ISBN: 9781450318990. DOI: 10.1145/2470654.2481318.

[54] Ian K. Duncan, Shi Tingsheng, Simon T. Perrault, and Michael T. Gastner. "Task-Based Effectiveness of Interactive Contiguous Area Cartograms". In: *IEEE Transactions on Visualization and Computer Graphics* (2020), pp. 1–1. DOI: 10.1109/TVCG.2020.3041745.

[55] P. Eichmann, D. Edge, N. Evans, B. Lee, M. Brehmer, and C. White. "Orchard: Exploring Multivariate Heterogeneous Networks on Mobile Phones". In: *Computer Graphics Forum* 39.3 (2020), pp. 115–126. DOI: 10.1111/cgf.13967.

[56] Stef van den Elzen and Jarke J. van Wijk. "Small Multiples, Large Singles: A New Approach for Visual Data Exploration". In: *Computer Graphics Forum* 32.3pt2 (2013), pp. 191–200. DOI: 10.1111/cgf.12106.

[57] Geraldine Fitzpatrick and Gunnar Ellingsen. "A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas". In: *Computer Supported Cooperative Work (CSCW)* 22 (2013), pp. 609–665. DOI: 10.1007/s10606-012-9168-0.

[58] Gapminder Foundation. *Free Data via Gapminder.org, CC-BY LICENSE*. July 2022. URL: https://www.gapminder.org/data/.

[59] Leo D. Frishberg. "Interactive Sparklines: A Dynamic Display of Quantitative Information". In: *CHI 11 Extended Abstracts on Human Factors in Computing Systems*. CHI EA 11. Vancouver, BC, Canada: Association for Computing Machinery, 2011, pp. 589–604. ISBN: 9781450302685. DOI: 10.1145/1979742.1979655.

[60] Hugo Fuks, Heloisa Moura, Debora Cardador, Katia Vega, Wallace Ugulino, and Marcos Barbato. "Collaborative Museums: An Approach to Co-Design". In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. CSCW 12. Seattle, Washington, USA: Association for Computing Machinery, 2012, pp. 681–684. ISBN: 9781450310864. DOI: 10.1145/2145204.2145307.

[61] S. Goodwin, J. Dykes, S. Jones, I. Dillingham, G. Dove, A. Duffy, A. Kachkaev, A. Slingsby, and J. Wood. "Creative User-Centered Visualization Design for Energy Analysts and Modelers". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), pp. 2516–2525. ISSN: 2160-9306. DOI: 10.1109/TVCG.2013.145.

[62] Google. *Google Maps*. https://www.google.com/maps. Accessed: 07/2020. 2021. URL: https://www.google.com/maps.

[63] Arno P. J. Gourdol, Laurence Nigay, Daniel Salber, and Joëlle Coutaz. "Two Case Studies of Software Architecture for Multimodal Interactive Systems: VoicePaint and a Voice-enabled Graphical Notebook". In: *Proceedings of the IFIP TC2/WG2.7 Working Conference on Engineering for Human-Computer Interaction*. NLD: North-Holland Publishing Co., 1992, pp. 271–284. ISBN: 0444899049.

[64] Tear Marie Green and Brian Fisher. "Towards the Personal Equation of Interaction: The impact of personality factors on visual analytics interface interaction". In: *2010 IEEE Symposium on Visual Analytics Science and Technology*. Salt Lake City, UT, USA: IEEE, 2010, pp. 203–210. DOI: 10.1109/VAST.2010.5653587.

[65] Georges Grinstein, Theresa O'Connell, Sharon Laskowski, Catherine Plaisant, Jean Scholtz, and Mark Whiting. "VAST 2006 Contest - A Tale of Alderwood". In: *2006 IEEE Symposium On Visual Analytics Science And Technology*. Baltimore, MD, USA: IEEE, 2006, pp. 215–216. DOI: 10.1109/VAST.2006.261420.

[66] Ursula S Grissemann and Nicola E Stokburger-Sauer. "Customer co-creation of travel services: The role of company support and customer satisfaction with the co-creation performance". In: *Tourism management* 33.6 (2012), pp. 1483–1492.

[67] The World Bank Group. *World Bank Open Data*. https://data.worldbank.org/indicator. 2022. URL: https://data.worldbank.org/.

[68] Jonathan Grudin. "Computer-Supported Cooperative Work: History and Focus". In: *Computer* 27.5 (1994), pp. 19–26. DOI: 10.1109/2.291294.

[69] François Guimbretière, Maureen Stone, and Terry Winograd. "Fluid Interaction with High-Resolution Wall-Size Displays". In: *Proceedings of the 14th Annu. ACM Symp. on User Interface Software and Technol.* Orlando, Florida: ACM, 2001, pp. 21–30. ISBN: 158113438X. DOI: 10.1145/502348.502353.

[70] N. Henry and J. Fekete. "MatrixExplorer: a Dual-Representation System to Explore Social Networks". In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 677–684. ISSN: 2160-9306. DOI: 10.1109/TVCG.2006.160.

[71] Nathalie Henry and Jean-daniel Fekete. "MatrixExplorer: a Dual-Representation System to Explore Social Networks". In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 677–684. DOI: 10.1109/TVCG.2006.160.

[72] Sebastian Herholz, Lewis L Chuang, Thomas G Tanner, Heinrich H Bülthoff, and Roland W Fleming. "Libgaze: Real-time gaze-tracking of freely moving observers for wall-sized displays". In: *13th International Fall Workshop on Vision, Modeling, and Visualization*. Akademische Verlags-Gesellschaft AKA. 2008, pp. 101–110.

[73] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. "Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions". In: *Proc. CHI*. Toronto, Ontario, Canada: ACM, 2014, pp. 1063–1072. ISBN: 9781450324731. DOI: 10.1145/2556288.2557130.

[74] Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddington, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. "Pen + Touch = New Tools". In: *Proc. Annu. ACM Symp. on User Interface Software and Technol.* New York: ACM, 2010, pp. 27–36. ISBN: 9781450302715. DOI: 10.1145/1866029.1866036. URL: https://doi.org/10.1145/1866029.1866036.

[75] Uta Hinrichs and Sheelagh Carpendale. "Gestures in the Wild: Studying Multi-Touch Gesture Sequences on Interactive Tabletop Exhibits". In: *Proc. CHI*. Vancouver, BC, Canada: ACM, 2011, pp. 3023–3032. ISBN: 9781450302289. DOI: 10.1145/1978942.1979391.

[76] Enamul Hoque, Vidya Setlur, Melanie Tory, and Isaac Dykeman. "Applying Pragmatics Principles for Interaction with Visual Analytics". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), pp. 309–318. DOI: 10.1109/TVCG.2017.2744684.

[77] Tom Horak, Sriram Karthik Badam, Niklas Elmqvist, and Raimund Dachselt. "When David Meets Goliath: Combining Smartwatches with a Large Vertical Display for Visual Data Exploration". In: *Proc. CHI*. New York: ACM, 2018, pp. 1–13. ISBN: 9781450356206. DOI: 10.1145/3173574.3173593.

[78] Kasper Hornbæk and Antti Oulasvirta. "What Is Interaction?" In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 5040–5052. ISBN: 9781450346559. DOI: 10.1145/3025453.3025765.

[79] Picovoice Inc. *Train, develop and deploy custom voice features - Picovoice*. July 2023. URL: https://picovoice.ai/.

[80] Interaction Design Foundation - IxDF. *What is Co-Creation?* https://www.interaction-design.org/literature/topics/co-creation. [Online; accessed 29-February-2024]. 2021.

[81] Petra Isenberg, Anastasia Bezerianos, Nathalie Henry, Sheelagh Carpendale, and Jean-Daniel Fekete. "CoCoNutTrix: Collaborative Retrofitting for Information Visualization". In: *IEEE Comput. Graphics Appl.: Special Issue on Collaborative Visualization* 29.5 (Sept. 2009), pp. 44–57. DOI: 10.1109/MCG.2009.78.

[82] Petra Isenberg, Niklas Elmqvist, Jean Scholtz, Daniel Cernea, Kwan-Liu Ma, and Hans Hagen. "Collaborative Visualization: Definition, Challenges, and Research Agenda". In: *Inf. Visualization Journal (IVS), Special Issue on Information Visualization: State of the Field and New Research Directions* 10.4 (Oct. 2011), pp. 310–326. DOI: 10.1177/1473871611412817.

[83] Petra Isenberg, Niklas Elmqvist, Jean Scholtz, Daniel Cernea, Kwan-Liu Ma, and Hans Hagen. "Collaborative Visualization: Definition, Challenges, and Research Agenda". In: *Information Visualization Journal (IVS), Special Issue on Information Visualization: State of the Field and New Research Directions* 10.4 (Oct. 2011), pp. 310–326. DOI: 10.1177/1473871611412817.

[84] Petra Isenberg and Danyel Fisher. "Collaborative Brushing and Linking for Co-located Visual Analytics of Document Collections". In: *Comput. Graphics Forum* 28.3 (June 2009), pp. 1031–1038. DOI: 10.1111/j.1467-8659.2009.01444.x.

[85] Petra Isenberg, Danyel Fisher, Meredith Ringel Morris, Kori Inkpen, and Mary Czerwinski. "An Exploratory Study of Co-located Collaborative Visual Analytics around a Tabletop Display". In: *Proceedings of Visual Analytics Science and Technology (VAST)*. Los Alamitos: IEEE, 2010, pp. 179–186. DOI: 10.1109/VAST.2010.5652880.

[86] Petra Isenberg, Florian Heimerl, Steffen Koch, Tobias Isenberg, Panpan Xu, Chad Stolper, Michael Sedlmair, Jian Chen, Torsten Möller, and John Stasko. "vispubdata.org: A Metadata Collection about IEEE Visualization (VIS) Publications". In: *IEEE Transactions on Visualization and Computer Graphics* 23.9 (Sept. 2017), pp. 2199–2206. DOI: https://doi.org/10.1109/TVCG.

2016.2615308. URL: https://tobias.isenberg.cc/VideosAndDemos/Isenberg2017VMC.

[87] Alaul Islam, Ranjini Aravind, Tanja Blascheck, Anastasia Bezerianos, and Petra Isenberg. "Preferences and Effectiveness of Sleep Data Visualizations for Smartwatches and Fitness Bands". In: *Proc. CHI*. New Orleans: ACM, 2022. ISBN: 9781450391573. DOI: 10.1145/3491102.3501921.

[88] Mikkel R. Jakobsen and Kasper Hornbæk. "Up Close and Personal: Collaborative Work on a High-Resolution Multitouch Wall Display". In: *ACM Trans. Comput.-Hum. Interact.* 21.2 (Feb. 2014). ISSN: 1073-0516. DOI: 10.1145/2576099.

[89] Mikkel R. Jakobsen, Yonas Sahlemariam Haile, Søren Knudsen, and Kasper Hornbæk. "Information Visualization and Proxemics: Design Opportunities and Empirical Findings". In: *IEEE Trans. Visual Comput. Graphics* 19.12 (2013), pp. 2386–2395. DOI: 10.1109/TVCG.2013.166.

[90] Raphaël James, Anastasia Bezerianos, and Olivier Chapuis. "Evaluating the Extension of Wall Displays with AR for Collaborative Work". In: *Proc. CHI*. Hamburg, Germany: Association for Computing Machinery, 2023. ISBN: 9781450394215. DOI: 10.1145/3544548.3580752.

[91] Stefan Jänicke, Pawandeep Kaur, Pawel Kuzmicki, and Johanna Schmidt. "Participatory Visualization Design as an Approach to Minimize the Gap between Research and Application". In: *VisGap - The Gap between Visualization Research and Visualization Software*. Ed. by Christina Gillmann, Michael Krone, Guido Reina, and Thomas Wischgoll. The Eurographics Association, 2020, pp. 35–42. ISBN: 978-3-03868-125-0. DOI: 10.2312/visgap.20201108.

[92] Juliane Jarke. *Co-creating digital public services for an ageing society: Evidence for user-centric design*. Public Administration and Information Technology, 6. Springer Nature, 2021. DOI: 10.1007/978-3-030-52873-7.

[93] Juliane Jarke, Herbert Kubicek, Ulrike Gerhard, Lucas Introna, Niall Hayes, Noel Cass, Frank Berker, Frank Reins, and MobileAge Project. *Interactive Co-Creation Good Practice Guide*. https://co-creation.mobile-age.eu. Accessed: 01/2019. 2019.

[94] Anick Jesdanun. *The consumer tablet is dying; long live the business tablet*. 2016. URL: https://www.ctvnews.ca/sci-tech/the-consumer-tablet-is-dying-long-live-the-business-tablet-1.2829557.

[95] Jaemin Jo, Sehi LYi, Bongshin Lee, and Jinwook Seo. "TouchPivot: Blending WIMP & Post-WIMP Interfaces for Data Exploration on Tablet Devices". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI 17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 2660–2671. ISBN: 9781450346559. DOI: 10.1145/3025453.3025752.

[96] Robert Johansen. *Groupware: Computer Support for Business Teams*. Series in communication technology and society. United States: The Free Press, Nov. 1988. ISBN: 9780029164914. URL: https://books.google.de/books?id=TIcTAQAAMAAJ.

[97] M. Kahng, P. Y. Andrews, A. Kalro, and D. H. Chau. "ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), pp. 88–97. ISSN: 2160-9306. DOI: 10.1109/TVCG.2017.2744718.

[98] D.A. Keim. "Information visualization and visual data mining". In: *IEEE Transactions on Visualization and Computer Graphics* 8.1 (2002), pp. 1–8. DOI: 10.1109/2945.981847.

[99] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. "Visual Analytics: Definition, Process, and Challenges". In: *Information Visualization: Human-Centered Issues and Perspectives*. Ed. by Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 154–175. ISBN: 978-3-540-70956-5. DOI: 10.1007/978-3-540-70956-5_7.

[100] E. Kerzner, S. Goodwin, J. Dykes, S. Jones, and M. Meyer. "A Framework for Creative Visualization-Opportunities Workshops". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019), pp. 748–758. ISSN: 2160-9306. DOI: 10.1109/TVCG.2018.2865241.

[101] Young-Ho Kim, Bongshin Lee, Arjun Srinivasan, and Eun Kyoung Choe. "Data@Hand: Fostering Visual Exploration of Personal Data On Smartphones Leveraging Speech and Touch Interaction". In: *Proc. CHI*. Yokohama, Japan: ACM, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445421.

[102] U. Kister, K. Klamka, C. Tominski, and R. Dachselt. "GraSp: Combining Spatially-aware Mobile Devices and a Display Wall for Graph Visualization and Interaction". In: *Comput. Graphics Forum* 36.3 (2017), pp. 503–514. DOI: 10.1111/cgf.13206.

[103] Ulrike Kister, Patrick Reipschläger, Fabrice Matulic, and Raimund Dachselt. "BodyLenses: Embodied Magic Lenses and Personal Territories for Wall Displays". In: *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. ITS 15. Madeira, Portugal: Association for Computing Machinery, 2015, pp. 117–126. ISBN: 9781450338998. DOI: 10.1145/2817721.2817726.

[104] L. C. Koh, A. Slingsby, J. Dykes, and T. S. Kam. "Developing and Applying a User-Centered Model for the Design and Implementation of Information Visualization Tools". In: *2011 15th International Conference on Information Visualisation*. July 2011, pp. 90–95. DOI: 10.1109/IV.2011.32.

[105] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. "Empirical Studies in Information Visualization: Seven Scenarios". In: *IEEE Trans. Visual Comput. Graphics* 18.9 (2012), pp. 1520–1536. DOI: 10.1109/TVCG.2011.279.

[106] J. Landstorfer, I. Herrmann, J. Stange, M. Dörk, and R. Wettach. "Weaving a carpet from log entries: A network security visualization built with co-creation". In: *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Oct. 2014, pp. 73–82. DOI: 10.1109/VAST.2014.7042483.

[107] Johannes Landstorfer, Ivo Herrmann, Jan-Erik Stange, Marian Dörk, and Reto Wettach. "Weaving a carpet from log entries: A network security visualization built with co-creation". In: *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 2014, pp. 73–82. DOI: 10.1109/VAST.2014.7042483.

[108] Ricardo Langner, Ulrike Kister, and Raimund Dachselt. "Multiple Coordinated Views at Large Displays for Multiple Users: Empirical Findings on User Behavior, Movements, and Distances". In: *IEEE Trans. Visual Comput. Graphics* 25.1 (2019), pp. 608–618. DOI: 10.1109/TVCG.2018.2865235.

[109] Bongshin Lee, Petra Isenberg, Nathalie Henry Riche, and Sheelagh Carpendale. "Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2689–2698. ISSN: 1941-0506. DOI: 10.1109/TVCG.2012.204.

[110] Bongshin Lee, Greg Smith, Nathalie Henry Riche, Amy Karlson, and Sheelagh Carpendale. "SketchInsight: Natural data exploration on interactive whiteboards leveraging pen and touch interaction". In: *Proc. PacificVis*. 2015, pp. 199–206. DOI: 10.1109/PACIFICVIS.2015.7156378.

[111] Bongshin Lee, Arjun Srinivasan, Petra Isenberg, and John Stasko. "Post-WIMP Interaction for Information Visualization". In: *Foundations and Trendső in HumanComputer Interaction* 14.1 (2021), pp. 1–95. ISSN: 1551-3955. DOI: 10.1561/1100000081.

[112] Jung-Joo Lee, Miia Jaatinen, Anna Salmi, Tuuli Mattelmäki, Riitta Smeds, Mari Holopainen, et al. "Design choices framework for co-creation projects". In: *International Journal of Design* 12.2 (Aug. 2018), pp. 15–31. URL: http://www.ijdesign.org/index.php/IJDesign/article/view/2782/814.

[113] Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, and Eric Lecolinet. "CoReach: Cooperative Gestures for Data Manipulation on Wall-Sized Displays". In: *Proc. CHI*. Denver: ACM, 2017, pp. 6730–6741. ISBN: 9781450346559. DOI: 10.1145/3025453.3025594.

[114] Zhengliang Liu, R. Jordan Crouser, and Alvitta Ottley. "Survey on Individual Differences in Visualization". In: *Computer Graphics Forum* 39.3 (2020), pp. 693–712. DOI: 10.1111/cgf.14033.

[115] D. Lloyd and J. Dykes. "Human-Centered Approaches in Geovisualization Design: Investigating Multiple Methods Through a Long-Term Case Study". In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011), pp. 2498–2507. ISSN: 2160-9306. DOI: 10.1109/TVCG.2011.209.

[116] Alan M MacEachren, Guoray Cai, Rajeev Sharma, Ingmar Rauschert, Isaac Brewer, Levent Bolelli, B Shaparenko, Sven Fuhrmann, and Hongmei Wang. “Enabling collaborative geoinformation access and decisionmaking through a natural, multimodal interface”. In: *International Journal of Geographical Information Science* 19.3 (2005), pp. 293–317. DOI: 10.1080/13658810412331280158.

[117] Narges Mahyar, Ali Sarvghad, and Melanie Tory. “Note-taking in co-located collaborative visual analytics: Analysis of an observational study”. In: *Information Visualization* 11.3 (2012), pp. 190–204. DOI: 10.1177/1473871611433713.

[118] Jessica L. Maples-Keller, Rachel L. Williamson, Chelsea E. Sleep, Nathan T. Carter, W. Keith Campbell, and Joshua D. Miller. “Using Item Response Theory to Develop a 60-Item Representation of the NEO PIR Using the International Personality Item Pool: Development of the IPIPNEO60”. In: *Journal of Personality Assessment* 101.1 (2019). PMID: 29087223, pp. 4–15. DOI: 10.1080/00223891.2017.1381968.

[119] S. Mckenna, D. Staheli, and M. Meyer. “Unlocking user-centered design methods for building cyber security visualizations”. In: *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*. Oct. 2015, pp. 1–8. DOI: 10.1109/VIZSEC.2015.7312771.

[120] Ulises A Mejías and Nick Couldry. “Datafication”. In: *Internet Policy Review* 8.4 (2019). DOI: 10.14763/2019.4.1428.

[121] Miriah Meyer and Jason Dykes. “Criteria for Rigor in Visualization Design Study”. In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), pp. 87–97. DOI: 10.1109/TVCG.2019.2934539.

[122] Christophe Mignot, Claude Valot, and Noelle Carbonell. “An experimental study of future natural multimodal human-computer interaction”. In: *INTERACT and CHI Conference Companion*. 1993, pp. 67–68.

[123] Val Mitchell, Tracy Ross, Andrew May, Ruth Sims, and Christopher Parker. “Empirical investigation of the impact of using co-design methods when generating proposals for sustainable travel solutions”. In: *CoDesign* 12.4 (2016), pp. 205–220. DOI: 10.1080/15710882.2015.1091894.

[124] Gabriela Molina León. "Advancing Inclusive Design in Multiple Dimensions". In: *CHI Workshop on Inclusive Design of CUIs Across Modalities and Mobilities*. 2023. arXiv: 2303.16790 [cs.HC].

[125] Gabriela Molina León, Anastasia Bezerianos, Olivier Gladin, and Petra Isenberg. *Talk to the Wall: The Role of Speech Interaction in Collaborative Work*. To submit. 2024.

[126] Gabriela Molina León and Andreas Breiter. "Co-creating Visualizations: A First Evaluation with Social Science Researchers". In: *Computer Graphics Forum* 39.3 (2020), pp. 291–302. DOI: 10.1111/cgf.13981.

[127] Gabriela Molina León, Petra Isenberg, and Andreas Breiter. "Eliciting Multimodal and Collaborative Interactions for Data Exploration on Large Vertical Displays". In: *IEEE Transactions on Visualization and Computer Graphics* 30.2 (2024), pp. 1624–1637. ISSN: 1941-0506. DOI: 10.1109/TVCG.2023.3323150.

[128] Gabriela Molina León, Petra Isenberg, and Andreas Breiter. "Talking to Data Visualizations: Opportunities and Challenges". In: *VIS Workshop on Multimodal Experiences for Remote Communication Around Data Online*. 2023. arXiv: 2309.09781 [cs.HC].

[129] Gabriela Molina León, Michael Lischka, and Andreas Breiter. "Mapping the Global South: Equal-Area Projections for Choropleth Maps". In: *2020 IEEE Visualization Conference (VIS)*. 2020, pp. 91–95. DOI: 10.1109/VIS47514.2020.00025.

[130] Gabriela Molina León, Michael Lischka, Wei Luo, and Andreas Breiter. "Mobile and Multimodal? A Comparative Evaluation of Interactive Workplaces for Visual Data Exploration". In: *Computer Graphics Forum* 41.3 (2022), pp. 417–428. DOI: 10.1111/cgf.14551.

[131] Gabriela Molina León, Gabriella Skitalinskaya, Nils Düpont, Jonas Klaff, Anton Schlegel, Hendrik Heuer, and Andreas Breiter. "Co-Creating a Research Data Infrastructure with Social Policy Researchers". In: *Proceedings of 20th European Conference on Computer-Supported Cooperative Work*. European Society for Socially Embedded Technologies (EUSSET), 2022. DOI: 10.48340/ecscw2022_p03.

[132] Meredith Ringel Morris. “Web on the Wall: Insights from a Multimodal Interaction Elicitation Study”. In: *Proc. ITS*. Cambridge: ACM, 2012, pp. 95–104. ISBN: 9781450312097. DOI: 10.1145/2396636.2396651.

[133] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel m. c., and Jacob O. Wobbrock. “Reducing Legacy Bias in Gesture Elicitation Studies”. In: *Interactions* 21.3 (May 2014), pp. 40–45. ISSN: 1072-5520. DOI: 10.1145/2591689. URL: https://doi.org/10.1145/2591689.

[134] Meredith Ringel Morris, Anqi Huang, Andreas Paepcke, and Terry Winograd. “Cooperative Gestures: Multi-User Gestural Interactions for Co-Located Groupware”. In: *Proc. CHI*. Montréal, Québec, Canada: ACM, 2006, pp. 1201–1210. ISBN: 1595933727. DOI: 10.1145/1124772.1124952.

[135] Mozilla and individual contributors. *Web Speech API*. 2019. URL: https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API.

[136] Tamara Munzner. *Visualization Analysis and Design*. A.K. Peters visualization series. A K Peters, 2014. ISBN: 978-1-466-50891-0. URL: http://www.cs.ubc.ca/%5C%7Etmm/vadbook/.

[137] Brad A. Myers, Rishi Bhatnagar, Jeffrey Nichols, Choon Hong Peck, Dave Kong, Robert Miller, and A. Chris Long. “Interacting at a distance: measuring the performance of laser pointers and other devices”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Minneapolis, Minnesota, USA: Association for Computing Machinery, 2002, pp. 33–40. ISBN: 1581134533. DOI: 10.1145/503376.503383.

[138] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. “Memorability of Pre-Designed and User-Defined Gesture Sets”. In: *Proc. CHI*. Paris, France: ACM, 2013, pp. 1099–1108. ISBN: 9781450318990. DOI: 10.1145/2470654.2466142.

[139] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. “Mid-Air Pan-and-Zoom on Wall-Sized Displays”. In: *Proc. CHI*. Vancouver, BC, Canada: ACM, 2011, pp. 177–186. ISBN: 9781450302289. DOI: 10.1145/1978942.1978969.

[140] Michael Nebeling, Alexander Huber, David Ott, and Moira C. Norrie. "Web on the Wall Reloaded: Implementation, Replication and Refinement of User-Defined Interaction Sets". In: *Proc. ITS*. Dresden, Germany: ACM, 2014, pp. 15–24. ISBN: 9781450325875. DOI: 10.1145/2669485.2669497.

[141] Thomas Neumayr, Hans-Christian Jetter, Mirjam Augstein, Judith Friedl, and Thomas Luger. "Domino: A Descriptive Framework for Hybrid Collaboration and Coupling Styles in Partially Distributed Teams". In: *Proc. ACM Hum.-Comput. Interact.* 2.CSCW (Nov. 2018). DOI: 10.1145/3274397.

[142] Nate Nichols and Homer Wang. *What is Tableau AI?* 2023. URL: https://www.tableau.com/blog/what-is-tableau-ai.

[143] Jakob Nielsen. "Noncommand user interfaces". In: *Communications of the ACM* 36.4 (1993), pp. 83–99.

[144] OpenAI. *ChatGPT can now see, hear, and speak*. 2023. URL: https://openai.com/blog/chatgpt-can-now-see-hear-and-speak.

[145] Organisation for Economic Co-operation and Development. *OECD Data*. OECD Data Portal. 2021. URL: https://data.oecd.org/.

[146] Esteban Ortiz-Ospina, Sandra Tzvetkova, and Max Roser. "Womens employment". In: *Our World in Data* (2018). https://ourworldindata.org/female-labor-supply.

[147] Sharon Oviatt. "Mulitmodal Interactive Maps: Designing for Human Performance". In: *HumanComputer Interaction* 12.1-2 (1997), pp. 93–129. URL: https://www.tandfonline.com/doi/abs/10.1080/07370024.1997.9667241.

[148] Sharon Oviatt. "Ten Myths of Multimodal Interaction". In: *Commun. ACM* 42.11 (Nov. 1999), pp. 74–81. ISSN: 0001-0782. DOI: 10.1145/319382.319398.

[149] Sharon Oviatt, Antonella DeAngeli, and Karen Kuhn. "Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction". In: *Proc. CHI*. Atlanta, Georgia, USA: Association for Computing Machinery, 1997, pp. 415–422. ISBN: 0897918029. DOI: 10.1145/258549.258821.

[150] Vanessa Peña-Araya, Anastasia Bezerianos, and Emmanuel Pietriga. "A Comparison of Geographical Propagation Visualizations". In: *CHI'20-38th SIGCHI conference on Human Factors in computing systems*. Association for Computing Machinery, 2020, pp. 223–1. DOI: 10.1145/3313831.3376350.

[151] Madhawa Perera, Tom Gedeon, Matt Adcock, and Armin Haller. "Towards Self-Guided Remote User Studies - Feasibility of Gesture Elicitation using Immersive Virtual Reality". In: *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2021, pp. 2576–2583. DOI: 10.1109/SMC52423.2021.9658673.

[152] Donna J. Peuquet. "It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems". In: *Annals of the Association of American Geographers* 84.3 (1994), pp. 441–461. DOI: 10.1111/j.1467-8306.1994.tb01869.x.

[153] Ken Pfeuffer, Ken Hinckley, Michel Pahud, and Bill Buxton. "Thumb + Pen Interaction on Tablets". In: *Proc. CHI*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 3254–3266. ISBN: 9781450346559. DOI: 10.1145/3025453.3025567.

[154] Plotly. *Plotly Express in Python*. 2022. URL: https://plotly.com/python/plotly-express/.

[155] Arnaud Prouzeau, Anastasia Bezerianos, and Olivier Chapuis. "Evaluating Multi-User Selection for Exploring Graph Topology on Wall-Displays". In: *IEEE Trans. Visual Comput. Graphics* 23.8 (2017), pp. 1936–1951. DOI: 10.1109/TVCG.2016.2592906.

[156] Yan Qu and Derek L Hansen. "Building shared understanding in collaborative sensemaking". In: *Proc. CHI 2008 Sensemaking Workshop*. 2008.

[157] Patrick Reipschlager, Tamara Flemisch, and Raimund Dachselt. "Personal Augmented Reality for Information Visualization on Large Interactive Displays". In: *IEEE Trans. Visual Comput. Graphics* 27.2 (2021), pp. 1182–1192. DOI: 10.1109/TVCG.2020.3030460.

[158] Patrick Riehmann, Gabriela Molina León, Joshua Reibert, Florian Echtler, and Bernd Froehlich. "Short-Contact Touch-Manipulation of Scatterplot Matrices on Wall Displays". In: *Comput. Graphics Forum* 39.3 (2020), pp. 265–276. DOI: 10.1111/cgf.13979.

[159] George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko. "Effectiveness of Animation in Trend Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (Nov. 2008), pp. 1325–1332. ISSN: 2160-9306. DOI: 10.1109/TVCG.2008.125.

[160] Yvonne Rogers and Siân Lindley. "Collaborating around vertical and horizontal large interactive displays: which way is best?" In: *Interact. Comput.* 16.6 (Sept. 2004), pp. 1133–1152. ISSN: 0953-5438. DOI: 10.1016/j.intcom.2004.07.008.

[161] Max Roser, Hannah Ritchie, and Bernadeta Dadonaite. "Child and Infant Mortality". In: *Our World in Data* (2013). URL: https://ourworldindata.org/child-mortality.

[162] Ramik Sadana and John Stasko. "Designing and Implementing an Interactive Scatterplot Visualization for a Tablet Computer". In: *Proc. AVI*. Como, Italy: ACM, 2014, pp. 265–272. ISBN: 9781450327756. DOI: 10.1145/2598153.2598163.

[163] Ramik Sadana and John Stasko. "Designing Multiple Coordinated Visualizations for Tablets". In: *Comput. Graphics Forum* 35.3 (2016), pp. 261–270. DOI: 10.1111/cgf.12902.

[164] Ayshwarya Saktheeswaran, Arjun Srinivasan, and John Stasko. "Touch? Speech? or Touch and Speech? Investigating Multimodal Interaction for Visual Network Exploration and Analysis". In: *IEEE Trans. Visual Comput. Graphics* 26.6 (2020), pp. 2168–2179. DOI: 10.1109/TVCG.2020.2970512.

[165] EBN Sanders. *An evolving map of design practice and design research. interactions 15, 6 (Nov. 2008), 13-17*. 2008.

[166] Elizabeth B.-N. Sanders and Pieter Jan Stappers. "Co-creation and the new landscapes of design". In: *CoDesign* 4.1 (2008), pp. 5–18. DOI: 10.1080/15710880701875068.

[167] Kai Sassenberg and Gordon B. Moskowitz. "Dont stereotype, think different! Overcoming automatic stereotype activation by mindset priming". In: *J. Exp. Social Psychol.* 41.5 (2005), pp. 506–514. ISSN: 0022-1031. DOI: 10.1016/j.jesp.2004.10.002.

[168] Bojan Savric, Tom Patterson, and Bernhard Jenny. “The Equal Earth map projection”. In: *International Journal of Geographical Information Science* 33.3 (2019), pp. 454–465. DOI: 10.1080/13658816.2018.1504949.

[169] Carina Schmitt, Hanna Lierse, Herbert Obinger, and Laura Seelkopf. “The Global Emergence of Social Protection: Explaining Social Security Legislation 18202013*”. In: *Politics & Society* 43.4 (2015), pp. 503–524. DOI: 10.1177/0032329215602892.

[170] Douglas Schuler and Aki Namioka. *Participatory design: Principles and practices*. CRC Press, 1993.

[171] Stacey D. Scott, M. Sheelagh T. Carpendale, and Kori Inkpen. “Territoriality in collaborative tabletop workspaces”. In: *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*. Chicago, Illinois, USA: Association for Computing Machinery, 2004, pp. 294–303. ISBN: 1581138105. DOI: 10.1145/1031607.1031655.

[172] M. Sedlmair, M. Meyer, and T. Munzner. “Design Study Methodology: Reflections from the Trenches and the Stacks”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (Dec. 2012), pp. 2431–2440. ISSN: 2160-9306. DOI: 10.1109/TVCG.2012.213.

[173] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. “Eviza: A Natural Language Interface for Visual Analysis”. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST ’16. Tokyo, Japan: Association for Computing Machinery, 2016, pp. 365–377. ISBN: 9781450341899. DOI: 10.1145/2984511.2984588.

[174] B. Shneiderman. “The eyes have it: a task by data type taxonomy for information visualizations”. In: *Proceedings 1996 IEEE Symposium on Visual Languages*. Sept. 1996, pp. 336–343. DOI: 10.1109/VL.1996.545307.

[175] Jesper Simonsen and Toni Robertson. *Routledge International Handbook of Participatory Design*. Routledge, 2012.

[176] Arjun Srinivasan, Bongshin Lee, Nathalie Henry Riche, Steven M. Drucker, and Ken Hinckley. “InChorus: Designing Consistent Multimodal Interactions for Data Visualization on Tablet Devices”. In: *Proc. CHI*. New York: ACM, 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376782.

[177] Arjun Srinivasan, Bongshin Lee, and John Stasko. "Interweaving Multimodal Interaction With Flexible Unit Visualizations for Data Exploration". In: *IEEE Trans. Visual Comput. Graphics* 27.8 (2021), pp. 3519–3533. DOI: 10.1109/TVCG.2020.2978050.

[178] Arjun Srinivasan and John Stasko. "Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks". In: *IEEE Trans. Visual Comput. Graphics* 24.1 (Jan. 2018), pp. 511–521. ISSN: 1941-0506. DOI: 10.1109/TVCG.2017.2745219.

[179] Marc Steen, Menno Manschot, and Nicole De Koning. "Benefits of co-design in service design projects". In: *International Journal of Design* 5.2 (2011). URL: http://www.ijdesign.org/index.php/IJDesign/article/view/890.

[180] Tableau Software. *Tableau Mobile*. https://www.tableau.com/products/mobile. Accessed: 03/2022. 2022. URL: https://www.tableau.com/products/mobile.

[181] Tableau Software. *Tableau Software*. https://www.tableau.com. Accessed: 03/2022. 2022. URL: https://www.tableau.com.

[182] Tableau YouTube Channel. *How to Revert Keep Only or Exclude on Tableau Desktop*. Accessed: 01/2024. Tableau, 2018. URL: https://www.youtube.com/watch?v=bK3MGEV0OlU&ab_channel=Tableau.

[183] Anthony Tang, Melanie Tory, Barry Po, Petra Neumann, and Sheelagh Carpendale. "Collaborative Coupling over Tabletop Displays". In: *Proc. CHI*. Montréal, Québec, Canada: ACM, 2006, pp. 1181–1190. ISBN: 1595933727. DOI: 10.1145/1124772.1124950.

[184] Christian Tominski and Heidrun Schumann. *Interactive visual data analysis*. New York: CRC Press, 2020. ISBN: 9781315152707. DOI: 10.1201/9781315152707.

[185] Theophanis Tsandilas. "Fallacies of Agreement: A Critical Review of Consensus Assessment Methods for Gesture Elicitation". In: *ACM Trans. Comput.-Hum. Interact.* 25.3 (June 2018). ISSN: 1073-0516. DOI: 10.1145/3182168.

[186] Theophanis Tsandilas and Pierre Dragicevic. *Accounting for Chance Agreement in Gesture Elicitation Studies*. Research Report 1584. LRI - CNRS, University Paris-Sud, Feb. 2016, p. 5. URL: https://hal.archives-ouvertes.fr/hal-01267288.

[187] Edward Tse, Saul Greenberg, Chia Shen, Clifton Forlines, and Ryo Kodama. "Exploring True Multi-User Multimodal Interaction over a Digital Table". In: *Proceedings of the 7th ACM Conference on Designing Interactive Systems*. Cape Town, South Africa: ACM, 2008, pp. 109–118. ISBN: 9781605580029. DOI: 10.1145/1394445.1394457.

[188] Wallace C. Ugulino and Hugo Fuks. "Prototyping Wearables for Supporting Cognitive Mapping by the Blind: Lessons from Co-Creation Workshops". In: *Proceedings of the 2015 Workshop on Wearable Systems and Applications*. WearSys 15. Florence, Italy: Association for Computing Machinery, 2015, pp. 39–44. ISBN: 9781450335003. DOI: 10.1145/2753509.2753515.

[189] Radu-Daniel Vatavu. "The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies". In: *Proc. CHI*. New York: ACM, 2019, pp. 1–13. ISBN: 9781450359702. DOI: 10.1145/3290605.3300454.

[190] Radu-Daniel Vatavu and Jacob O. Wobbrock. "Clarifying Agreement Calculations and Analysis for End-User Elicitation Studies". In: *ACM Trans. Comput.-Hum. Interact.* 29.1 (Jan. 2022). ISSN: 1073-0516. DOI: 10.1145/3476101.

[191] Radu-Daniel Vatavu and Jacob O. Wobbrock. "Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit". In: *Proc. CHI*. Seoul, Republic of Korea: ACM, 2015, pp. 1325–1334. ISBN: 9781450331456. DOI: 10.1145/2702123.2702223.

[192] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. "A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?" In: *Proc. DIS*. Eindhoven, Netherlands: ACM, 2020, pp. 855–872. ISBN: 9781450369749. DOI: 10.1145/3357236.3395511.

[193] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Bühner, and Heinrich Hussmann. "Developing a Personality Model for Speech-Based Conversational Agents Using the Psycholexical Approach". In: *Proceedings of the 2020 CHI Conference on Hu-*

man Factors in Computing Systems. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–14. ISBN: 9781450367080. DOI: 10.1145/3313831.3376210.

[194] W. H. Voorberg, V. J. J. M. Bekkers, and L. G. Tummers. "A Systematic Review of Co-Creation and Co-Production: Embarking on the social innovation journey". In: *Public Management Review* 17.9 (2015), pp. 1333–1357. DOI: 10.1080/14719037.2014.930505.

[195] Jagoda Walny, Bongshin Lee, Paul Johns, Nathalie Henry Riche, and Sheelagh Carpendale. "Understanding Pen and Touch Interaction for Data Exploration on Interactive Whiteboards". In: *IEEE Trans. Visual Comput. Graphics* 18.12 (2012), pp. 2779–2788. DOI: 10.1109/TVCG.2012.275.

[196] Colin Ware. *Information Visualization: Perception for Design*. 2nd. San Francisco: Morgan Kaufmann Publishers Inc., 2004. DOI: 10.1016/B978-155860819-1/50001-7.

[197] Wesley Willett, Qi Lan, and Petra Isenberg. "Eliciting Multi-touch Selection Gestures for Interactive Data Graphics". In: *EuroVis - Short Papers*. Ed. by N. Elmqvist, M. Hlawitschka, and J. Kennedy. The Eurographics Association, 2014. ISBN: 978-3-905674-69-9. DOI: 10.2312/eurovisshort.20141161. URL: https://doi.org/10.2312/eurovisshort.20141161.

[198] Adam S. Williams, Jason Garcia, Fernando De Zayas, Fidel Hernandez, Julia Sharp, and Francisco R. Ortega. "The Cost of Production in Elicitation Studies and the Legacy Bias-Consensus Trade off". In: *Multimodal Technologies and Interaction* 4.4 (2020). ISSN: 2414-4088. DOI: 10.3390/mti4040088.

[199] Adam S. Williams, Jason Garcia, and Francisco Ortega. "Understanding Multimodal User Gesture and Speech Behavior for Object Manipulation in Augmented Reality Using Elicitation". In: *IEEE Trans. Visual Comput. Graphics* 26.12 (2020), pp. 3479–3489. DOI: 10.1109/TVCG.2020.3023566.

[200] Adam S. Williams and Francisco R. Ortega. *A Concise Guide to Elicitation Methodology*. 2021. DOI: 10.48550/ARXIV.2105.12865.

[201] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. "Maximizing the Guessability of Symbolic Input". In: *CHI Extended Abstracts*. Portland: ACM, 2005, pp. 1869–1872. ISBN: 1595930027. DOI: 10.1145/1056808.1057043.

[202] Jacob O. Wobbrock and Julie A. Kientz. “Research contributions in human-computer interaction”. In: *Interactions* 23.3 (Apr. 2016), pp. 38–44. ISSN: 1072-5520. DOI: 10.1145/2907069.

[203] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. “User-Defined Gestures for Surface Computing”. In: *Proc. CHI*. Boston: ACM, 2009, pp. 1083–1092. ISBN: 9781605582467. DOI: 10.1145/1518701.1518866.

[204] World Economic Forum. *Unlocking the Shared Value of Smart City Data: A Protocol for Action*. June 2022. URL: https://www.weforum.org/publications/unlocking-the-shared-value-of-smart-city-data-a-protocol-for-action/.

[205] Ji Soo Yi, Youn ah Kang, John Stasko, and J.A. Jacko. “Toward a Deeper Understanding of the Role of Interaction in Information Visualization”. In: *IEEE Trans. Visual Comput. Graphics* 13.6 (2007), pp. 1224–1231. DOI: 10.1109/TVCG.2007.70515.

[206] Amy X. Zhang, Michael Muller, and Dakuo Wang. “How do Data Science Workers Collaborate? Roles, Workflows, and Tools”. In: *Proc. ACM Hum.-Comput. Interact.* 4.CSCW1 (May 2020). DOI: 10.1145/3392826.

[207] Caroline Ziemkiewicz, Alvitta Ottley, R. Jordan Crouser, Krysta Chauncey, Sara L. Su, and Remco Chang. “Understanding Visualization by Understanding Individual Users”. In: *IEEE Computer Graphics and Applications* 32.6 (2012), pp. 88–94. DOI: 10.1109/MCG.2012.120.

[208] Caroline Ziemkiewicz, Alvitta Ottley, R. Jordan Crouser, Ashley Rye Yauilla, Sara L. Su, William Ribarsky, and Remco Chang. “How Visualization Layout Relates to Locus of Control and Other Personality Factors”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.7 (2013), pp. 1109–1121. DOI: 10.1109/TVCG.2012.180.