The Shape of Agency: Fostering Agency in Qualitative Research through Data Visualization

by

Larissa Ugaya Mazza

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Public Health and Health Sciences

Waterloo, Ontario, Canada, 2023

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Qualitative data analysis is important to the field of healthcare since it allows researchers to understand the lived experience of patients, practitioners, and everyone in between. However, qualitative data requires time and effort, which is not always available. A potential way to overcome this barrier is to use artificial intelligence as a tool to help researchers with data analysis. However, many qualitative researchers do not have the programming skills to use AI and are reluctant to lose their sense of agency when conducting research. As a potential way to bridge this gap, we explored the use of data visualizations to foster researcher agency and make using AI more accessible.

We used Design Science Research and developed a datavis tool prototype to map out how researchers perceive agency. A user centered design approach was used to design a non-functional data visualization tool with the assistance of 5 qualitative heath researchers. Two semi-structured interviews were used to facilitate the user centered design, the first to provide guidelines for the prototype and the second for testing the tool and altering any features considered confusing or lacking.

The results showed that qualitative researchers have a wide range of cognitive needs when conducting data analysis and for that, need a variety of visualizations to best accommodate their needs. Additionally, they place high importance upon choices and freedom, wanting to feel autonomy over their own research and not be replaced or hindered by AI. Despite this, participants were open to the idea of delegating tasks, so long as they could maintain the final choice on results.

Seven barriers were identified for the fostering of agency when conducting research with AI: full AI delegation, lack of transparency with results, no choice in how results are reached, excessive freedom with no guidance, lack of ability to make edits, no guidance on how a tool works, and restricted movements.

As potential solutions for these issues, five facilitators were found during the interviews. Those being: providing choices for different kinds of data visualization, explaining the AI process in simple language, the addition of co-creation tools, addition of guidance in navigation, and the ability to enable free movement.

Land Acknowledgement

Firstly, I would like to acknowledge that I am an international student who lives and works on the Territory of the Neutral, Anishinaabeg, and Haudenosaunee Peoples. I have spent the better part of the last two years at the University of Waterloo, which is situated on the Haldimand Tract, the territory granted to the Six Nations that includes six miles on each side of the Grand River.

Additionally, I acknowledge that I was born and grew up in Brazil, a colonized country and I spend the majority of my life in the territories of the Tekohá (Guarani), Mbya, and Kaingang. I acknowledge that much of my academic journey is because of my privileges, including that of growing up on land that belongs to Indigenous peoples, both here in Canada, as well as back home.

Acknowledgements

I would like to thank my supervisors Dr. Jim Wallace and Dr. Plinio Morita for providing me guidance during these last 2 years and helping he develop this thesis into something I could be proud of.

Additionally, I would like to thank my two Labs, the Ubiquitous Health Technology Laboratory, and the Games Institute for all of their support during my Masters.

I also would like to thank the University of Waterloo and MITACS for helping to fund my research.

Special thanks for all of my family and friends for keeping me sane and productive during the last 2 years.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Healthcare is a field that can benefit from the collection of data about lived experience of patients and healthcare practitioners for a variety of reasons [1]. One way that researchers can glean insights into this kind of data is by using qualitative methods of data collections and analysis [1]. Qualitative research is better suited for the analysis of rich, interconnected, and often contradictory datasets, which makes it valuable when dealing with people [2, 3].

In the field of healthcare, qualitative research is especially important to complement quantitative results [4]. For example, if quantitative methods can measure the effectiveness of an intervention, then qualitative methods can help researchers understand why the results might be impractical to implement in clinical practice [4]. This is especially relevant because changes in healthcare are highly dependent upon the people participating in their creation and implementation, making it very important to collect information about lived experiences [5]. The main questions tackled by qualitative research are very different than the ones that quantitative research focuses on. Qualitative research in healthcare is more focused on understanding meanings and experiences and it is particularly useful when dealing with the intricacies of improving the quality of a system [6].

Additionally, this type of method allows for researchers to understand the meanings, experiences, and views of participants, giving voice to the experiences of people that have to deal with the realities of the healthcare system [4]. Rice (1996) described that qualitative research is important for describing and explaining behavior, understanding how people understand their own health experience, the reasons behind people's behaviors, and how to best design a health program taking into consideration those that will use it [3].

However, analyzing qualitative data is a labor-intensive, rigorous, and exhausting process [7, 8]. It can take weeks for researchers to manually sort through datasets and often this process is even further complicated by the lack of available time [2, 7].

The field of quantitative data analysis, which usually deals with large datasets, solves this issue by making use of tools such as AI, but for qualitative researchers, this is less common [7]. However, there is potential for the use of AI for the analysis of qualitative data [2, 7]. For example, Chapmen et al. (2015) talks about the use of grounded theory to analyze data

related to healthcare to provide a more holistic view of the healthcare system. At the same time, grounded theory follows similar patterns to Machine Learning (ML), a type of Artificial Intelligence that focuses on pattern finding [7]. This opens possibilities of combining elements of qualitative data analysis with AI.

However, despite this potential, it is important to highlight that the presence of a researcher is paramount to qualitative research, and no software should look to replace it [4, 8]. This type of method depends on the critical and creative thinking of the researcher, as well as their ability to interpret the results and form meanings [4, 8].

Software packages exist to help researchers organize complex data more efficiently, amongst them, the most wildly used are QSR NUD*IST, now replaced with NVIVO, and ATLAS.ti [8]. Using software can assist with the laborious and repetitive tasks, as well as increase rigor to the process by making it systematic [8]. However, software must always be used with caution as computer assisted analysis is not able to replace the role of the researcher in making sense of the data collected [8].

1.1 AI for qualitative research

In this world of big data, it is necessary to develop tools to help researchers properly process it and keep pace with the current developments [2]. Despite many researchers' disliking the use of AI for qualitative research, many were found to already work alongside AI for assistance in data sampling, inductive code development, general coding, and examination of patterns [2]. The place for AI to assist qualitative researchers is reportedly very specific, and must not remove researcher agency or serendipity [2, 7, 9]

Overall, researchers are much more open to AI coming in after initial data analysis, as they theorize on how they could use AI for comparing and picking up on themes that were missed [2]. Researchers also are open to the idea of the AI coding alongside them, both influencing each other [2]. In this, the AI would act as a partner and make suggestions instead of just informing the researcher of what it did, which is an approach that has been used to develop software for qualitative research aided by AI [2, 9].

Currently, there are tools available that offer the automation of AI and ML to qualitative researchers. One of these includes the Computational Thematic Analysis Toolkit [10], of

which this research is a continuation of. The Computational Thematic Analysis Toolkit (CTA) seeks to make using Machine Learning more accessible to qualitative researchers with no programming experience. The purpose of this tool is to assist researchers and guide them in the process of conducting their analysis, without replacing interpretation. The software runs one of the three types of topic model sampling and then presents the user with a data visualization, which the researchers can use to help assist in data interpreting (Figure 1).

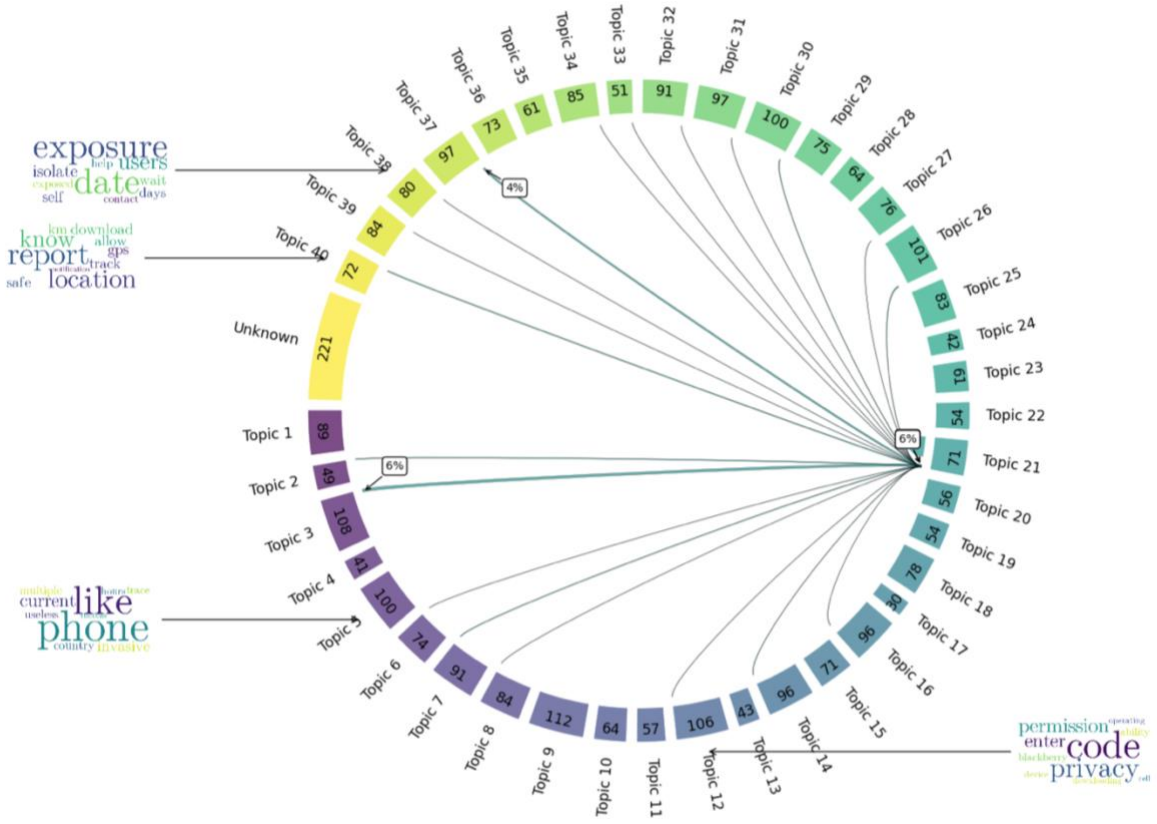


Figure 1: Computational Thematic Analysis Toolkit (Gauthier & Wallace, 2022) data visualization

Another example is Scholastic, a tool that seeks to put the human in the center of the research and use the AI only as assistance [9]. The researchers retain agency over when the machine will assist in the research and in which phases [9]. However, it is important to highlight how, in qualitative research with AI, there is no one size-fits-all and the tool should be flexible to accommodate the multitude of different needs from researchers [2, 9].

However, a big issue with those tools is that they can be complicated to use for non-programmers, or people who are not familiar with programming language. In fact, the lack of understanding of programming language is a big hurdle in the incorporation of using Machine Learning to assist in qualitative data analysis and researchers have expressed that tools that assist in data analysis should be intuitive and easy to learn [7]. Therefore, it becomes necessary to propose a tool that can act as a translator between programmers and qualitative researchers to help researchers feel more agency over their own research. For this, we propose the use of data visualizations.

Data visualizations (datavis) function as resources to assist cognition and foster understanding that raw data cannot provide [11, 12]. Datavis has many benefits, such as: aiding in objectivity, impartiality, allowing for the user to see impossible or complicated data, simplification, predicting patterns, supporting decision making, informed choices, and taking social and political action [12–17].

Datavis has great potential for helping qualitative researchers with no knowledge of Artificial Intelligence to use AI during their research [18], becoming an essential tool in bridging the gap between data production and consumption [19]. In particular, data visualization's potential for supporting change is interesting for the field of healthcare. Because datavis can assist people in making informed decisions and taking social and political action [17], they can become important tools for healthcare practitioners to analyze health data and with it enact change.

Datavis has been explored as a tool to assist in AI for qualitative analysis, however, it mostly centers around the exploratory phases of data analysis and coding [2]. Furthermore, when it comes to visualizations, their use cannot address all issues of AI trust or lack thereof [18]. However, they are powerful tools of communication and can make the process of working of AI more transparent [18]. In especial, visualizations that are adaptative and dynamic can support a wide range of stakeholders, including those that are not familiar with using AI or programming language [18].

Of course, data visualizations are not perfect. Some purely because of the nature of representing things, such as being limited in inferring causation [20], while others are

because of the people designing said visualizations, like data manipulation [21–23], and designer bias [11].

There are many models looking for ways to automatize the process of qualitative research, or assist the researchers, however, few consider how the qualitative researchers feel when using said tools. Jiang et al. (2021) talks about the requirements that qualitative researchers have for using AI in their data analysis, and a surprising amount of them pertain to how they feel when carrying out their research. They want to feel ownership over their data, preserve those eureka moments when the pieces slot into place, they want to be serendipitous when coding and still feel agency over their own research [7].

Agency itself is an important concept when it comes to research, since it is the feeling that enables people to take ownership of their situation and act upon it [24]. For this purpose, datavis has the potential to assist researchers that are not familiar with programming language to use AI during their research and preserve a sense of agency [18]. Additionally, the importance of designing tools for qualitative research that retain researcher agency over the AI has been noted to be important [9]. Therefore, for this thesis, we will be focusing on the fostering of agency as a feeling.

However, the way a data visualization is designed mediates how user will perceive their agency, and therefore designers need to understand how the design elements employed affect user agency to choose them accordingly [11, 25, 26].

However, for a person to become aware of their agency, they need to have enough literacy of a determined language to be able to interact with it [27]. In order to have proper literacy, a person needs to be able to recognize the elements that belong to a specific language, act upon them, and be able to reflect on what they say [27]. Our hypothesis is that datavis can be used to increase perceived agency for qualitative researchers when using AI.

1.2 Research Question

There are many ways to conduct qualitative research in healthcare, but the most common one is interviews, especially semi-structured interviews [4–6]. Semi-structured interviews allow health researchers the flexibility to produce richer data, as well develop a good rapport with the participants. It is appropriate for the collection of experiences and attitudes and

provide depth to the issues being researched, which is highly important for the field of healthcare [4–6]. Since this method of qualitative research is so widespread, we have elected to explore the use of data visualizations to assists researchers in conducting data analysis from interviews. Additionally, despite visualizations not being able to fix all the issues of miscommunication and lack of trust between humans and AI, they are still a powerful and flexible tool that could bridge that gap [18].

Furthermore, since the way designers design datavis will impact on how users interact with it [28], it becomes necessary to explore modes of data visualization that will foster researcher agency when conducting data analysis in order to better align with the needs and wants of qualitative researchers, especially since the use of data visualizations has been considered crucial for human agency when it comes to working with AI [18]. There are technological options available for researchers to use, but no clear guideline on how each different option suffices different needs of qualitative researchers. With this, the question that arises is: How can data visualizations be used as knowledge translation tools for fostering agency in qualitative researchers performing interview data analysis with the assistance of AI.

1.3 Objectives

1.3.1 Main objective

The main objective of this thesis is to: Map out the main barriers for fostering agency that qualitative researchers perceive when conducting data analysis using data visualizations.

1.3.2 Secondary Objectives

The secondary specific objectives are to:

1. Determine how qualitative researchers use visual techniques to make sense of their data analysis.
2. Compile what visual elements qualitative researchers find important to have in their data visualizations for qualitative analysis.
3. Determine what kinds of data visualization satisfy which needs of the qualitative healthcare researcher community.

4. Determine how qualitative researchers perceive their sense of agency when doing qualitative research.
5. Develop a template for using data visualizations as knowledge translation tools for qualitative researchers performing interview data analysis with the assistance of automated tools.

Chapter 2
Literature Review

In this chapter, three concepts will be discussed, the place of AI in qualitative research, data visualizations and their design, and the concepts of agency and how it can be fostered in media.

2.1 AI delegation and issues

Before diving into the possibilities for human-AI collaboration, it is important to discuss the issues that can come from automation, as well as what kinds of tasks should be delegated to it [29]. Human-AI collaboration can be something positive and likely necessary in the coming years [30]. Humans and AI see the world differently and there is value in their unique perspectives and the solutions they can come up together [30].

However, caution is necessary when discussing the place of AI in performing tasks. For example, Lubars and Tan (2022) highlight some of the issues that come from using AI indiscriminately, including an incident in which a program to detect people's sexual orientation via their images caused backlash and a repudiation letter from GLAAD and the Human Rights Campaigns, pointing out how dangerous the study was for LGBTQ+ and non-LGBTQ+ people [29, 31].

Even more recently, the use of AI has caused a commotion in the artistic world with the deployment of many AI generators that can create intricate piece of art in a short amount of time [32]. The main argument center around the nature of art and the fact that the AI cannot create anything, but only copy art that belongs to another artist and combine them [32].

Netflix Japan also had a controversial usage of AI when they fired artists in mass, only to then announce a short, animated series that used AI art generators to create the background [33]. This has brought up many issues of ownership and plagiarism, with artists claiming that they did not give permission for their art to be used in this way by the AI [32].

Despite the usefulness of AI for some tasks, should not be delegated to AI for a variety of reasons. For example, civil surveillance would have ethical, privacy and legal concerns if delegated to AI [29]. The same thing would happen to creative areas which require human motivation, such as creative writing, art, etc. [29]. Because of this it is important to establish

when to delegate tasks to AI and the appropriateness of this delegation. Deciding on which level of automation to assign to a task is not a hard science but there are parameters that can be used to make decision-making more solid [29, 34].

Automation of tasks is not a new concept, but the nature of AI brings some unique issues to the field. Frameworks to delegate tasks to machines existed for a long time. For example, Parasuraman et al. (2000) developed a framework for defining different kinds of human-machine interaction and their levels of automation. The authors analyze the automation in different levels of a task, which is divided into: 1) information acquisition, 2) information analysis, 3) decision and action selection, and 4) action implementation.

For the purposes of this thesis, we will focus on step 2, information analysis since we want to find out how to improve upon qualitative data analysis. For analysis automation, the algorithms can be applied to allow to extrapolation over time and prediction [34]. More automation means integration of various data points into a single value. In this stage, automation should help humans with perception and cognition, as well as management of information [34].

The framework presented by the authors seeks to help with automation by guiding people into deciding if the automation is needed. Six aspects are highlighted for consideration when wanting to use automation:

1. Mental workload: automation can lessen mental workload to a level that is healthier to the human as better for the completion of the task.
2. Situational awareness: high automation might lessen the awareness humans have over a system if they take away decision-making tasks from the human.
3. Complacency: when humans stop checking for errors in the machine and end up missing vital moments when the algorithm makes a mistake.
4. Skill degradation: if decision-making is always left to the machine (i.e., high automation), then the skill of the human operator can deteriorate.
5. Reliability: If the machine is not reliable than there will be no benefits to automation as the human operators do not trust it.

6. Cost of decision: low level decisions that have very low risks are better candidates for automation since they lessen the load of the human operator without risking causing skill degradation or loss of situational awareness.

Additionally, automation might be useful for very high-risk situations that are time sensitive and cannot rely on human unpredictability [34]. When it comes to AI it can be more helpful the investigate the roles that humans would like to give to AI and how to conciliate that want [29]. Lubars and Tan (2022) developed a framework to outline what tasks should be performed by AI based on four factors: motivation, perceived difficulty, perceived risks, and trust.

Motivation for doing the task take into consideration intrinsic motivation, goals, and utility as ways to measure motivation [29]. Ryan and Deci (2000, p. 69) defend that *"Motivation concerns energy, direction, persistence and equifinality--all aspects of activation and intention".* There are many aspects to motivation. People can become motivated because there is value to an activity, or strong external factors that convince them of the tasks' importance [35]. There are two main kinds of motivation: intrinsic and extrinsic, with the first one being the most relevant for this thesis. Intrinsic motivation is a powerful human feeling that drives change and is feeds on novelty and challenges to improve and explore [35]. It is a feeling that comes from within and cannot be replicated by external pressures, however, when there is no intrinsic motivation, behavior can be regulated via external motivation [35].

Motivation can also be regulated by way of the flow theory [36]. Flow is a state of mind in which intense focus is experienced and can be categorized by: (1) development of goals and monitoring of feedback, (2) immersion, (3) focus on task, and (4) enjoyment of experience [36].

Goals are defined as the interest someone has in learning to master a task and can be divided into leaning goals and performance goals, and utility is how important the task is considered by the person accomplishing it [29, 35]. The more perceived utility it has than the easier it is to accomplish. However, certain neurodevelopmental disorders can cause people to have difficulty in perceiving utility of tasks and thus low overall motivation [37].

Perceived difficulty of the task is divided into social skills required to complete it, creativity needed, whether the task will require a lot of effort – in terms of time or labor –, the expertise

that will take to accomplish the task, and the perceived competence of the people performing the task [29]. Perceived competence can also be assessed in terms of self-efficacy, which is a fundamental part of human agency [38].

Perceived risks associated with completing the task have three components: the accountability for the outcome of the task (i.e., in case of failure who will be held accountable), the uncertainty around the task that may cause failure, and the impact of the task (i.e., whether failure will have a negative impact on someone's life) [29].

Lastly, trust in the AI is how people deal with risks and uncertainty. Trust is an important aspect of automation and seems to strongly govern the decision to delegate or not [29, 39]. People are more likely to rely on automation that they trust than those they do not [39]. Trust is here defined as *"the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"* (Lee and See, 2004, p. 54). Trust in AI is multifaceted issue that stems from sociological and psychological issues that are often neglected to be considered, as well as dynamic and constantly changing [18].

Trust is divided by Lubars and Tan (2022) into 4 aspects: performance, process, purpose, and value alignment [29]. Performance is the ability of the machine to reach the goals set by the human. Process has to do with the inner workings of automation (e.g., how dependable it is, what is the integrity of the process and how does it conduct interpretation). Purpose is the intent behind automation and how it aligns with the goals of the human using it and depends on the machine providing explanations to its own internal processes. Lastly, value alignment means that humans need to trust that the AI's actions will protect and align with their interests and values.

Within these 4 parameters, motivation and difficulty directly influence one another. The more difficult a task is perceived as, the less motivation a person has to complete it, and the less motivation a person has, the more difficult seems the task [29]. Similarly, the level of risks involved in the task will inform the trust a human would have in the AI [29]. Risks have three components: the accountability for the outcome of the task (i.e., in case of failure who will be held accountable), the uncertainty around the task that may cause failure, and the impact of the task (i.e., whether the failure of doing the task will have a negative impact on someone's life).

With these 4 components Lubars and Tan (2022) consider 4 stages of AI delegation: (1) No AI assistance, (2) Human leads and AI assists (i.e., human does most of the work, but the AI assists where appropriate, for example when there are blockages or AI predicts a possible mistake), (3) AI leads and human assists (i.e., AI performs tasks and humans confirms and suggests changes where appropriate), and (4) Full AI automation. This framework is summarized in Table 1.

Motivation	Intrinsic motivation	Whether people are motivated internally to accomplish a task
	Goals	Interest in learning the necessary skills to accomplish a task
	Utility	How useful the accomplishment of the task is regarded
Perceived difficulty	Social skills	What social skills are needed to accomplish a task
	Creativity	How much creativity the task requires?
	Effort	Whether the task will be laborious or time-consuming
	Expertise required	What is the expertise required of the people accomplishing the task
	Perceived competence	How does the person doing the task perceived their own competence for fulfilling the task
Risk	Accountability	Who will be accountable in case of failure?
	Uncertainty	Will the task fail?
	Impact	Will the failure of the task have negative impacts?
Trust	Performance	Can the machine reach the goals set?
	Process	How dependable is the process?
	Purpose	Does the machine align with the human's goals and communicate effectively what those goals are?
	Value alignment	Will be machine protect the human's intents, interests, and values?

Table 1: AI delegation framework (Lubars and Tan, 2022).

Generally, people seldom prefer full automation and tend to opt for the human to be the one leading a task [29], however, they are open to AI assistance [2, 7, 29]. Lubars and Tan (2022) highlight that research should focus on ways to add AI to tasks while leaving the humans at the center, leveraging machines as support. This framework is not perfect and

there is plenty about human preference that cannot be encompassed in it. Human preferences are dynamic and might change over time, but the framework is a start into the field of human-machine cooperation so that delegation can be chosen with a greater degree of certainty [29, 34].

With these circumstances in mind, we characterize qualitative research as having potential for the integration of AI as a tool for assisting researchers, while keeping the focus on the researcher and not the machine. Qualitative research requires creativity and expertise that makes it impossible to be fully automated [4, 8]. However, it is also time consuming and laborious, which opens the possibility for integration of AI as an assistant. In fact, research shows that qualitative researchers are open to collaborating with AI, as long as their agency is not removed from the research process [2, 7]. The specifics of using AI to assist in qualitative research are furthered covered in the next session.

2.2 Data Visualization

We are currently living in an era of extreme data generation, collection, and storage, making data more accessible than ever [16]. With all this production, data visualization has taken in the role of providing insights into patterns and trends for the future, which means it becomes an essential tool to bridge the gap between people that consume information and those that collect it [16, 19].

Data means very little without interpretation, and to be made into something that can become meaningful it needs to be manipulated and transformed into information [40]. Data is only the base of the DIWK (data, information, wisdom, knowledge) hierarchy, in which data is interpreted into information, which is then processed as wisdom, and then transformed into knowledge [41]. There are many ways to follow this process, but we will be focusing on the perspectives from the field of Information Design, which has the purpose of organizing and simplifying data into information to increase accessibility [40, 42].

Datavis has a lot of potential for data analysis, as it does more than simply present data, instead creating a space where researcher can explore and interpret meanings for themselves [13]. Specifically, interactive data visualizations show promise of being effective sources of fostering user agency [43]. Additionally, datavis can provide cognitive assistance for researchers and help translate information for pattern finding and correlation [11, 16].

One of the biggest strengths of data visualization is its ability to simplify and synthetize [13], which requires the filtering and removal of some information. This process risks being vulnerable to the biases of those making the data visualizations, be that human or machine, since they are the ones judging what needs to be kept or removed [11]. If simplification can allow for the finding of patterns and trends [13, 16, 19], then it is also vulnerable to the biases of people making those data visualizations [11].

Additionally, simplification can also risks reinforcing the idea of the “normal” or “average”, which not only perpetuates the myth of an attainable standard, but also risks compressing information into a summary that reveals less about the data than the individual assessments [20]. Cooley (2020) discusses this when talking about how the data visualizations portraying death during the COVID-19 pandemic could make people lose sights of the meaning of real human death.

For the purposes of assisting with qualitative healthcare data, this simplification could clash with the initial purpose of using qualitative data to highlight lived experiences [3, 4, 6]. However, datavis can be particularly interesting for qualitative researchers when they focus on storytelling. When a datavis focuses on smaller narratives of individuals, they can garner feelings of empathy and engagement [44]. This goes hand in hand with qualitative research’s aim to highlight individual stories and prioritize lived experiences in the healthcare field [1].

Data visualizations have many benefits, such as allowing people to visualize concepts that otherwise would be too difficult, as well as large and abstract information [13, 45]. When data visualization is at its best it can allow for the conveyance of complex information, using simplification to allow for users to grasp at concepts quickly, however, at its worst it can confuse and manipulate those that use it [16].

Additionally, data visualizations also allow for data exploration and interpretation of meanings [13]. It can provide cognitive assistance and act as a knowledge translation tools, especially when the person using it is not familiar with the information being presented [11, 13, 16].

On the other hand, data visualization has a history of being used for manipulation and misinformation, since it can so easily promote persuasion and shape minds at a

subconscious level [16, 21–23]. In politics, data visualizations can prioritize specific ideologies and use visual language to further agendas through instating meanings, signs and values [17]. Since people depend so heavily on it for making decisions, their design can shape the future [20]. We have seen this during the pandemic when so many healthcare measures were being taken based on the models and data visualizations made to monitor the pandemic [28, 44].

Specifically for research in partnership with AI, researchers have expressed the desire to use visualizations to help them find patterns and connections [7, 9]. Hong et al. (2022) have especially taken data visualizations into consideration when developing their platform, Scholastic [9]. Their tool is entirely mediated by a map data visualization, which allows the users to navigate through the tool using the visuals as a guide.

2.2.1 The design of data visualizations

Data visualizations communicate information using their own set of grammatical rules and internal structure, which can be harder to understand when one does not have literacy in them [27]. Much like writing, they create meaning by using a set of established rules between those that create it and those that consume it [46]. However, different types of design elements will communicate different meanings, and it is important to understand how each works to be successful [47].

Data visualizations are deeply influenced by the time, place, and cultural background they are inserted into, but despite that, there are still underlying principals essential to their design that should be considered when making and analysis them [47, 48]. One of these principals is the visual language of the media, which is the combination or words, images, and shapes to create a language [48]. It is separated into morphology (the most basic visual elements that can be used) [47, 48], syntax (how these base elements combine), semantic (how the elements convey meaning), and pragmatics (how language is used in real world applications).

Visual language is a complicated field to study, since it encompasses many different areas, such as data science, information designers, human-data interaction, among others [49]. Because of this, different authors can use different terms when talking about the same phenomenon [49].

Furthermore, data visualizations function not only as informational tools, but a storytelling medium [21, 44]. Their design has the potential to enforce one narrative or another, weaving tails out of data, which can be used both for deception [21, 23, 50], or for cognitive assistance and empathy [11, 14, 44].

Data visualizations as storytelling tools is especially interesting when trying to foster agency, since stories are such ubiquitous parts of society and can aid in informing, explaining, understanding, persuading, and making sense of the world [51]. In data visualizations, Cooley (2020) highlights how there were better results for datavis in fostering empathy and motivation when they incorporated storytelling elements.

To support storytelling in datavis, 4 elements are the most common: narrators, sequentiality, temporal dimensions, and tellability [51]. To create a narrator the data visualization can make use of voice narrations, commentary inferring emotion, animations, and more [51]. Sequentiality is when a data visualization makes explicit the sequence of events necessary to trigger a situation and the correlation between them [52]. In datavis this can happen in many ways, such as through scroll telling, when scrolling and zoom features are combined to generate a narrative [51]. This can happen with or without user input.

Temporal dimension is when the media shows the passage of time and the changes that happen during that time [52]. Lastly, tellability is how meaningful the user finds the story being presented to them [52]. In datavis, this happens if the users connect to the message being shows and believe it to be significant.

Ugaya Mazza (2022) proposes a way of analyzing datavis with the elements of visual language by different authors [14, 47, 49, 51, 53], which are separated into syntactic, semantic, classification of graphic representations, and elements of engagement [28].

The syntactic elements are the characteristic of individual elements (i.e., color, orientation, size, position in space, proximity, etc.), while the semantic elements are for example, the metaphors used to make sense of those elements, or the interactions one can have with them (i.e., zoom, filter, distortion, search, and locate). The classifications are the types of data visualization and how they combine (e.g., maps, linking diagrams, tables, etc.), as well as the needs that are fulfilled by distinct types of visualization. For example, maps

are better suited to see the geospatial data [47, 49]. Lastly, there are the elements of engagement, which are how data visualization can foster storytelling.

Those elements are how a data visualization communicates meaning, however, to foster the feeling of agency that qualitative researchers require out of data analysis, the design of datavis needs to go beyond and consider how agency can be fostered.

2.3 Agency

Agency is defined by Murray (1997, p. 126) as the “satisfying power to take meaningful action and see the results of our decisions and choices”. However, agency is not equal for everyone, as people are born in different environments and perceive agency differently [24].

Acting alone is not agency. People can still sense agency even when their actions are limited (i.e., a game of chess). This happens because those limited actions have immediate consequences and cause change [54]. The more a choice changes an environment, the more people would experience agency, regardless of if the choice had a positive or negative impact [54].

Even more than acting, agency is also the possibility of taking an action if so desired, and sometimes, this perceived agency is just as powerful as action-taking [24]. Complementing this concept, Tanenbaum and Tanenbaum (2009) define agency as a commitment to meaning, emphasizing that it happens when a person chooses to immerse into a narrative and believes in what they are doing [55]. Therefore, agency is not simply acting within a world, but expressing your intentions of acting and receiving feedback from your surroundings [24, 54, 55].

2.3.1 Fostering agency

Any piece of media can foster agency, and this can be done in many ways. For example, games use interactivity to foster agency in the players, while movies tend more towards perceived interaction through using tropes [56]. Eichner (2014) divides agency into three levels: personal, creative, and collective. Personal agency is fostered through the mastery of narrative, choice, action, and space and must be perceived before someone can perceive creative and collective agency [24].

Mastering narrative happens when one identifies the genre of media, they are interacting it and can predict what will happen next, thus gaining pleasure from seeing a narrative unfold in the way they predicted. For example, in graphs, the lines act as a narration point that guides the reader through the data like it's a story [28], which makes a line graph foster greater agency than the same data in a table.

Mastering choice is when the user realizes they can make a choice within the environment they are in [24]. This type of agency works best when there are rules that restrict the world around the user, instead, when there's too much freedom, agency is more likely to be fostered through a commitment to meaning [55].

Mastering of action happens when the user moves their body and sees the same movement reflected on the media they are interacting with [54]. For example, in the game A Ceremony of Innocence, the movement of the mouse on the screen acts as an extension of the human body and the user perceives agency through its movement [57]. In datavis this can be perceived though moving a mouse over the screen and having the visualization respond to it (e.g., pop ups when the mouse hovers over a data point).

Lastly, mastering of space is when the user perceives agency through the act of moving through space [54]. In order for this to happen seamlessly, the user needs to have enough knowledge of the media they are interacting with to use movement. For example, if you try to move and the zoom does not function in the way you are used to, then the agency is traded for frustration [28, 57].

Creative agency comes from the creation of material for the media one is interacting with, for example, how people create MODS for video games [24], while collective agency is fostered through the creation of a community around a piece of media (e.g., Comic Con). For this thesis, we will be focusing on only personal agency since it is the first step before perceiving the other two kinds of agency [24]. A summary of the different kinds of agency and how they are perceived can be found in Table 2.

Types of Agency		
Personal agency	Mastering of narrative	When one identifies the genre of the media they are interacting with and can predict what will happen next by following the breadcrumbs left by the media.

	Mastering of choice	When one is given a choice to make.
	Mastering of action	When one acts in the real world and sees that action reflected in the media they are interacting with.
	Mastering of space	When one can move through a physical or digital space through an understanding of movement mechanics.
Creative agency		Agency that comes through the creation of objects that related to the principal media.
Collective agency		Agency that comes from the interaction with a community that consumes the same principal media.

Table 2: types of Agency.

Specifically for data visualizations, agency is also fostered through the process of creation, as in, the user perceives their agency because they create the visualization with the designer through interactions [43]. It might be a simple zoom, or an input of numbers, but since user can take actions that have effects on the visualization agency is perceived through that mutual crafting of meaning [43, 55].

Additionally, the agency of the user is mediated by the agency of the designer, thus making the end product (visualization) a collaboration between the designer, visualization, and the end user, each operating with their respective constraints and context [43]. Figure 1 illustrates this by showing the perception of agency of the designer and user compared to what the data visualization shows. When the user is given no control over the message, data, and design, then they experience very limited agency. When they are given unconstrained control over these three elements, then their sense of agency overpowers the designers' [43].

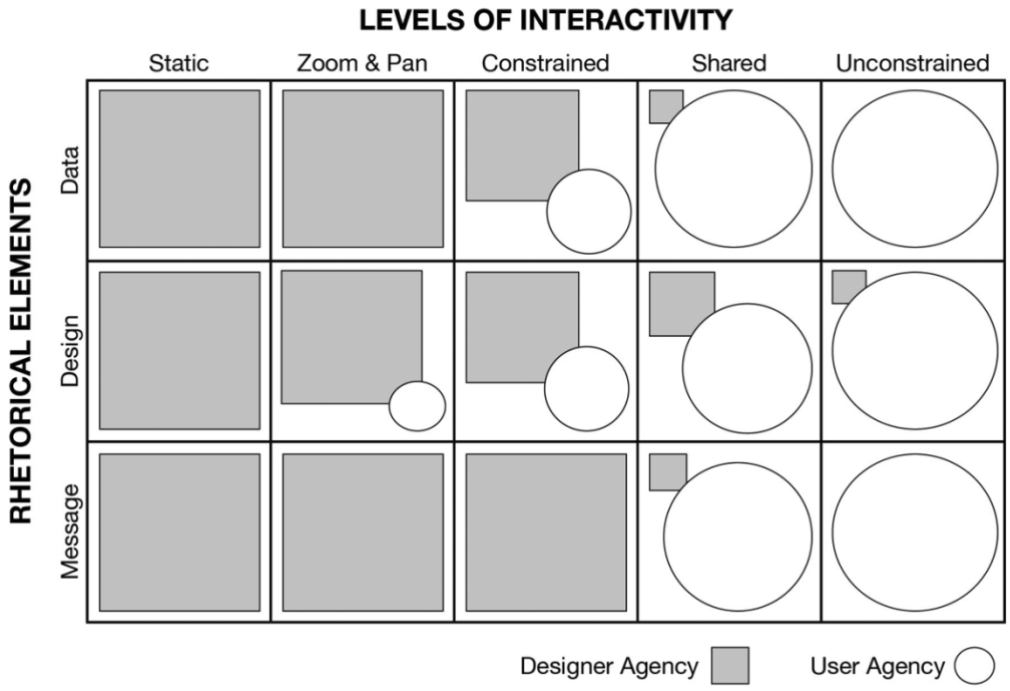


Figure 2: user vs. designer agency (Rawlins & Wilson, 2014)

For example, when it comes to the data visualization provided by the CTA the user (i.e., researchers) have unconstrained interactivity with the data, but static interactivity with the design and constrained with the message, which means out of the three elements of interactivity to foster agency, this datavis only fulfills one.

Chapter 3
Overview of Research Process

When it comes to designing objects, there are many benefits to taking extra care to ensure usable systems, such as increased productivity, reduced errors, reduced training and support, improved acceptance, and enhanced reputation [58]. Maguire (2001) discusses how to achieve usable systems from the perspective of human-centered methods. Human-centered design has 5 key principles that need to be followed to ensure the creation of usable systems [58]. Those are:

1. Active involvement of users during the development of the system;
2. Clear understanding of user requirements, as well as the tasks that need to be accomplished;
3. Appropriate delegation of tasks between human and machine (i.e., knowing what are the tasks that need to be done by the machine and which ones should be left to the user). This point is especially valid since for this thesis it will combine with AI task-delegation [29, 59].
4. Iterative process of design solutions, and;
5. Multi-disciplinary development teams.

A more specific method stemming from Human-Centered design is Design Science Research (DSR), which was used for this thesis. Development methods like DSR are iterative and generally consist of 5 phases: planning the process, understanding specific context of use, specifying requirements, producing prototypes, testing the prototypes against requirements, and then looping back in case of necessity [58, 60].

Design Science Research (DSR) is an iterative and cyclical method that intertwines the creation of a solution with the comprehension of a problem [60]. This was considered vital since the iterative nature of the method and the multiple points of contact with the user are important elements for designing usable systems [58]. During this section I will discuss how the DSR was adapted to fit my research project, as well as the 4 phases we used, and the techniques employed during each phase.

My method was divided into 4 phases: (1) Problem comprehension, (2) Specifying the user needs and requirements, (3) Development of prototype, and (4) Beta-testing of the prototype. Phases 2 and 3 were heavily influenced by user centered design [58, 61, 62], which has been used to develop products with positive reception by the users [62].

Phase 1 was dedicated to further understanding the problem proposed through a literature review and analysis of other similar software out in the market though a technique called Close Reading. Then, phase 2 was dedicated to understanding the specific needs and requirement of the users, which was achieved through a semi-structured interview with users, in this case qualitative researchers.

With those phases completed, guidelines were developed for phase 3, which was the development of a paper prototype. Following that, phase 3 was a beta-testing of the prototype in which the participants of the interview in phase 2 returned for another session of user input. This phase had the goal of both testing the prototype and confirming that the design choices that came out of the first interview were what the participants wanted and needed. A visual representation of the entire method can be seen in Figure 4.

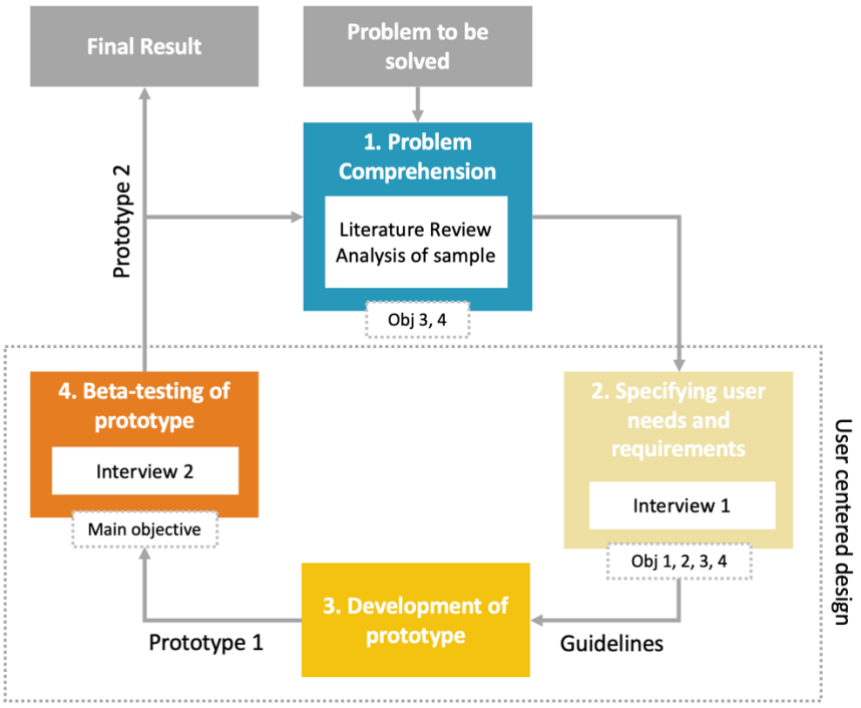


Figure 3: method

The next 5 chapters will be dedicated to explaining the process of each of these 4 phases and the final prototype, as well as the results garnered from each of them.

3.1 Limitations and mitigation strategies

This research is qualitative and follows the Design Science Research Method, which means the results will be heavily context specific and generalization will be limited. In addition, only one designer will be participating in the brainstorming and developing the prototype, which risks designer bias. This will be mitigated with the guidelines derived from the interviews with researchers and the literature review.

Chapter 4
Problem Comprehension

This phase was important for identifying the specific context of use of the object as well as some of the needs and requirements from the stakeholders [58]. It comprises two parts for this thesis. First, a literature review was done to better understand the problem and propose a solution. Then, an analysis of available software was done to assess what are the currently available data visualizations and how they are incorporated into their respective software.

The analysis of available software was done with the purpose of exploring the extent of which currently available options meet the needs of the users, as well as map out usability issues to incorporate in future designs [58]. It was guided by the concepts reviewed during the literature review using a technique called Close Reading. This technique allows for the researcher to develop meaningful connections that come from their own experience with a piece of media [56, 63, 64]. It is especially useful when dissecting mediatic objects to map out how their individual elements contribute to fostering specific emotions [56, 64]. For this thesis, it was used to assess how currently available software could come to foster feelings of agency in researchers. Agency, in this case, referred to what is fostered through the interaction with data visualizations, as opposed to control or input/output of the AI.

Close Reading is a method that has been used previously to analyze both interactive media, and specifically data visualizations [28, 63]. It consists of approaching a piece of media with theories and as you analyze it, construct an analytic protocol for the specific situation at hand, which can be used in the future by other researchers [28]. In the case of this thesis, the theories used to analyze the sample were agency and the visual language principles of data visualization design. This phase partially completed the specific objectives 3 (determining what kinds of data visualization satisfy which needs of the qualitative healthcare researcher community) and 4 (determining how qualitative researchers perceive their sense of agency when doing qualitative research).

4.1 Sample of data visualization for the analysis

In total 4 different software were considered for the sample: The Computational Thematic Analysis Toolkit [10], NVIVO, ATLAS.ti, and Scholastic [9]. Only completely or partial open access platforms were included for reasons of convenience, or ones that had a free trial before purchase. Scholastic was the only data visualization that was not available for use at the time of the analysis. However, the research team has publicly available YouTube videos showcasing the interface and data visualization, so they were included in the sample. Additionally, only the data visualization aspect of the sample was analyzed.

ATLAS.ti, Scholastic and the CTA are software that make use of AI to assist qualitative researchers, however, NVIVO does not. Still, the software was considered for the analysis for its ubiquity in the world of qualitative research.

Thirteen different data visualizations were analysed, 2 from the Computational Thematic Analysis Toolkit, 7 from NVIVO, 3 from ATLAS.ti, and 1 from Scholastic. All images from the sample are available in appendix A.

4.2 Analysis protocol

The Protocol was created based on the 4 types of personal agency fostering [24]. The elements that were present in the sample and were observed to foster agency were documented accordingly. The elements were based on the proposed syntactic, semantic, classification, and elements of engagement by Von Engelhardt [47], Börner et al. [49], Weber [51], and Ugaya Mazza [28], which were explained in chapter 2.2. However, any additional elements that were discovered during the analysis were added to the table iteratively. The protocol can be found in appendix B.

4.3 Results from sample analysis

The visualizations were analyzed iteratively with the construction of the analytic table. The analysis focused on the three kinds of fostering agency in media interaction, mastery of: action, narrative, choice, and space. After the analysis, it was found that different graphic elements helped foster each of the different kinds of mastery.

For narrative mastery, the elements that helped foster it were using proximity to establish connections, focusing on showing connections, and the use of frequency as a storytelling tool. For choice, three elements were found to be positive: editing capabilities, explanations of machine processes, and guidance. For action the use of filtering elements and the ability to reveal more data if chosen were considered positive. Lastly, for space, the use of color-coding, nesting of different windows, breadcrumbs and the clustering of similar elements helped with navigation.

Elements that were found to hinder in the fostering of personal agency were: over focusing on quantitative data in detriment to the narrative, lack of freedom for editing, lack of explanation of how the machine generated results, inconsistent and meaningless colors, inability to zoom in, inability to click and drag, and inability to move and manipulate the data visualization. These elements will be further explained below as it pertains to each of the four software in the sample.

4.3.1 ATLAS.ti

ATLAS.ti is a software that uses AI to generate codes and then allows for the exploration of those codes through a data visualization. However, the machine does not allow the researcher to add any new codes independently, limiting editing capabilities to changing the codes that were already created. The codes and data visualizations are generated based on the dataset imported by the user, but there are no elements that explain what process the machine went through to generate results. Therefore, to understand the results the users need to spend significant time and cognitive effort to make sense of the codes the AI generates.

All data visualizations organize the codes by frequency, using bar graphs (Figure 5), pie charts (Figure 6), or treemaps (Figure 7) so the research can see what the most prevalent themes were. The interface also separates the codes by color but lacks any explanation of why the colors are applied in that way. Two different codes can have the same color for no apparent reason.

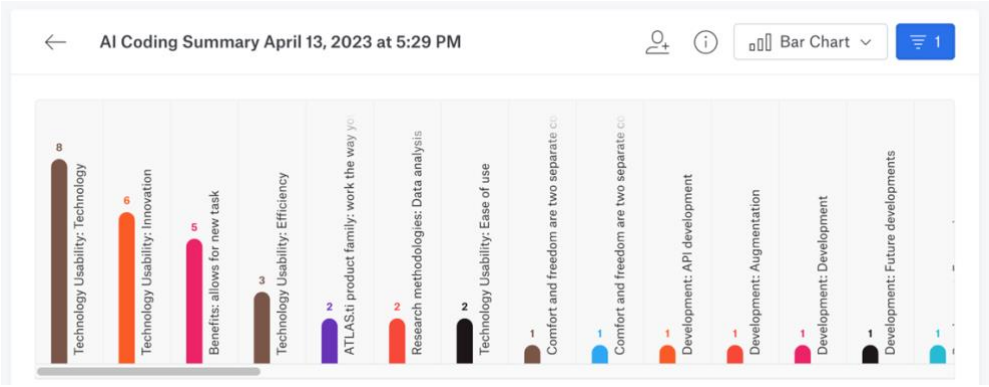


Figure 4: ATLAS.ti line graph.

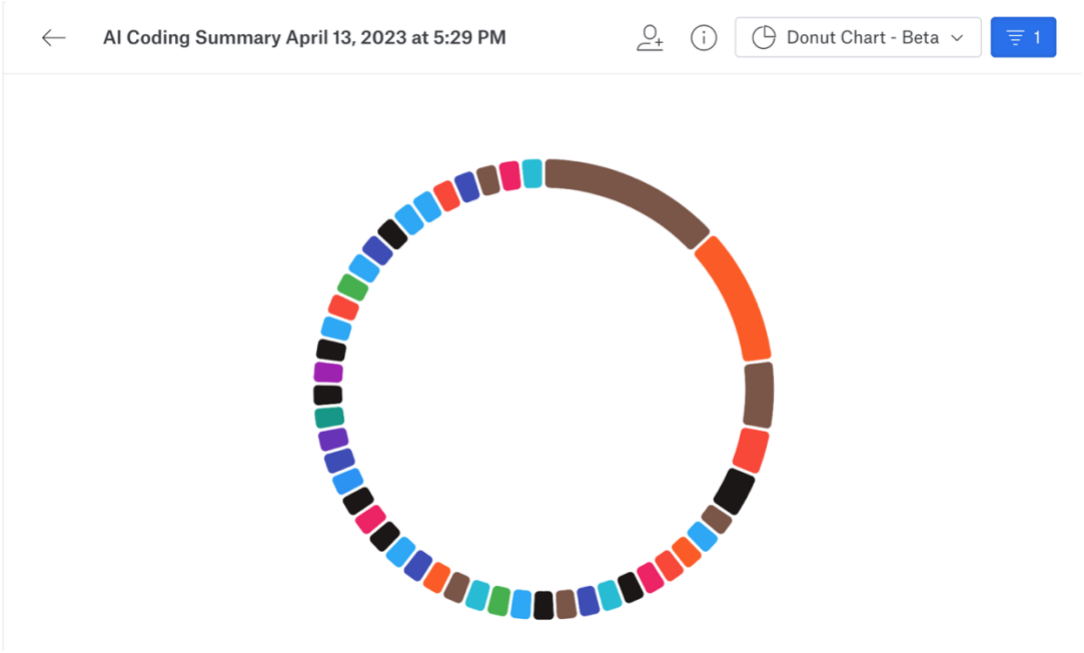


Figure 5: ATLAS.ti statistical pie chart.

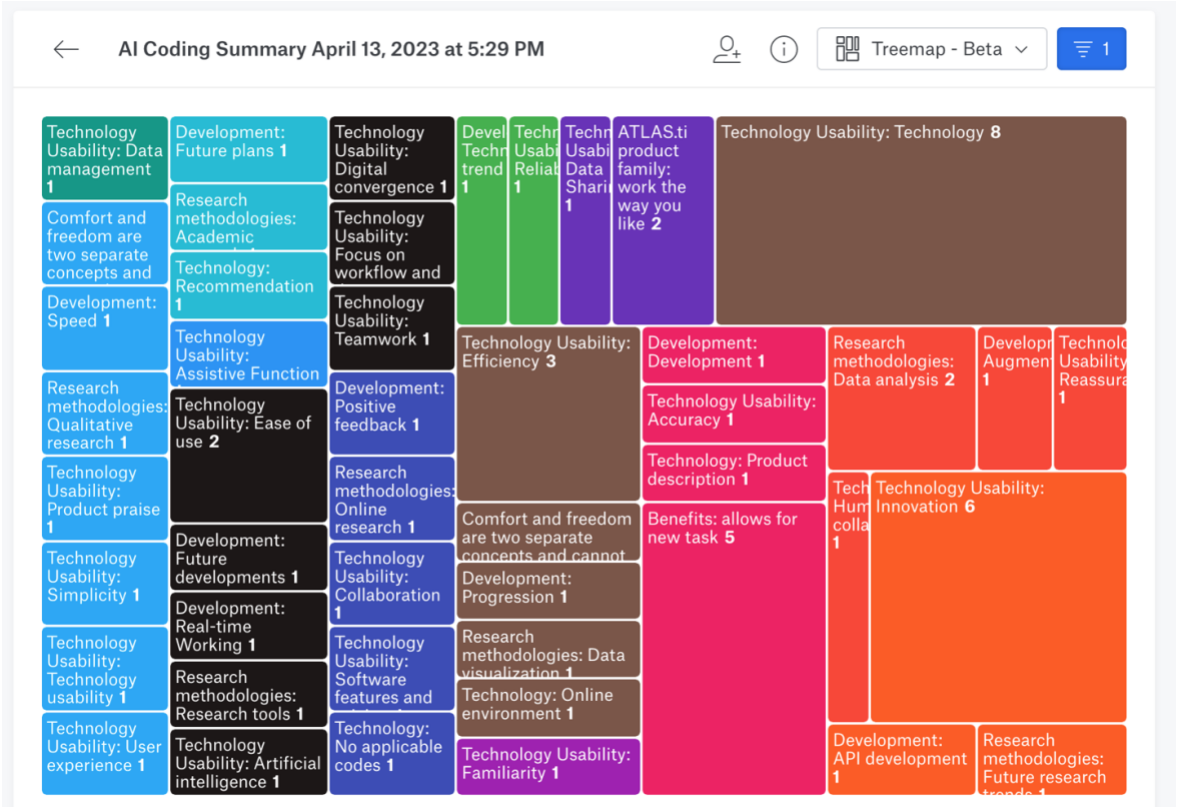


Figure 6: ATLAS.ti treemap.

In terms of navigation, most elements in the interface open separate windows to edit the data relating to the element that was clicked. To return to the original page, all the user needs to do is close the pop-up, allowing for an easy navigation. A filter is also available so the user can search for only what they want to see.

When the user selects a portion of text, the text is highlighted, and a window opens so codes can be edited, and comments can be added. However, despite having an option for creating new codes, this is not possible in the trial version, which was the one that was used for the analysis.

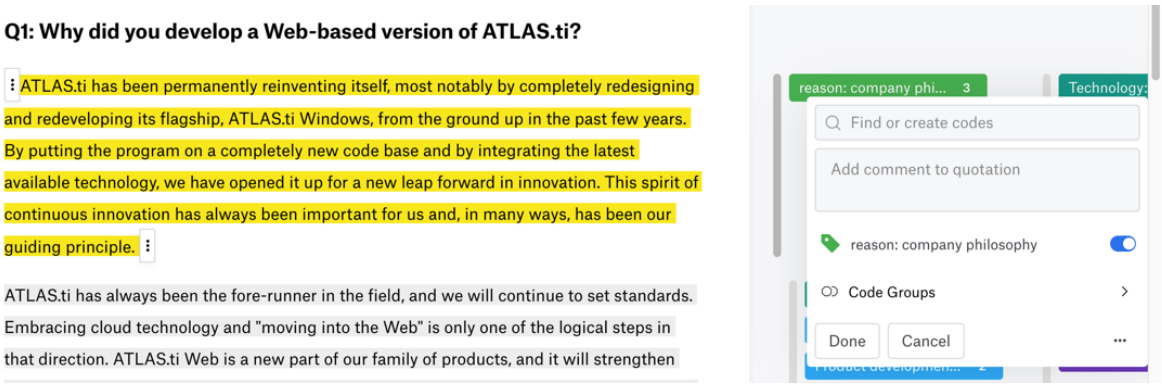


Figure 7: ATLAS.ti, selection of text and editing of codes.

The software is separated into 3 different visualizations: a bar chart, a pie chart, and a treemap. For all of those, the data is organized based on frequency and best suited for ranking and ordering data points. There is a limited option for drawing connections when the user clicks on a section of text and is presented with Figure 9. There, the user can see all codes that have been applied to the same section of text, therefore drawing connections between two different codes. However, the software only does this on the individual level, providing no visualization for connections throughout the dataset.

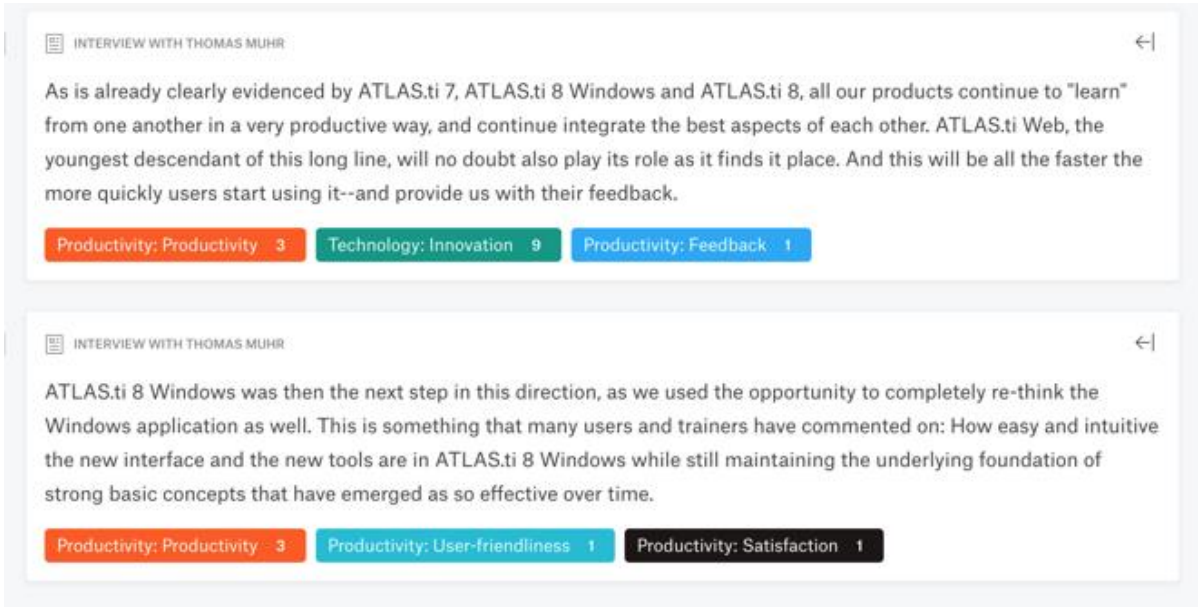


Figure 8: ATLAS.ti, connections of codes.

Additionally, the software makes use of color coding and saturation as navigation and grouping tools. For example, in Figure 10 the user can select different codes in the bar graph and the software will indicate they were selected using color and saturation. After the selection, ATLAS.ti presents the user with all the strands of text that have been coded using that theme. The change is instantaneous and as the user clicks on more codes, the strands of tact below are updated at the same time.

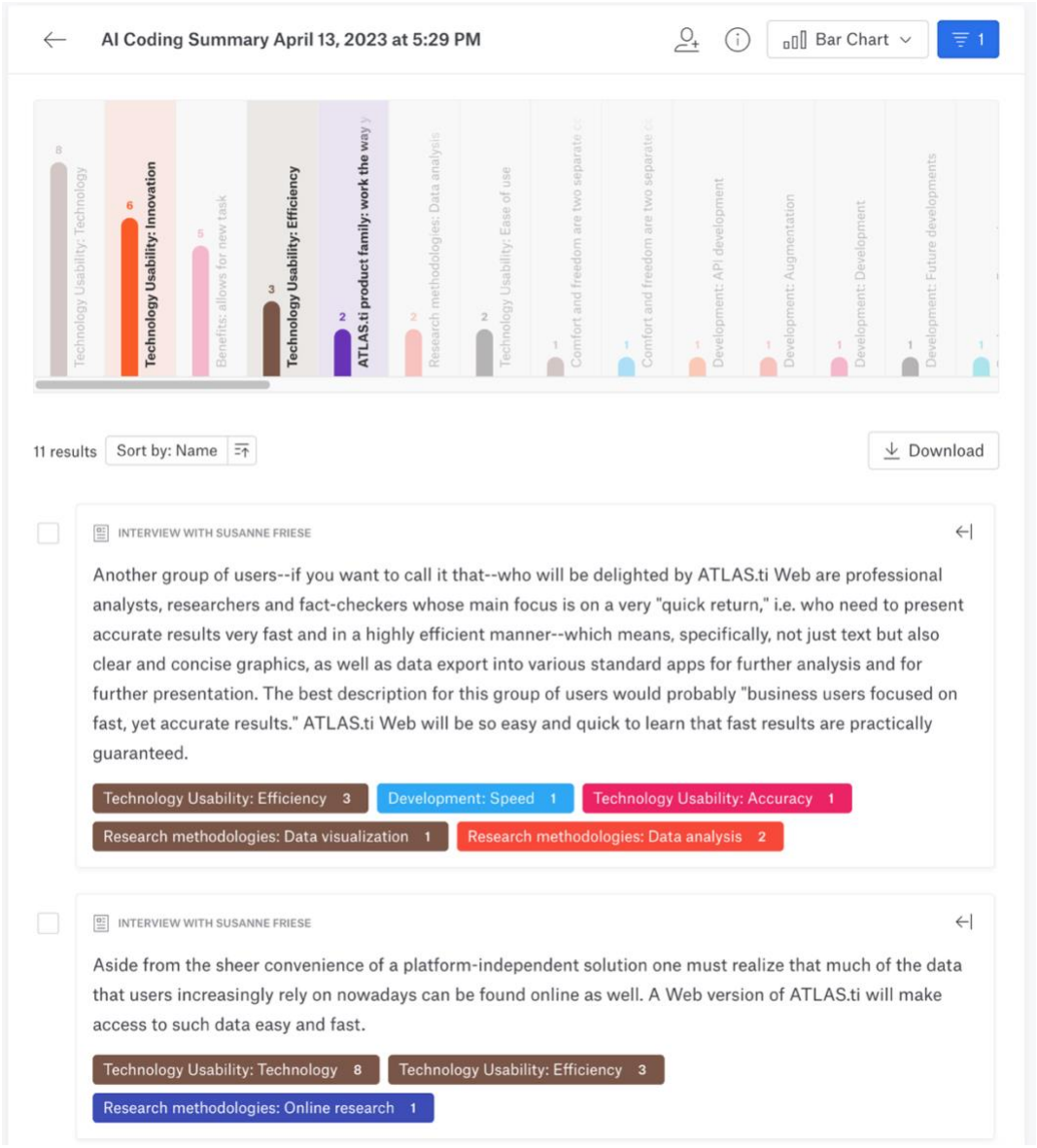


Figure 9: ATLAS.ti color as a navigation and grouping tool.

4.3.2 NVIVO

NVIVO is a qualitative analysis software that can either allow the researchers to code their own data, or to use AI to generate automatic coding. It was the software with the greatest variety in data visualizations available, having options for a word cloud, bar chart, linking diagram, treemap, pie chart, and table.

NVIVO's data visualizations give greater emphasis to building connections through different kinds of linking diagram. For example, the user has the option to view the connection in their data by checking which codes are applicable to which participant (Figure 11), or by checking the connections specific words have with others (Figure 12). In addition to that, the software also has options for data visualizations that highlight frequency, such as word clouds, bar graphs, and tables.

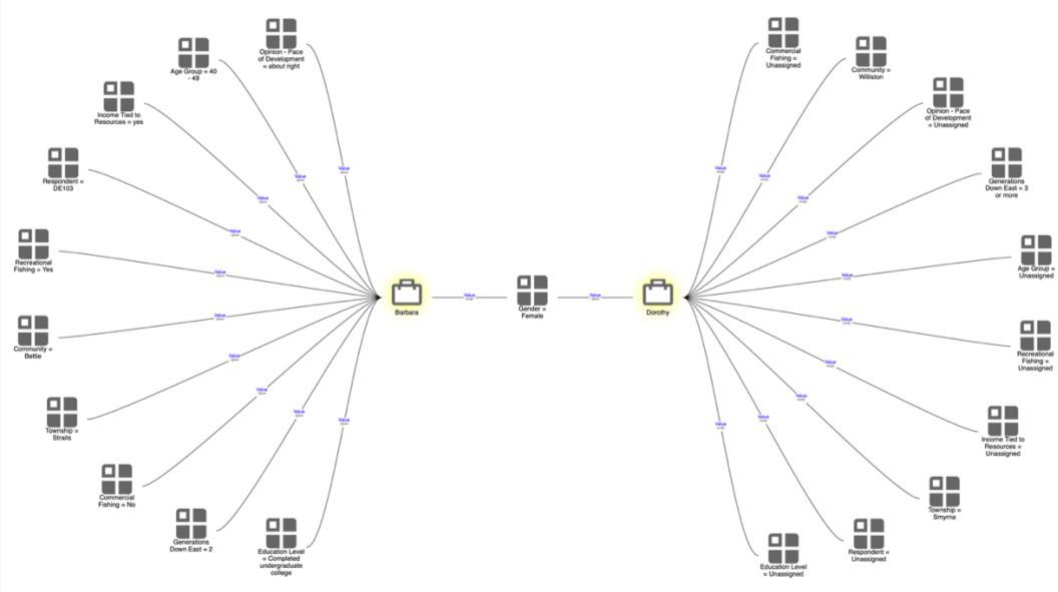


Figure 10: NVIVO linking diagram for codes.

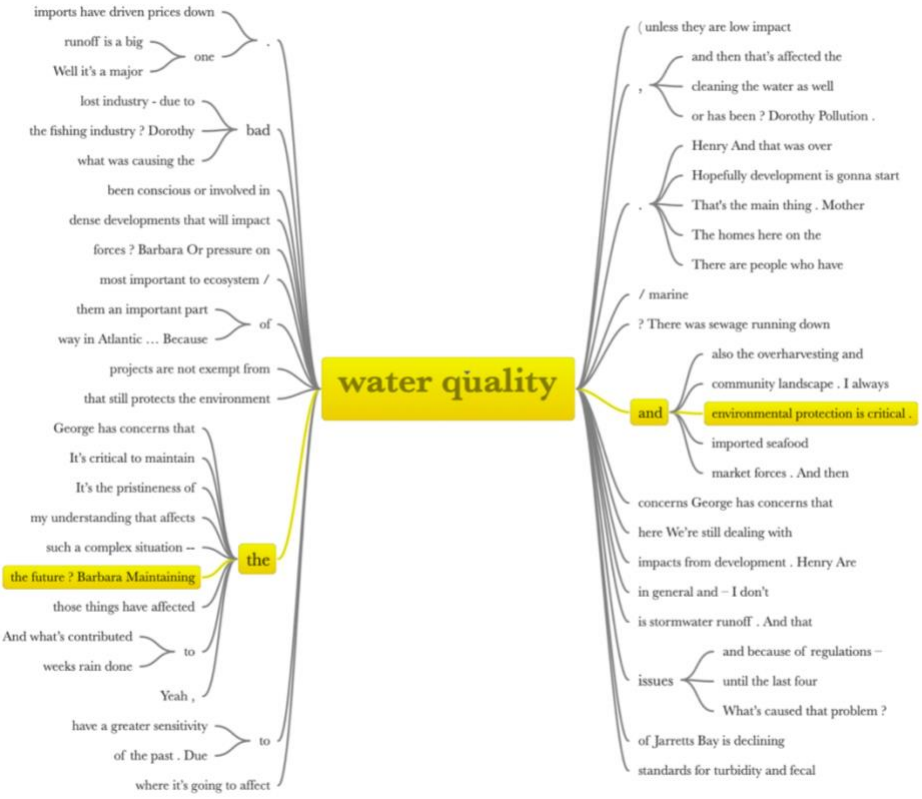


Figure 11: NVIVO linking diagram for words.

The software provides some interactivity through clicking to reveal more information. For example, in Figure 12, yellow highlights appear to show the path of the word tree, guiding the user through the machine's process of organizing the data into that kind of visualization. The colors act as both an exploration and a guidance tool.

In terms of navigation, NVIVO has some features that facilitate the mastery of space. One such feature is color-coding, which is highly used both to derive connections (such as the one mentioned before), and to group similar elements, and highlight frequency. Colors are used for coding, for example, in the treemaps and sunbursts, where each theme is connected to its sub-themes by color and proximity (Figure 13). When it comes to frequency, the tables are equipped with heat maps that show counts via color saturation (Figure 14). These uses of color are elements that help guide the user and facilitate navigation within the software.

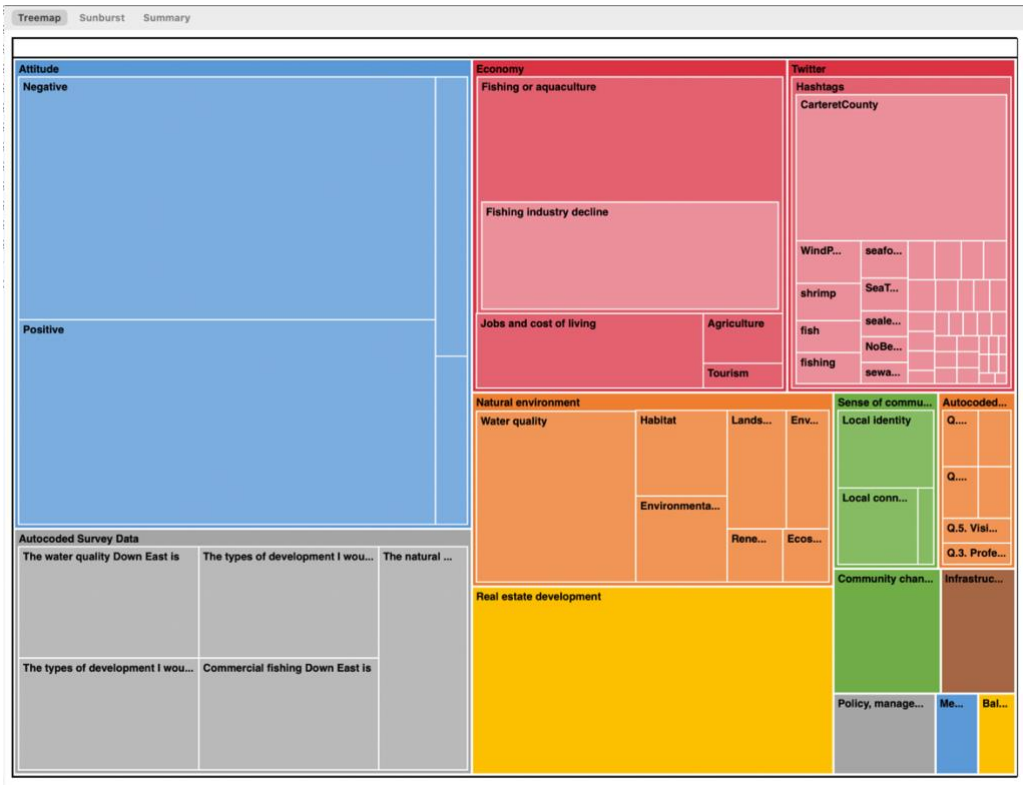


Figure 12: NVIVO treemap using color-coding and proximity to form groups.

	A : Agriculture	B : Fishing or aquaculture	C : Jobs and cost of living	D : Tourism
1 : Mixed	0	7	4	0
2 : Negative	5	97	20	1
3 : Neutral	0	7	4	0
4 : Positive	0	31	20	1

Figure 13: NVIVO using color-coding for heatmaps in a table.

Another form of navigation tools is the ability to filter information by codes, files, connections, classification, etc. NVIVO provides the researcher with a wide selection of filters and editing capabilities that make data manipulation more accessible. The software also shares the quantity of each of the filters in a circle to the right, so the user can have an overall idea of how much data each filter contains (Figure 15).

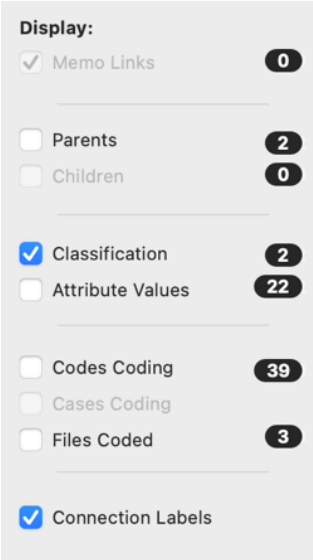


Figure 14: NVIVO filtering capabilities.

In addition to that, NVIVO is equipped with a click and drag feature for most of its linking diagrams that allows the users to move the elements in the visualization to where they please. To further guide the user, the elements that are clicked appear with a blue square around them (Figure 16).

Figure 15: NVIVO click and drag features.

However, not all data visualizations have the same manipulation capabilities. Linking diagrams can be zoomed in, moved around with the scrolling bars, and have their elements moved around, while other kinds of datavis have very limited interaction. For example, bar graphs and word clouds have almost no interaction other than hovering over data points to see more information about data (i.e., bar graph), or right-click to get more information (i.e., word cloud).

Even for datavis that allows some movement, it can be difficult since they reach large sizes and the only way to move through it is to use the scrolling bars, which can be frustrating (Figure 17). This inconsistency in editing capabilities was highlighted in the

analysis as an element that could come to hinder agency when it came to interactions with the data visualization tool.

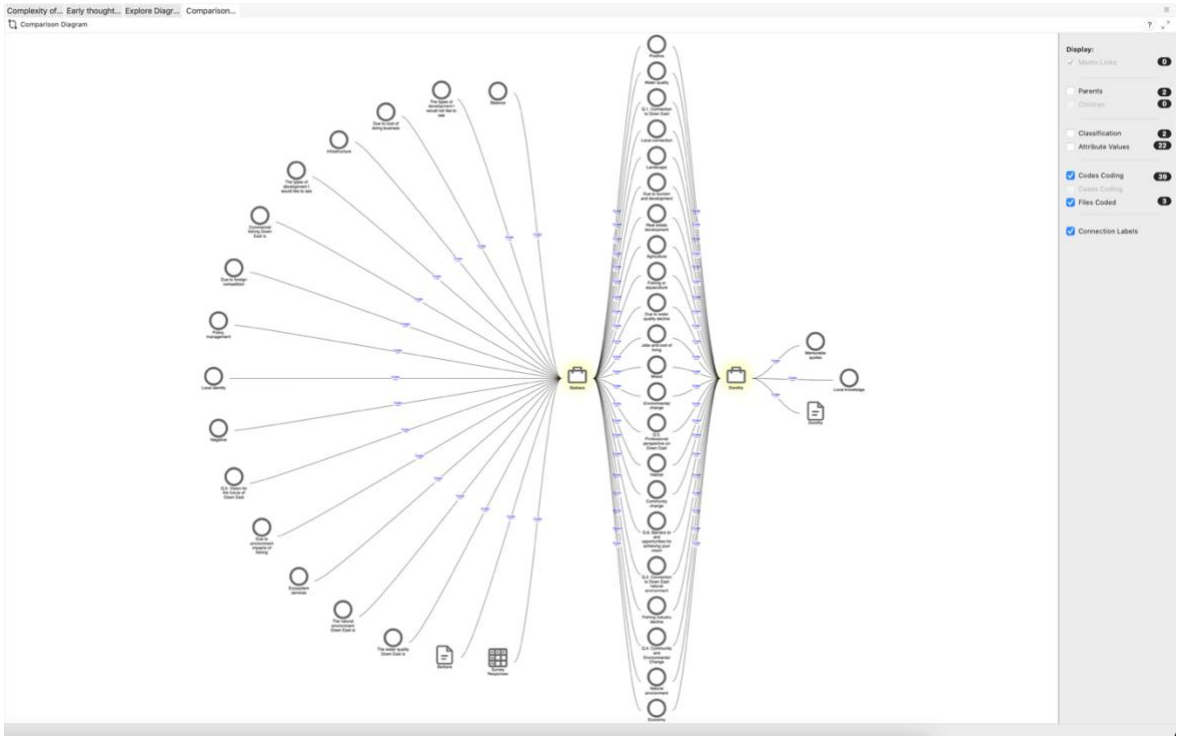


Figure 16: NVIVO example of how large the data visualizations can get.

Right-clicks are also available for most of the data in the data visualizations. Doing so will open a separate window and allow for the user to further explore their data with other kinds of data visualization. The example in Figure 18 illustrates how the user can right-click on a word in the word-cloud and open an option for running a text search query, which in term generates a word tree data visualization. However, the software fails to communicate this is an option and the user has to rely on previous knowledge, experimentation, or outside tutorials in order to learn that the feature even exists.

Figure 17: NVIVO right-click function for data exploration.

Overall, lack of guidance is an issue that is present throughout NVIVO. The software has many different functionalities and tools for qualitative researchers, and it can be difficult to locate them all without external assistance. This is detrimental to fostering agency, as it limits the users' choices and possibilities of acting and makes navigation in space difficult.

4.3.3 Computational Thematic Analysis Toolkit

Unlike the other software included in this analysis, the Computational Thematic Analysis Toolkit (CTA), was originally developed for qualitative analysis of short text (i.e., social media posts). Similar to NVIVO, it allows for the user to code alone, or with the help of AI. CTA has two kinds of data visualization, one for displaying the information that is run by the AI, and one for when the user manually inputs data. The former is presented as a statistical linking diagram with word clouds for each section (Figure 19), and the latter as a mind map (Figure 20).

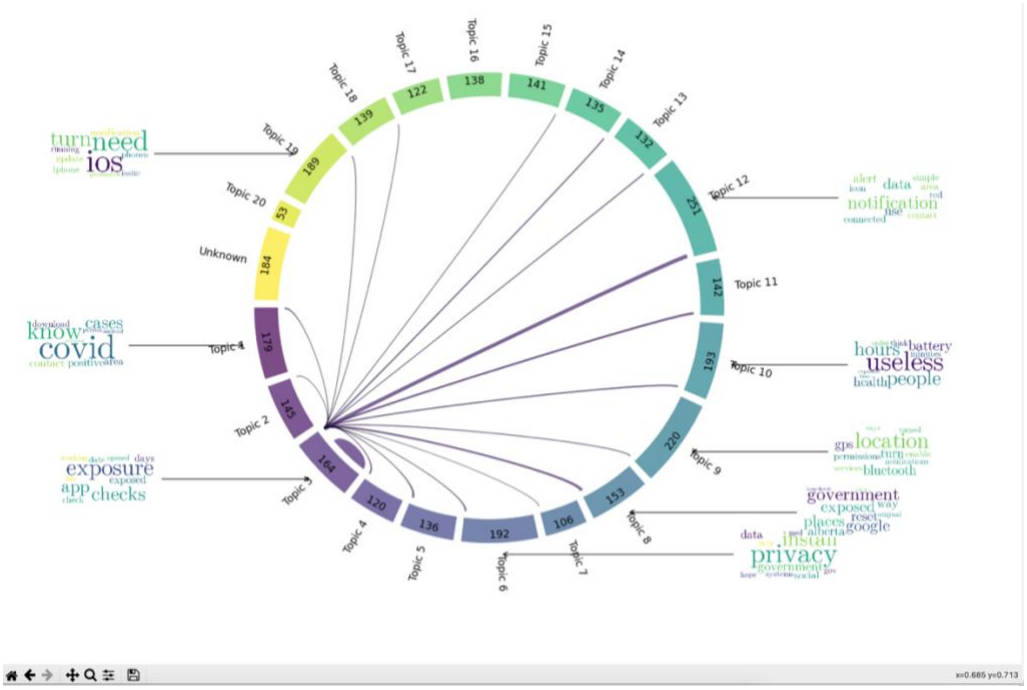


Figure 18: CTA statistical linking diagram with the word clouds.

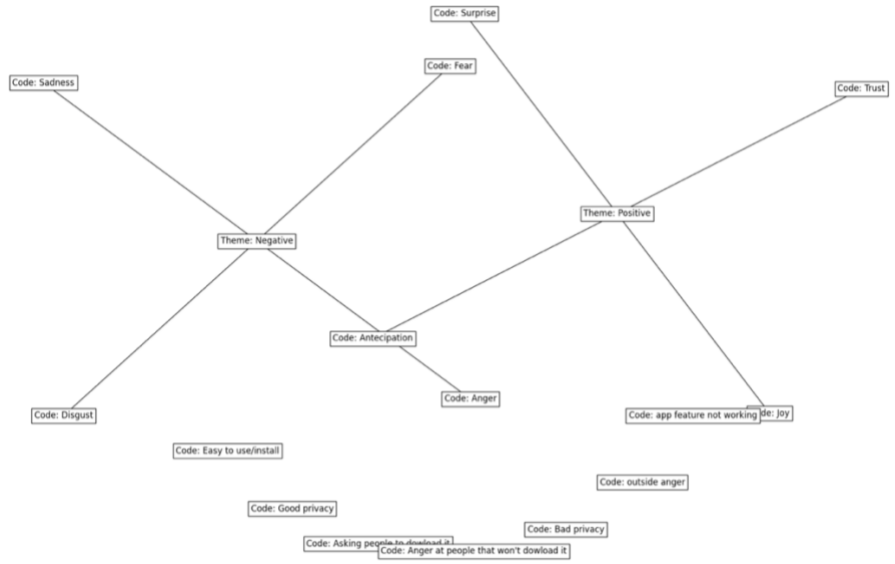


Figure 19: CTA mind map.

The software places emphasis on mapping out connections and their frequency, serving as a mixed method software by combining both qualitative (connections) and quantitative

(frequency). The size of the slices and the word cloud (Figure 19) are the elements that display frequency counts, and the lines that connect the slices are what shows connection.

This combination of elements can foster agency through narrative by creating a narrator out of the connections made by the data visualization. For example, the thickness of the lines tells the story that more people that talk about topic 3, also talk about topic 12 over all other topics. With this frequency of connection mapped out, the researcher can then begin to make inferences about reasons and motivations, crafting a narrative out of the data.

The CTA offers a limited amount of editing of the codes generated by the AI. It is possible to merge or unmerge the themes generated, but the user cannot separate a code into two, or create new codes from scratch. This limitation is detrimental to the autonomy of the users as it limits their choices and possibilities of action, both elements that are instrumental for fostering agency.

Additionally, the software does not provide any explanation as to how the AI came to the results it did. Similar to ATLAS.ti, the results are presented with a lack of context and explanation about the machines' processes. This can affect the fostering of agency by removing the sense of autonomy of the user relating to their own research.

When it comes to navigation, CTA allows for some clickable options that reveal more data. By clicking on the topics, a word cloud appears sharing the frequency of the specific words inside of that topic. Another option is to click on the slices and make the connections appear and disappear. These connections are organized by color-coding so the user can keep track of where the connections stem from. These elements assist in making navigation easy and can come to foster agency through the mastery of space, by making it intuitive and simple.

However, this color organization does not happen for the word clouds, which have colors, but they do not directly relate to the rest of the data visualization. For example, a word can be colored the same as a slice but have no relation to it. This lack of consistency can be confusing and hinder agency as it makes the user spend time and effort to figure out how the color-coding works.

Another element that assists in navigation is the CTA's ability to open customization windows on top of the main window (Figure 21). This allows for the user to edit the data and make changes without being taken away from that main window, facilitating the navigation by not taking the user out of the main screen.

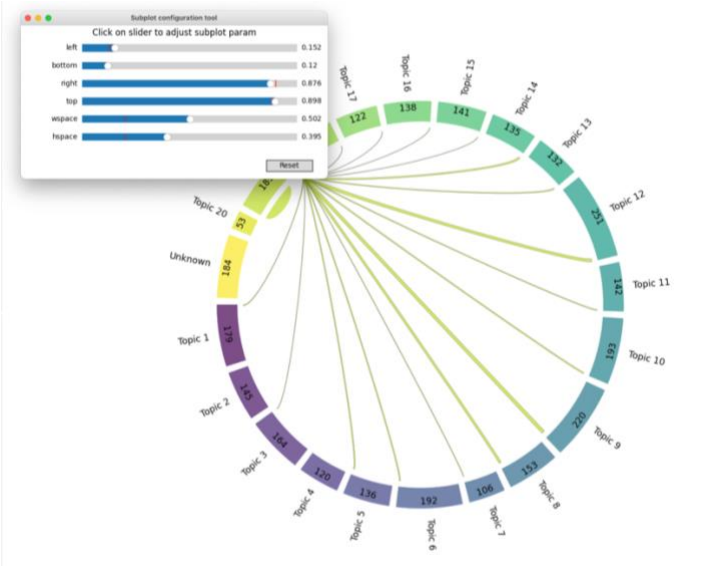


Figure 20: CTA nesting windows.

However, the manipulation of the data visualization itself is limited. The feature for zoom in/out does not work consistently and it depends on the user to select a section of the data visualization for the zoom to be applied to, instead of the gradual scroll zoom that most software uses. There is no way to move the visualization from one side to another other than opening a separate window and adjusting the sliders inside (Figure 21). Additionally, there are limited interactions the user can do to the visualization, only being able to click to reveal the word clouds and the connections. Any further editing is impossible.

This hard to manipulate features and limited editing capabilities hinder agency. The former does it by limiting the choices and actions that can be taken by the user, and the latter by making moving in space complicated and frustrating.

4.3.4 Scholastic

Scholastic is the only software in the sample to not be completely out. However, there are video demonstrations of how it will work in the future, and those display how the data visualization could be used to conduct the analysis.

It only has one option for a data visualization, which is a grouping map, separating each datapoint into clusters organized by the AI (Figure 22). The user can peruse through the clusters and see the individual elements inside each theme.

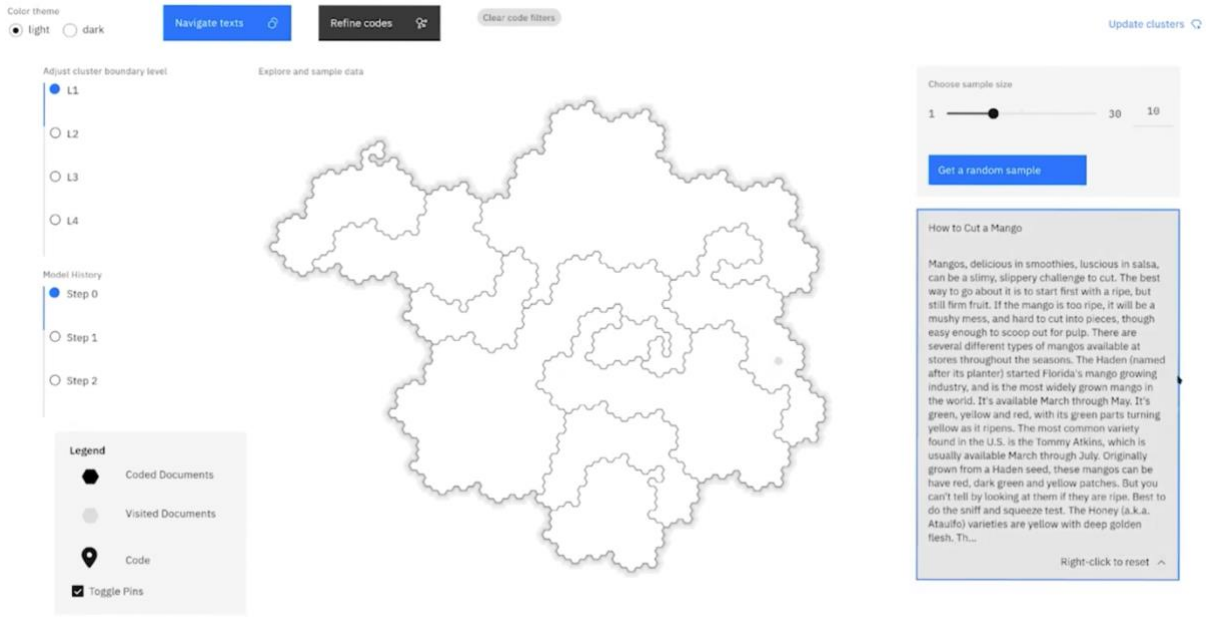


Figure 21: Scholastic grouping map.

Scholastic allows for the user to edit the coding made by the AI and the parameters the machine has used to arrive at the conclusions and then run the thematic modelling again. It also keeps track of each of the iterations of AI modelling (the model history window) so the user can go back and forth and see the changes between different kinds of coding.

Other kinds of filtering the user have access to is changing the sample size (scrolling bar, or choosing a random sample), and adjusting the cluster boundary level. The latter lets the user increase the number of clusters and create more detailed divisions between the sample. An example of creating more clusters using this filter can be found in Figure 23. These editing capabilities and the explanations the software provides about how the AI gets

results were both considered to be elements that fostered agency through allowing the research to feel autonomy over their own research and provide them with more choices and actions.

Figure 22: Scholastic creating smaller clusters.

Navigation capabilities were more complicated to assess since the software could not be experienced first-hand. However, the capabilities displayed by the video tutorial showed that the software makes use of nested windows to reveal more information and stick to using only the main screen to reveal most of the information, making it easier for the user to find themselves and fostering agency through the mastery of space.

4.4 Final observations from sample analysis

Through the analysis of the sample, it was possible to map out some of the elements that could come to foster agency in the users. Those elements were divided into those that fostered agency by the four main kinds of mastery: narrative, choice, action, and space. Each software in the sample had different elements that supported or hindered different kinds of mastery, all explained previously. Not all elements were present for all of the software in the sample. Table 3 and Table 4 have a summary of these elements by software.

		Sample of software			
Agency	**Elements that foster agency**	CTA	NVIVO	ATLAS.ti	Scholastic
Narrative	Proximity to establish connection			Codes applied to the same passage of text	
	Focus on showing connections	Use of linking diagrams and statistical linking diagrams	Use of linking diagrams		
	Use of frequency as storytelling tool	The use of word clouds and a statistical linking diagram	Use of bar graphs, treemaps, sunbursts, and tables	Use of bar graphs, pie charts, and treemaps	Use of clusters of similar information
Choice	Editing capabilities	Limited ability to merge and unmerge themes		Ability to edit the codes made by the AI	
	Explanations of AI processes				Tracker that displays the AI process
Action	Filtering	Filtering of words and themes	Filter by codes, files, words, and connections	Filter information in the transcripts	Filters change the size of the clusters and which version of AI work the user wants to see
	Ability to reveal more data	Click topics to reveal word clouds. Click or hover over clusters to see connections	Clickable options reveal more information in some datavis	Codes and datavis are clickable and enable edition	Click to change size of clusters. Click each data point to get more details
Space	Color coding	Color coding is used to connect clusters of words	Colors unite codes, guide the eye of the user to what is important, and	Color coding is used as navigation assistance	

			help determine frequency		
	Nesting different windows	Windows show up around and on top of the datavis, never taking the user to another page	More information and actions open on pop-ups if the user clicks on certain elements	Editing codes happen in pop-up windows, minimizing taking the user out of the main interface	
	Breadcrumbs		Breadcrumbs guide the user back to the initial data visualization when the user is exploring		
	Clustering of similar elements	Clusters are created, when clicked word-clouds belonging to that theme are revealed	Clustering oh themes so user can see their connections		Clusters of data belonging to the same theme

Table 3: Elements that foster agency.

		Sample of software			
Agency	**Elements that hinder agency**	CTA	NVIVO	ATLAS.ti	Scholastic
Narrative	Over focus on quantitative data		Some visualizations focus only on quantity (i.e., graph)	Visualizations are more suited for quantitative analysis. Focus on frequency over narrative	
Choice	No editing capabilities	Themes given by the AI have limited editing possibilities		Free version of software does not allow for the creation of new codes	

Action	Lack of explanation of AI processes	Tool does not provide explanations for the results given by the AI		Tool does not provide explanations for the results given by the AI	
	Lack of guidance		Lack of proper guidance on many of the software's features		
Space	Inconsistent and meaningless colors	Word clouds has meaningless color-coding compared to the clusters		Colors have no apparent meaning or connection	
	Lack of zoom in capabilities	Zoom in is only available by a tool that is not the most used form of zooming on digital platforms	Zoom is inconsistent across datavis and does not always work. It is not clear which datavis supports zooming features and which do not	Zoom in not available	N/A
	Inability of manipulate data visualization	To move the datavis in space a separate window needs to be opened and with sliders for adjustment	Movement is with scrolling bars, which make it hard to manipulate properly, especially when the datavis is too large	There are no click and drag features	N/A

Table 4: Elements that hinder agency.

Through the analysis of this sample, a series of guidelines was compiled for the development of the first prototype. Those were:

- Use proximity to establish connections.

- Include data visualizations that focused on showing connections.
- Allow for editing and personalization.
- Explain the process behind the AI's results.
- Include filtering tools.
- Use consistent color-coding.
- Use nested windows that reveal more information upon interaction.
- Use easy zoom in/out features (e.g., two fingers on trackboard, or a toggle bar).
- Allow for easy manipulation of data visualization (e.g., click and drag, move with two fingers on trackpad, etc.)

In combination with the guidelines taken from the results of interview 1, the prototype was developed. The agency that was mentioned during this chapter and the one that will be used for the prototype pertains to the agency people can perceive through interacting with the data visualization tool.

Chapter 5
Requirement Gathering

A semi-structured interview to determine what researchers need and want for a tool to help with their qualitative data analysis was conducted with 5 participants. The goal of it was to determine 3 aspects of data visualization interaction:

1. How do qualitative data researchers use visual elements to assist in their data analysis?
2. What visual elements do they find important to have in a data visualization?
3. How do qualitative researchers perceive their sense of agency while doing research?

The interview was done with the purpose of understanding the user and task requirement [58]. The sample of participants was gathered via purposive/snowball sampling seeking information power [65]. Information power is a theory that helps define sample size in qualitative research based on 5 characteristics, each with two polar possibilities: study aim (narrow or broad), sample specificity (dense or sparse), established theory (applied or none), quality of dialogue (strong or weak), and analysis strategy (case or cross-case) [65]. Those will be further explained below.

For study aim, broad studies require larger samples, while narrow require less participants because of the specificity of the situation. Sample specificity refers to how specific the experience being researched is. It is more beneficial to gather different and conflicting opinions than to have a larger sample of similar data. Established theory means that studies with more applied theories in the background need less participants than those that use no theories. Quality of dialogue pertains to the people collecting the data from the participants. If they are more experienced and able to gather the necessary information efficiently than less participants are needed. Lastly, analysis strategy dictates that more exploratory research requires a larger sample, while more descriptive research need less participants. Figure 24 better describes the relationship between each topic and its relation to sample size.

< higher information power

Narrow	<	AIM	>	Broad
Dense	<	SPECIFICITY	>	Sparse
Applied	<	THEORY	>	None
Strong	<	DIALOGUE	>	Weak
Case	<	ANALYSIS	>	Cross-case

Larger Sample size needed >

Figure 23: Information Power (Malterud et al. 2016).

This project has a narrow aim since the situation being observed is highly specific. It also utilized previously established theories to conduct data collection and analysis, as well as be descriptive in nature. The data was be collected and analyzed by researchers that have experience in dealing with qualitative data, therefore the sample sought higher levels of information power instead of opting for a comprehensive large sample. To guarantee more variety, the participants were selected with purpose for trying to gather a wide array of different experiences. For this, multiple labs and areas of healthcare were selected.

5.1 Recruitment criteria

For all interviews, participants were recruited through university and personal connections. The sampling was collected through snowballing from several different laboratories in the University of Waterloo. Participants were qualitative researchers with limited to no experience with programming, and different levels of interest and experience in using AI for their data analysis.

Five participants were recruited in total, each from a different laboratory, and the five completed both instances of the interview. All were in the field of healthcare, but had varied specialties, and different levels of qualitative research experience and experience with data visualizations.

Participant 1 studies vision science and its health service delivery, she reported being new to qualitative research and showed moderate experience with datavis during the interview. Participant 2 works with social media, ethics, vaccine opinions, and trust. She

reported a high experience in qualitative analysis, as well as moderately high experience with datavis.

Participant 3 studies the use of technology and IoT (Internet of Thing) in the care of older adults, she reported high experience in qualitative methods and using datavis. Participant 4 works with gerontology and the care of patients with dementia, she claimed to have moderate experience with qualitative methods and low experience with datavis. Lastly, participant 5 studies the experiences of aging and older adults belonging to ethical minorities, she reported moderate experience in qualitative methodologies and low experience with datavis.

5.2 Interview protocol

The interview protocol consisted of 8 open-ended questions and had the goal of assisting in understanding how qualitative researchers use data visualizations to assist in data analysis. The questions in the protocol were meant to make the participant share what their process for conducting data analysis is, what kind of data visualization they are familiar with, and what software (if any) they use.

The examples of data visualization were taken from Von Engelhardt [47], with some changes that apply to the situation at hand of conducting research. Von Engelhardt separates data visualizations into primary and hybrid modes of visualization, in which the hybrid is a combination of one or more primaries. For the purposes of this project, this separation will be classified as: maps, diagrams, table, and charts, with each having subdivisions, totaling 10 kinds of data visualization.

Maps are visualizations that anchor their meaning in the physical space they represent. They can be divided into statistical maps, path maps, and statistical time maps. Diagrams are visualizations that derive meaning from the interaction between its visual elements. They are divided into link diagram, grouping diagram, chronological link diagram, and statistical link diagram.

Tables comprise of organizing information into rows and columns and establishing relationships between elements based on them. Lastly, we have charts, which derive their

meaning from its metaphorical connections to the real world. It is divided into two kinds: time chart and statistical chart.

Additionally, the interview had the purpose of collecting sketches from the participants of what they believed their ideal data visualization tool to look like. This was done following user centered design principles to include the participants during the design process [61]. User centered design has been notably a successful method in developing design objects for a specific population [62], therefore we decided to employ some of the same techniques to ensure our prototype sufficed the needs of qualitative researchers. The full interview protocol can be found in Appendix C.

The results from both the sample analysis and the interview were then compiled into guidelines, which were used for the next phase. The interview specifically was conducted with the goal of including the users in the design method, which has better results than designing alone [66].

This phase satisfied the objectives of 1 (determine how qualitative researchers use visual techniques to make sense of their data analysis), 2 (compile what visual elements qualitative researchers find important to have in their data visualizations for qualitative analysis), 3 (determine what kinds of data visualization satisfy which needs of the qualitative healthcare researcher community), and 4 (determine how qualitative researchers perceive their sense of agency when doing qualitative research).

5.3 Results from interview

The results from the interview were divided into 2 parts, data visualizations being used as cognitive assistance and how researchers perceived their own agency when conducting research.

5.3.1 Data Visualizations as Cognitive Assistance

There were three kinds of cognitive assistance that data visualizations could provide according to the participants: those that helped them figure out their own data, those that served as knowledge translation tools for third parties, and those that were a mix of the two. Participants cited using data visualizations as cognitive assistance for themselves when they were either organizing or exploring data. Specifically, when organization was the goal,

participants mentioned 7 separate needs: checking frequency, making connections, comparing, tracking changes over time, grouping, seeing processes, and mapping out paths. Often participants mentioned more than one need at the same time, highlighting how they used the same kind of visualization for more than one purpose.

When it came to using it as knowledge translation tools, participants highlighted their importance for presentation, and summary. Additionally, two uses of datavis were mentioned to be used both for personal and third-party cognitive assistance. Those were the datavis meant for assisting in collaboration and supporting narratives.

In terms of presenting and summarizing, participant 1 shared her preference for using Microsoft PowerPoint tools to make diagrams instead of presenting them in plain text. According to her, presenting information in a visual way instead of plain text is more engaging for her style of presenting information. She shared *"I'm obsessed with smart art chart designs when I'm doing presentations. Obsessed! like instead of typing all my speaking points in just on the slide or whatever, I just love putting it in a diagram or a hierarchy."*

Similarly, participant 3 highlighted the knowledge translation capabilities of data visualization, especially when it came to presenting information to industry partners that do not have the same grasp on research as the researchers. She shared *"That's the big thing, especially I think for me, since I work with a lot of industry partners and… things that would be very of interest to an academic, such as theory and, you know, precise methods, they're really not concerned with it because they don't have background in it"* when explaining how data visualizations could be used for the kind of qualitative research she does. She frames this need as *"the story in action",* in which the researcher needs to present industry partners with visualizations that tell a narrative strong enough to enact change, answering questions like *"what's the point? What can I do with it? How do we act going forward?"*. Another participant voiced a similar experience, saying she likes using visualizations to make her research more digestible and compatible with the narrative she wants to highlight.

The use of datavis as a narrative tool was expressed by all participants, both as a something they would use for the data they generated, and something they would add to research rationale to weave a story. When it came to using narrative elements as an

organization tool, two participants talked about how they made use of mind maps and flow diagrams when conducting data exploration to identify the story the themes in their interviews were telling, and thus help them organize their ideas and frame their research before starting the process of writing. For example, participant 4 illustrated this point by saying that she works with data visualization by *"having the data speak for itself and seeing like what kind of story it tells and then highlighting that accordingly."* Similarly, participant 5 also reported using data visualizations to find a narrative when organizing themes for her thematic coding of qualitative interviews, she quotes *"I try to make a story, like there should be, like a theme that makes sense."*

As for using data visualizations for augmenting a narrative outside of their data analysis, three participants talked about using data visualizations that were compatible with the narrative of their rationale and taking advantage of the storytelling abilities of datavis to convey it. For example, participant 4 talks about how she likes using data visualizations *by "feeding it into my literature, review, my argument, my rationale, in order to justify my topic,"* and therefore creating a narrative. Similarly, participant 3 talks about using data visualizations as augmentations for the text so that the reader can more easily process the story, saying *"in a glance with meaningful line weights and colors, that creates a lot more meaning."*

Another part of data visualization being used as a summary was the importance of framing any summary as novel information. Participant 3 discusses her belief that *"the role of data visualization more than anything else it as being a way to enhance the text and… they should be able to tell you something that you could not glean just by reading 6 paragraphs on a page,"* and expressed her desires for data visualization that not only presents information but also assists in framing a narrative. She gave the example of sharing the total number of a migration versus presenting a statistical path map that shows people the magnitude of what the number means. Similarly, participant 1 also voiced her desire for visualization that preserved context and narrative when reporting findings. She explains that *"I think you don't just want the words you want the context that it existed in the conversation."*

Aside from narrative, another topic mentioned was the use of datavis as a tool for assisting in collaboration. One participant mentioned how she would use the mind map data visualizations developed to organize themes to present information to her colleagues and supervisor and get their feedback and approval before continuing with the rest of the data analysis.

Additionally, three participants mentioned their frustration with the software NVIVO for limiting their collaboration efforts. All three participants complained about the lack of integration in NVIVO and shared having to use Microsoft Word to collaborate with other colleagues, which made them lose many of the features that NVIVO has to assist in qualitative research. For example, participant 3 talked about losing the ability to quickly open coded sentences and check the context around them when working with Microsoft Word. She voiced *"So that's the very annoying low-tech way. It's a lot harder to search. It results in a very long difficult to navigate document and it's a lot less convenient."*

When it came to using datavis as cognitive assistance for themselves, four participants talked about their process for organizing their thoughts in the middle of data analysis, and shared sketches that they used to organize, explore, and make connections. One of those three participants made use of mind maps, while two others preferred to use grouping diagrams. The fourth participant used both for her research, but for different purposes.

Participant 5 cited that *"I kind of, like, go and look back, you know, to the mind map and then. And I can reconnect it, you know, cause it's easier for me to. […] It makes more sense to me on paper,"* as a reason as to why she prefers to use a visual approach when she is conducting data analysis. Figure 25 shows the sketch she shared during the interview in which she divided the main themes into sub-themes and then subcategories. The green line was used to represent the fact that those themes are related in time and the ones happening on the left must happen before the ones on the right. According to her, being able to visualize all data at once helps to organize her thoughts and explore data in a way that makes sense to her. As she quoted *"if I can visualize it, I can easily understand it properly. So, like the web of chart like it makes sense to me."*

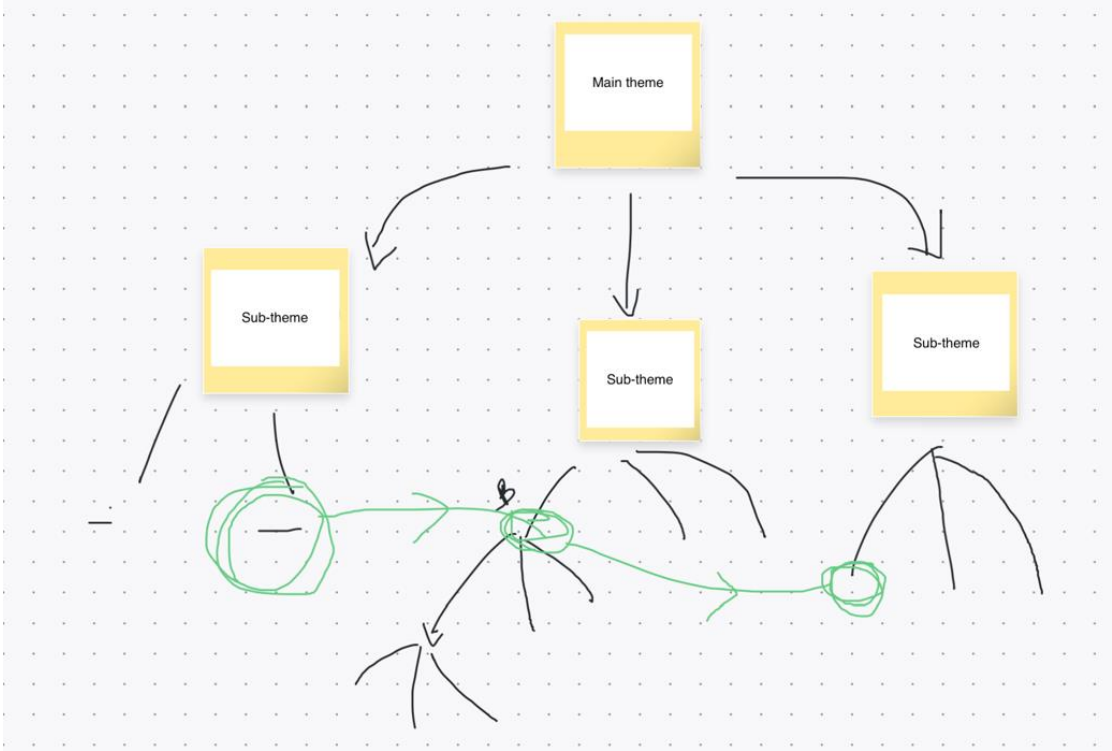


Figure 24: mind map sketched by participant 5.

Similarly, participant 4 shared a sketch of a mind map and talked about her process for the brainstorming phase of data analysis. Like participant 5, she also expressed her preference for organizing her ideas visually before converting it to text. She shared that *"That's how I get things out of my head so that I'm not just juggling them in my head so that they can be clear".* In her sketch (Figure 26) she explained her process of adding themes and connecting points to the same visualization and then mapping out their connections and relationships, which are represented in the sketch by the purple lines.

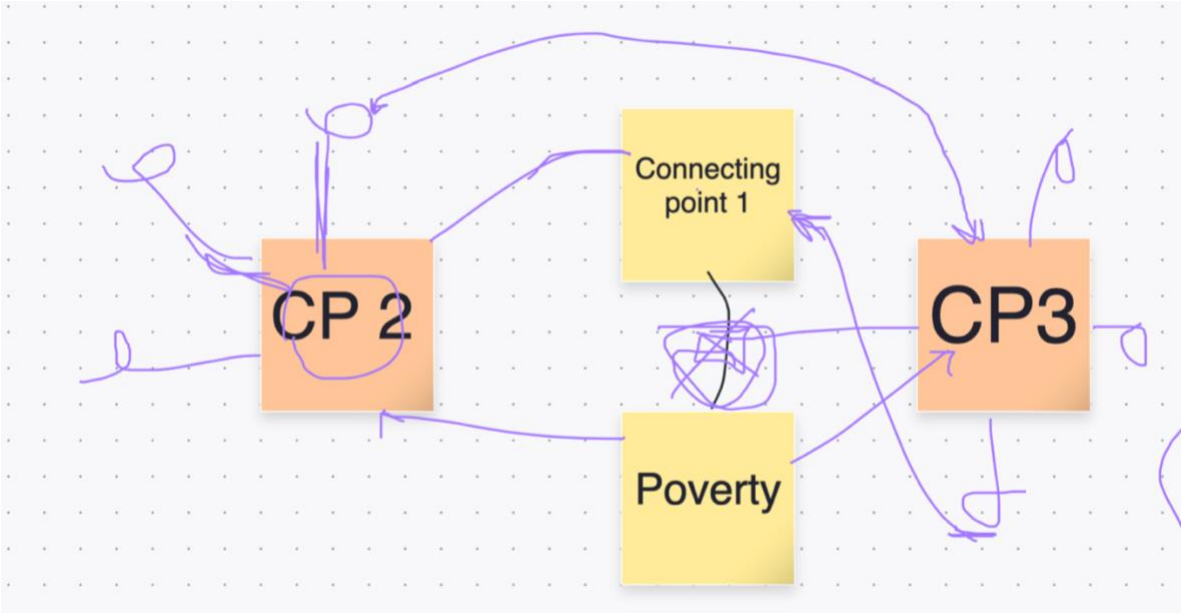


Figure 25: mind map sketched by participant 4.

Additionally, the same participant shared another sketch of how she uses datavis as an organization and exploration tool to come up with the themes for her analysis. In the sketch (Figure 27) she explained that each shape represented an idea from the transcript, and that by placing them in a board she could move them around and cluster them together based on similarities.

Those clusters, here represented by the purple lines, could then be transformed into the main themes. According to participant 4, having all ideas in front of here in a way that she could easily pick up and move it somewhere else helped with figuring out what belonged where. As she put it *"and as you begin to move things around, like maybe you decide that this person […] maybe they don't, like, belong here. Maybe they deserve to be. Maybe they deserve to be with this group over here."* Through this categorization, she reported being able to properly organize information into "little boxes, so then you can like properly connect them and write."

To further organize her thoughts, participant 4 also shared her preference for using colors, saying that even though *"the whole thing would be like color-coded so many colors everywhere it's messy, it looks ugly, but it made sense to me, and it looked… I could understand it and makes sense of it."* Her preference for color coding can be seen in Figure

26 and Figure 27, where the participant switched colors to sketch different elements depending on the organization she narrated through the interview.

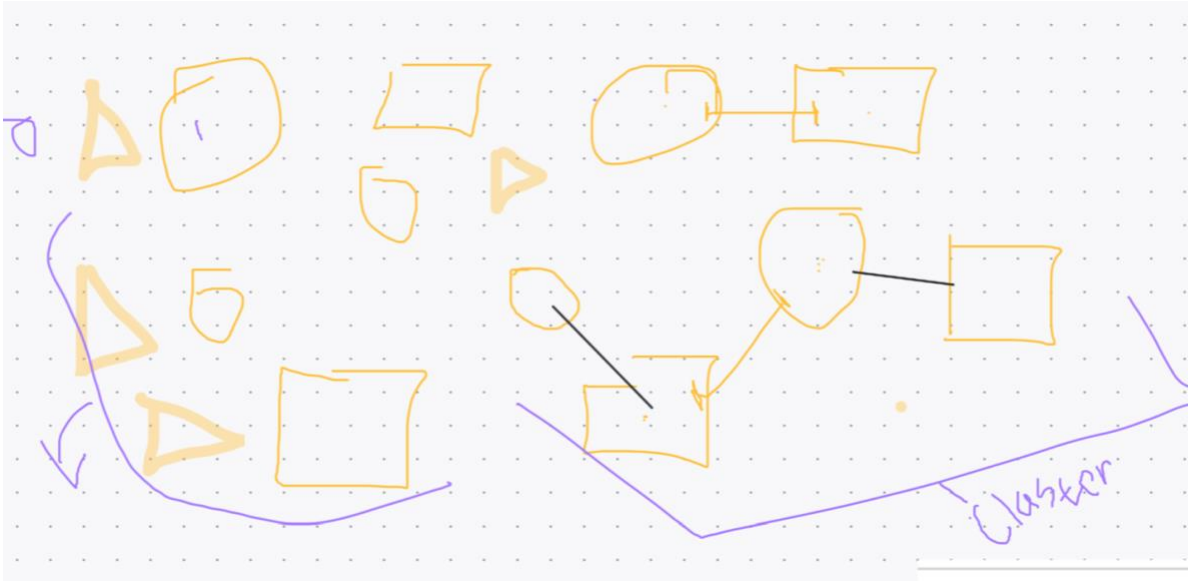


Figure 26: grouping diagram sketched by participant 4.

Similarly, participant 2 also used grouping diagrams to create clusters for her ideas and come up with the themes for her analysis (Figure 28). She quoted *"I write things out all the time. Like that, like in diagram form and tables. And that's how I figure out themes".* According to the participant, she usually uses tables as an organization tool, separating them into a column for explanations, themes, and codes. Then she puts each code into one of the buckets for themes until there's no left. If there are still codes left at the end that do not fit any of the themes, then she uses the table to re-organize her ideas until every element fits into their proper space. According to her, this way of organization makes sense to her as a table and *"that's just the only way I think you can visualize it in my mind, because that's the only way that makes sense."*

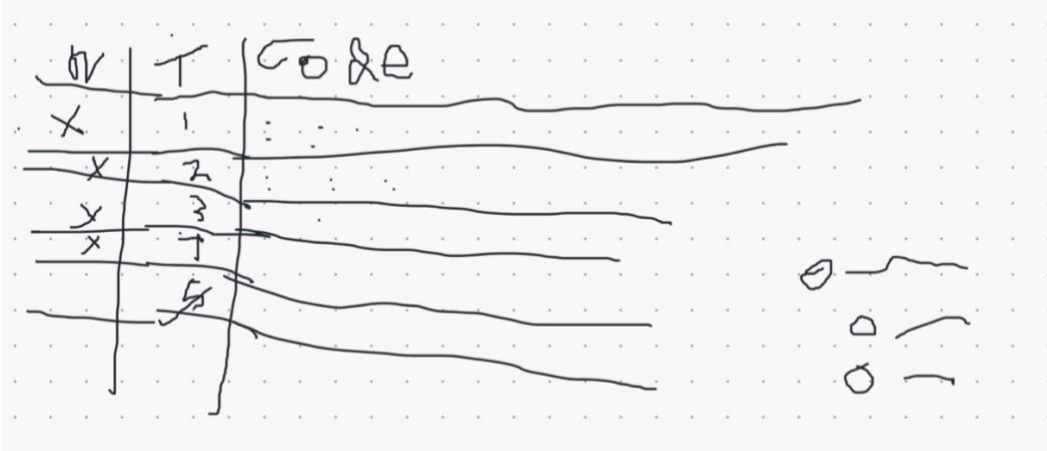


Figure 27: grouping diagram sketched by participant 2.

Another way of organizing thoughts in space came from participant 1, who shared the sketch in Figure 29. Since her research is very dependent upon processes and barriers, she explained her desires to visualize data organized in a nested grouping diagram in which each layer would be a different category of connected barriers.

The purpose of this representation was to illustrate the barriers around the goal, which would be placed right in the center of the visualization. Alongside this way of visualizing data, she also voiced her desire for making this kind of visualization quickly and with ease, saying that *"and sharing the considerations that need to go into that theme that are unique to the context, so being able to make something like that… efficiently!"*

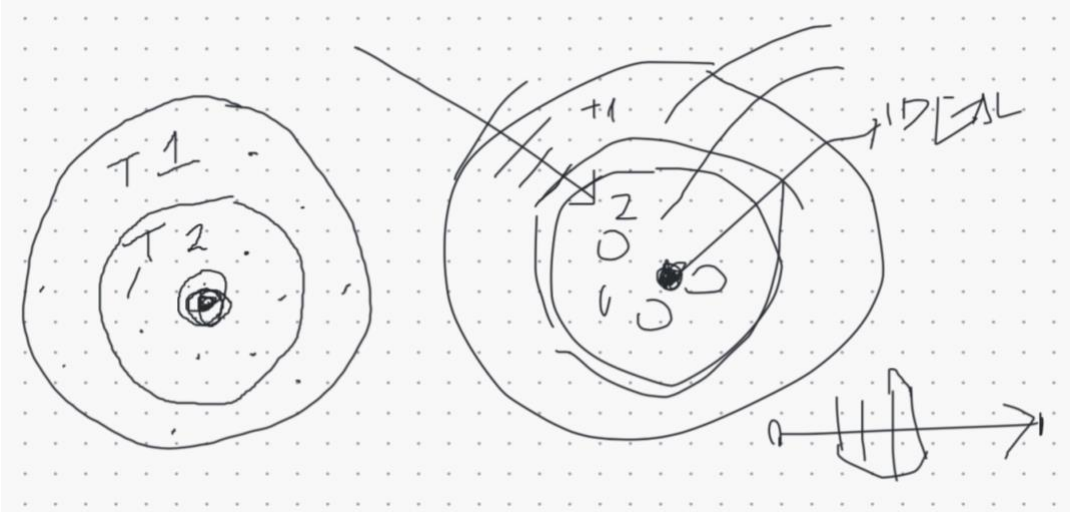


Figure 28: grouping diagram sketched by participant 1.

Another kind of organization that the participants voiced using were comparison data visualizations. Participant 1 shared her frustration with the tool provided by the software she chose to use to analyze her data, NVIVO, and making datavis that showed comparisons. According to her, her needs would be for *"comparing the differences in the qualitative data between two groups is what I would like a visualization to show me,"* but with the tools offered by NVIVO she finds herself often frustrated and lost, *quoting "so, like if the if NVIVO could easily tell me I… I haven't like quickly found a resolution to being able to see like I get overwhelmed when I'm trying to see which codes are… occurring in different… [locations], or the interviews from people from different [locations]."*

Another participant mentioned using comparisons in datavis when assessing the differences and commonalities in stakeholders. She quoted that *"part of that [finding the commonalities and differences in stakeholder groups] was finding like where their views contrasted, or how they approached issues differently"* and said that she would probably not use those kinds of visualization for publication, but instead for organizing her own thoughts when conducting research.

Four out of the five participants mentioned using data visualizations for checking frequency in their data. Participant 1 talked about how she appreciates having a software that can check what themes come up the most without having to do it manually. She commented on using the frequency features in NVIVO to check frequency and extract from that what themes were the most important to her participants. She quoted *"So, looking at how much space and the percentage of coverage of different themes is really helpful. I think to be able to like… actually get the take home messages of, okay, what is important to this population of survey for research participants?"*

Similarly, participant 4 also voiced her interest in using data visualizations that showed frequency when mapping out how common certain experiences are, but with the caveat of not ignoring a theme even if it is not being mentioned by all participants. She explains that *"as themes like emerge, if there's a theme that was highlighted by most participants, but not all participants, you're still going to talk about it in your findings. It doesn't have to be highlighted by every single individual. And so maybe if you wanted to see, OK, how common*

is this particular experience?" when talking about her personal preferences for using frequency checks in her data analysis process.

Participant 3 also talked about her preference for having visualizations that allow for easily checking of frequency, especially when it comes to content analysis. She voiced *that "it's a good idea to look at the data at a glance, especially if you're doing something like a content analysis where the frequencies of certain data point is of interest."* However, the same participant also showed concern over being forced to use graphs that show frequency in places that she believed did not fit the narrative. She gave the example of being asked to include a bar graph for theme prevalence, which she felt did not contribute to the reporting of results.

Participant 2 was the only one that mentioned not having much use for frequencies in her research. According to her, she prefers to steer away from graphs that display frequency as she never feels like she has enough data to make meaningful statistical data visualizations and that counting is not compatible with the kind of data analysis she does. She explained that *"I usually steer clear of like graphs and stuff because… I've been told before during peer review processes that because I am not using quantitative data in those… projects…. I can't visualize my data like that like I can't provide counts because counts don't have impact on the data."* However, she has also voiced a high interest in frequency if it comes with connections, such as in a linking diagram. With a combination of the two, she expressed that she would be able to both draw connections and see how prevalent these connections are, which is more in line with her kind of research.

Similarly, participant 3 also voiced a preference for using frequency data visualizations coupled with another kind of data insight. In her case, she explained how using it paired with the passage of time often garnered more interesting results. She shared that *"If you're able to do a histogram like this as opposed to a bar graph that is going to give you a lot more novel information at a glance,"* when explaining her preference for a combination of the two.

Passage of time was also a need that participants voiced, regardless of whether it was connected to frequency or not. Two participants shared how their main research contains heavy chronological elements. One of them shared *"I think about the ways that my, even, my qualitative research is chronological because I'm really discussing the step wise process*

[details of project]," as well as her overall preference for using line graphs for representing changes across time because of personal preference.

Participant 5 expressed similar feelings, realizing in the middle of the interview that her research fit the category of a chronological linking diagram, saying *"Now that I'm thinking about it like my research is basically a linking diagram depending on time."* Interestingly, this participant also voiced her interest in learning more about data visualizations because of her limited experience with them and low visual literacy, saying she would seek to know more about different ways to present the data for her research now that she was aware of the possibilities. She shared that *"I think I'm restricted with like linking, you know, diagrams visualization cause that's how I have always envisioned qualitative research to be. Now that you're showing me so many types of visualization, I'm like oh damn! Does that even exist?"*

Other participants shared their experiences needing data visualizations that depicted the passage of time for several needs, such as documenting changes, presenting it to industry partners, making decisions, and show processes. Specifically for process, participant 3 voices how data visualizations play a necessary role in communicating trajectories, decision trees and the continuous nature of what she researches. One of the biggest needs she voices was mapping out the trajectories of her participants in the healthcare system, and having a datavis to assist with that is valuable in her opinion. She quoted *"so having a clearer idea of what care trajectories look like. You know what? What settings they move through. What clinicians they engage with. That could be a helpful visualization of that."*

She shared a sketch during the interview explaining how she prefers to use data visualizations to depict processes (Figure 30). In the sketch she separates the tasks into branches and uses connections and hierarchical relationships to explain what must be done by who and in what order.

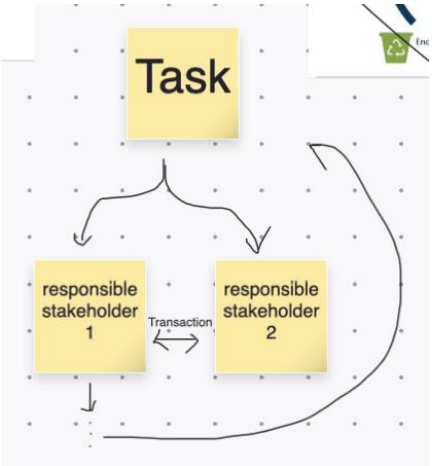


Figure 29: process diagram sketched by participant 3.

One final use the participants shared to being aware of for data visualizations was of paths. Participant 1, in particular, talked about how her research investigates the paths that long distance healthcare providers take when travelling to remote communities and how she has used maps to present data about her research before, both to academia and to her industry partners. Three other participants said they had no personal need for data visualizations that followed paths but were aware of other researchers in the field of gerontology and dementia that required them constantly.

Often, participants mentioned doing more than one thing at a time with the same kind of representation. For example, participant 4 talked about how she would use the same mind map to connect ideas, group them into themes, and then find the narrative that she would like to pursue in her research. Participant 5 also echoed the same feelings of using one visualization for multiple purposes, quoting *"I organize my thoughts in like a little notepad. I tried, for example, I try to make a story, like there should be, like a theme that makes sense, right?"* which implies she organizes, explores, and searches for a narrative in her data all using the same visualization.

Additionally, there was a lot of personal preferences and biases that the participants showed during the interviews, which they attributed to the ways they preferred to visualize information. For example, participants 5 and 3 showed a strong dislike for tables, quoting *"I don't know, when I see the table like 'nope!', this is not how I, like, my mind works,"* and *"You know, it's like I question sometimes how effective [tables] are as a data visualization*

tool”. On the other hand, other participants voiced their love for this kind of data visualization, for example, participant 5 shared *“I love tables. I think tables are always useful when it comes to presenting data like qualitative or quantitative kind of data. Or just like planning things out.”*

Participant 3 also shared her preferences for statistical linking diagrams, talking about how she found them to be the most useful when he is exploring her data and coming up with connections. According to her, she likes to use this kind of visualization coupled with AI to see what are the possible connections that the machine suggests and then look out for them when she is doing her own coding. She quotes *“[the statistical linking diagram] is a perfect way to visualize it and then it shows me: oh well, that one is connected to this big piece and this, this one is, can I like…. that to me is the best one.”*

Preferences were also prevalent for bar graphs and pie charts, where participant 3 shared her opinion that *“there is nothing a pie chart can demonstrate to you that a bar graph cannot do more clearly,”* while participant 5 said she *“wouldn't use [bar graphs] unless if it was, if it was required of me to be honest, because I feel like there are better ways to present your information when it comes to qualitative information.”*

In conclusion, the participants reported different uses for data visualizations depending on their needs, here separated into those that are used for providing cognitive assistance for others (presentation and summarization), those that served as cognitive assistance for themselves (grouping, frequency, comparison, connection, passage of time, path, and processes), and those that could be used for both (collaboration and narrative enhancers). In addition to that, the participant showed different preferences for how data is visualized.

5.3.2 Agency

The importance of agency was highlighted by all participants to different extents. Two facets of agency were commented on, which were the elements that made participants feel like they had agency over their research, and the barriers. The barriers were mostly related to the loss of autonomy, freedom, choice, and their lack of trust in machines or software.

Regarding trust, some of the participants voiced a lack of trust in a system that can automate the process of data analysis. This lack of trust mostly came out of fear of

replacement or loss of autonomy over their research. Additionally, others talked about the distrust in automation because of their beliefs that machines are not advanced enough to provide with any useful insights. Both are delegation related issues that stem from different sources, but have the same result, a lack of trust in automation.

Two participants voiced their fear of being replaced by machines and what that meant for their future and their job as a researcher. Participant 4 expressed her fears by saying that she does not want machines to progress to far and make her obsolete, quoting *"I wouldn't want it to analyze anything for me because then I become obsolete, so I don't want it to be too good at its job, where I become obsolete and it replaces me and takes over the world and it turns into the matrix."* While participant 4 explained her concerns over being out of a job once automation becomes too advanced for humans to compete with.

On the other side of the spectrum, four participants commented on the usefulness of using automated systems and if it even can generate useful results. Four participants talked about the richness of qualitative data and the need to preserve context when analyzing it, questioning how a machine could understand this context and how could it generate codes without that understanding. Participant 1 quoted the use of automatic coding as *"very, very bizarre,"* explaining that she did not believe that AI could understand the nuances of the data and therefore would have a hard time generating useful codes.

Similarly, participant 4 explained that the human element of qualitative analysis is very important and that immersing in data is a crucial step to any qualitative analysis that cannot be replaced by machines. Participant 3 echoed the same feeling, saying that *"Naturally, still, you know, being a researcher and knowing the context, being able to read between the lines, there would still be a lot of things that you would need to code manually because you know in a way the machine doesn't know."* She expressed her frustration with AI capabilities, calling it rudimentary and questioning whether it would even be useful for the kind of research it does, since her qualitative analysis is so context heavy and she does not believe an algorithm could pick up on these nuances.

Participant 1 illustrated the same point by saying *"I didn't look at it in that great detail because the codes that it produced, I think they were all single word and codes, which*

wasn't really helpful anyway," about the time she tried to use NVIVO's automatic coding system.

Additionally, to the important of preserving context, participant 4 also raised the issue of regulations, talking about how she wants there to be a balance between AI regulation and censorship. She explained by saying she does not want the machine to grow so smart it will replace her, but also fearing that because of censorship and regulation any results generated by the machine can be biased. She quotes *"Because you want it to be regulated enough where people don't use it for like harm and destruction, but you don't want it to be overly regulated where it conceals the truth."*

These issues raised the issue of AI delegation, which was brought up by all participants. Three participants were open to the idea of using AI, but very strict as to which tasks should be delegated to it and which should remain in human hands. Participant 1 shared how she enjoys having the machine assist in tasks such as counting, checking frequency, and helping her design data visualizations. She expressed the need to have a tool that could assist with the last part, making the creation of visualization easy, especially since her experience with data visualizations is limited.

Similarly, participant 3 shared the same need of using AI as a way of *"cutting down on the work of coding a little bit."* She explained that her ideal partnership with AI would be using it as a facilitator for checking the frequency of very simple sentiments. Since she questions the abilities of AI to fully understand context, she would prefer to keep automatic coding to simple sentiments that machines wouldn't have trouble processing, quoting that *"There's very simple sentiments that people express, like 'technology is expensive', '[name of devices] are expensive'. That's something that would be nice that you wouldn't have to manually tag like just those… those very simple, straightforward ideas."*

Participant 3 was the only one that was more openly receptive to the idea of doing coding with the assistance of AI. She explained that because of the nature of her type of research in working with social media, her samples end up being very large and very short in character amount. Because of this, she explained that coding by hand is simply not possible and if she could employ the assistance of AI to make the process more seamless, it would be ideal. She shared *"I prefer [coding with the assistance of AI]. I don't know how to…. when*

it's 500,000 data points, it's not possible for me to do it. I can do it with 4000 easy peasy, but 500,000 plus. I just don't know if it's possible and so, having the toolkit which would. It can do that for you." She clarified that the assistance of the AI would not negate the need for a human researcher, saying that she still codes herself, but enjoys delegating some of the work.

On the other hand, two participants shared their reluctance towards delegating at all. Participant 5 explained that she views qualitative research as untainted and would prefer it not to *"be in, 'invaded by machines.' I still want it to be in the hands of humans."* She explains that having a human be the one to derive codes to be the beauty of qualitative research and the by delegating to machines she fears losing the empathy and personal connection with her kind of data analysis. Similarly, participant 4 talked about her perceived sense of agency and how much it would be affected if she felt the machine was doing her work for her. She quotes *"I feel like a part of it is the agency element like I feel like. It'll replace me or like it does things very fast that like. You know, it's concerning."* This statement highlights her sense of loss of agency, as well as her fear of being replaced.

Despite the fear and apprehension expressed by the participants, there were also elements that they felt increased their sense of agency that were discussed during the interviews. Those were separated into: freedom of movement, ability to choose, autonomy, and suggestions and guidance.

Freedom to move was partially shown in chapter 4.2.1 by the way the participants talked about using data visualizations and how much emphasis they put in moving elements freely and having advanced editing options. For example, participant 5 shared how she *"wants that freedom to scribble and just vomit everything out and have it be ugly on paper so that my brain is organized,"* as an example of how she likes to have that freedom of movement when she is working with data visualizations for research. Similarly, participants 5 and 2 talked about liking the to use a whiteboard to draw out data design ideas because it is easy to move around elements and sketch concepts.

However, freedom was also discussed in the frustrations participants shared to existing software, which hinders their editing abilities and constricts their choices. Participants 1 and 2 explained their frustration when using the software NVIVO and how it can be hard to figure

out how to get what you want out of it easily. Participant 2 shared that she considers the software to be *"[Not] Easy to figure out and it's so not worth it for me to try and waste my time figuring it out,"* and that she would rather spend her time manually coding the data instead of trying to understand how to navigate through the software.

Additionally, participants 3 and 5 shared their experiences with Microsoft Word and how they had to use it to collaborate with their colleagues, since NVIVO does not allow for collaboration, and the exhausting process that was navigating the document.

Freedom often came related to the possibility of choice, which the participants expressed a preference to. Three participants discussed how they liked being presented with options when working with software, instead of given results that they could not edit. Participant 1 illustrated this point by saying her ideal interaction with AI would be through the machine offering her insights and suggestions, which she could then sort through and choose to accept. She quoted that *"I can see the different ones and then select the one that resonates with me."*

Similarly, participant 2 talked about liking to use AI as an assistant that could suggest possible themes and connections, while still leaving her in charge of whether to keep them or not. Participant 4 echoed the same sentiment, quoting that *"laying those options out for me and giving me suggestions. I think like, that would be very helpful."* She referred to her ideal for a data visualization tool to present her with layout options and guide her through the process of designing a data visualization, whilst still leaving her with enough choice to feel in control of the process.

The use of suggestions and guidance was a consistent theme through the interviews, and all participants mentioned wanting a tool that could offer it when dealing with data visualizations for research. This guidance was mostly mentioned when participants were talking about using a software to help them figure out how best to visualize their data, and what kind of visualization to use for which kind of need.

Participant 4 illustrated this point by sharing how she would like a tool that gave her enough options to guide her, but still allowed her the freedom to choose which option worked best for her needs. She quoted that as ideal tool would be *"trying to see how different people like to see their data and then giving people the option of seeing their data*

in all those, all of those different ways," in a way that still ensures autonomy and plenty of options.

Similarly, participant 1 described her need for a guide that could help her *"determine which software would produce the visualizations to show that data meaningfully,"* when talking about what kind of machine assistance she would be interested in. She shared her lack of experience with data visualizations is perceived lack of tech-savviness, and the belief that because of this she considered herself limited in choosing data visualizations and knowing how to best represent her data. Participant 5 echoed the same feeling, saying that because of her lack of expertise in data visualization she considers her skills to be restricted to only the diagrams that she knows. She quotes *"I don't really know if there's like a specific… because I haven't really like, have experience with data visualization in particular."*

In conclusion, participants expressed two different facets of agency perception during the interviews: barriers and facilitators. Barriers were related to loss of autonomy, lack of freedom to move and edit, limited choices, and lack of trust in the machine. Facilitators were elements that allowed them to move within the software, freely edit their visualizations, ensured transparency between researcher and AI, and provided with guidance and suggestions.

5.3.3 Final observations from interview

From the results of the interview, a few guidelines were developed for the creation of the prototype. Some of these guidelines were repetitions of the ones found during the analysis of the sample of similar software, but others were unique to the interview with qualitative researchers. Overall, the guidelines taken from the interview were:

- Categorizing data visualizations based on the needs the participants described (checking frequency, making connections, comparing, tracking changes over time, grouping, seeing processes, and mapping out paths).
- Giving the users editing capabilities so they can fine tune the visualizations to fit whatever narrative they desire (e.g., editing colors, moving elements, etc.).
- Providing multiple kinds of data visualizations to accommodate the different preferences and needs of different people.

- Allowing the user to move elements within the visualization for better organization and exploration.
- Preserving freedom of movement.
- Consistent Color-coding options.
- Providing guidance for what kinds of data visualization exist and how they could be used.
- Preserving autonomy (e.g., presenting choices and letting the user decide, explaining how the results were reached, etc.).

With that series of guidelines, a prototype was developed for a data visualization tool that could help qualitative researchers work with AI during the data analysis process.

Chapter 6
Development of prototype

Phase 3 was the development of the first version of the prototype following a brainstorming of data visualization ideas, based on the results from the interviews and literature review of the previous phases. Following Santos [60] the brainstorming included the generation of multiple alternatives before the selection and refinement of a final product.

After the brainstorming, a prototype was developed, which included 16 pages of examples of how the software would function and look like. With these templates we returned to the 5 participants that participated in the first interview and recruited them for a second session of user input.

6.1 Presentation of prototype

The first prototype was developed based on the insights gathered from the literature review, analysis of similar software, and results of the first interview. It followed the guidelines summarized in chapters 4.1.5 and 4.2.4.

The prototype was a series of sixteen images that depicted the process of utilizing the data visualization tool. Figure 31 has an example of the first screen of the prototype. The design was inspired by PowerPoint to ensure familiarity, as per the suggestion of participant 1. There are four sections for the prototype, the top bar (colored blue in Figure 32), side bar (colored green in Figure 32), lower bar (colored purple in Figure 32), and workspace (colored yellow in Figure 32). The full set of images can be found in appendix D.

Figure 30: prototype 1, main screen

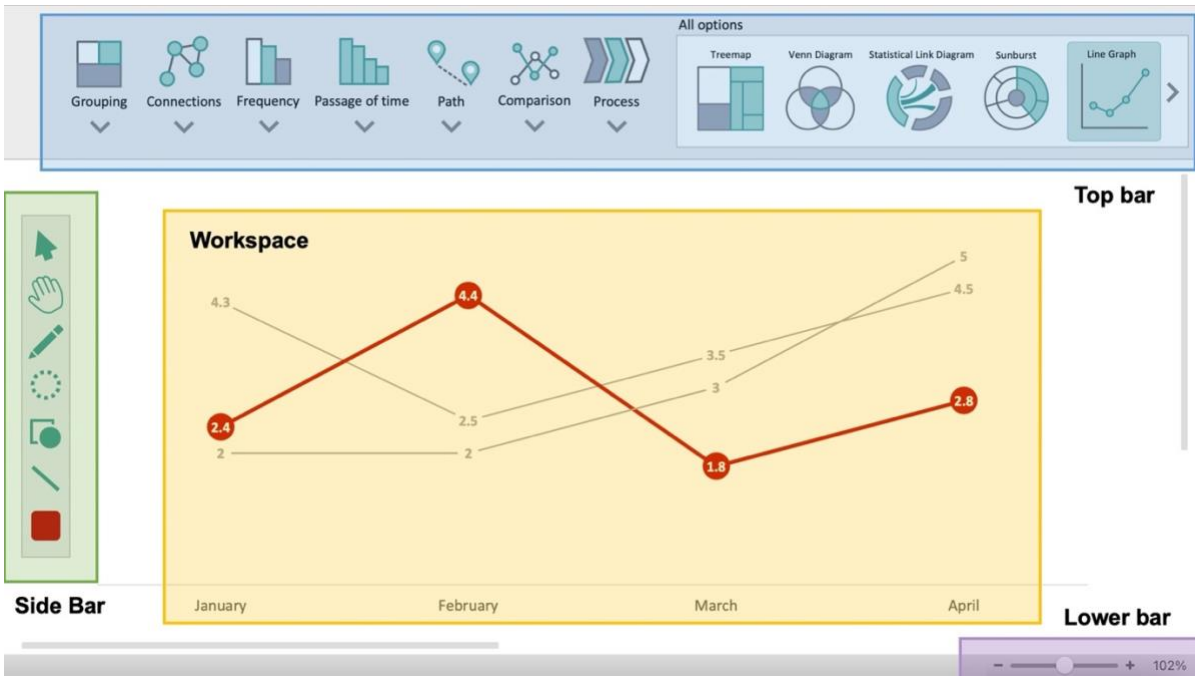


Figure 31: prototype 1, main screen color coded by section.

The top bar is separated into the categories the participants mentioned needing the most, and a section that contains all options for data visualization. Upon clinking on one of the

category buttons, a pop-up appears with all the individual datavis that fit that need. An example of what that looks like for the category ‘frequency’ can be seen in Figure 33.

Figure 32: prototype 1, pop-up with all options for 'frequency'.

Once the user selects an option, the software runs the data input and presents a visualization. Figure 34 has an example of what would happen if the user were to choose a pie chart. The colors are consistent, except for one slice of the pie, which is red, to represent the customization abilities user can have other individual elements in the data visualization.

A few interactions are available for the user to do with the datavis. The first one is an option to hover over individual elements and elicit a pop-up with explanations as to how the AI reached those results. The others are mediated by the side bar, which allows for editing capabilities.

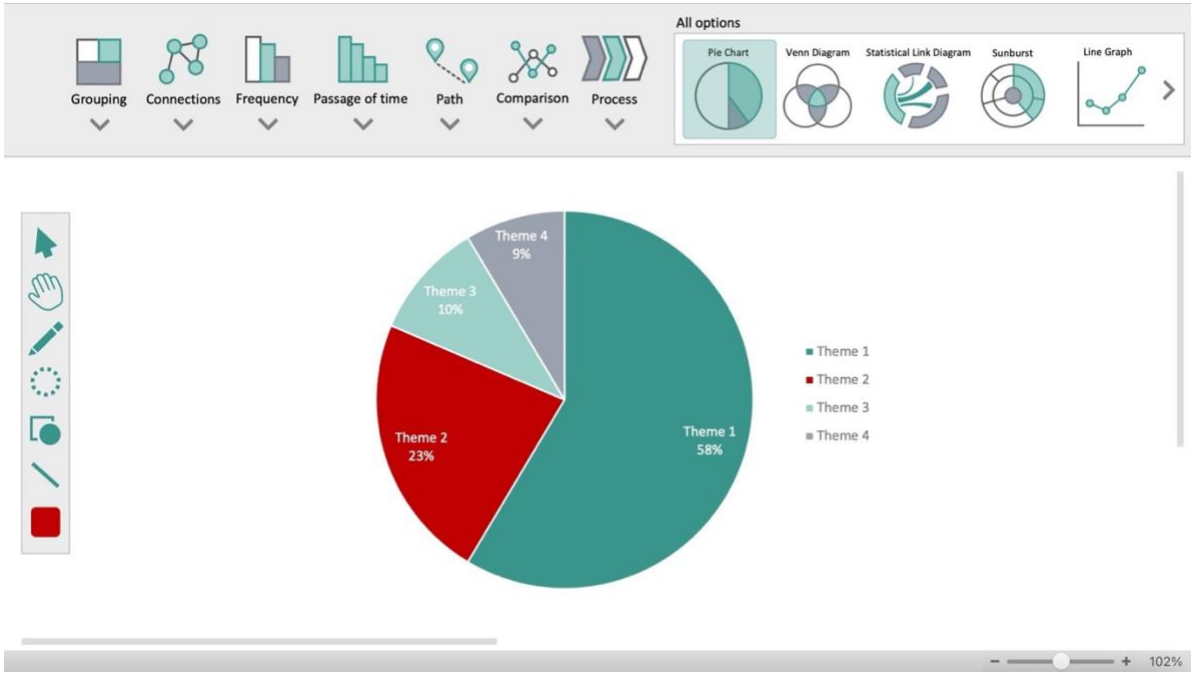
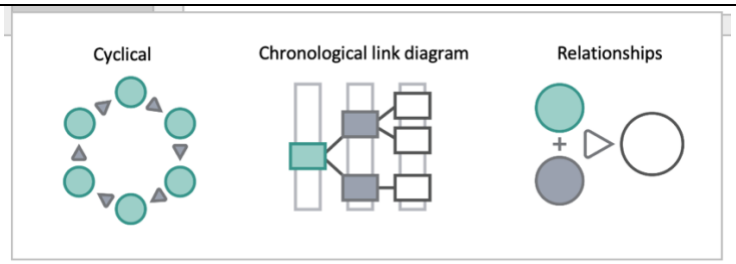
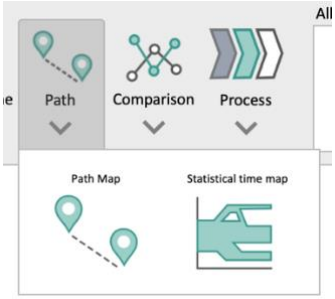


Figure 33: prototype 1, pie chart.

From top to bottom, the icons in the side bar allow the user to click on individual elements, move the entire screen, draw lines, select, and move areas, add shapes, add lines, and change colors. In addition to those, the user can also move through the workspace by placing two fingers on the trackpad, and zoom in/out via trackpad, or using the slider on the bottom bar.

Specific details about the design elements and how they satisfy the requirements gathered by the interview and analysis are presented in Table 5.

Guideline	Element(s) of prototype included to satisfy the requirement
Use proximity to establish connections.	The tool was separated into 3 areas, each responsible for a different function. The elements in these areas were kept close so the user understood that they belonged to the same family. For example, all visualizations that belonged to the category of ‘process’ were kept visually close.

	
Include data visualizations that focus on connections.	A category for connections was included with data visualization that focused on that specific need.
Allow for editing and personalization (e.g., editing colors, moving elements, etc.)	A side bar was included to allow for the users' different kinds of editing capabilities.
Explain the process behind the AI's results.	When the user hover over elements of the data visualization a pop-up appears with an explanation as to how that result was reached.
Include filtering tools.	Filtration was provided by the categories in the top menu. The types of data visualization were separated by category and the user could filter them by need.
Consistent color-coding.	The color pallet was kept to a minimal, using only one color and different variations. Options for editing color were available, but the user would have to edit those manually. 
Use of nested windows for easier navigation.	The options for different kinds of data visualization appear as pop-ups so the user never has to leave the main screen. 
Easy zoom in/out features.	Zoom in/out provided with a scale at the lower right corner and the two-finger motion on a trackpad. 
Easy manipulation of the data visualization to facilitate exploration.	The tool includes the ability to click and drag elements and move by placing two fingers on a trackpad.
Categorize data visualizations based on the needs the	Categories placed in the top menu bar.

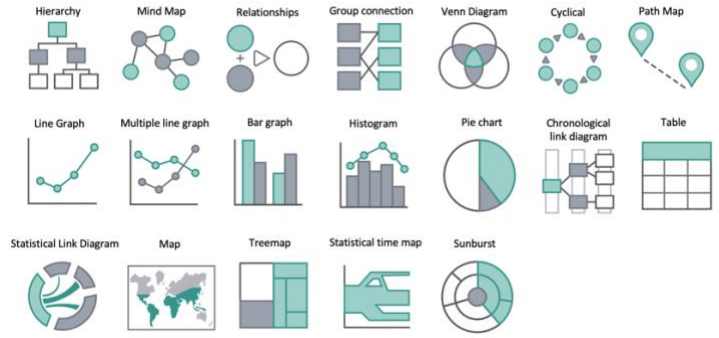
participants described (checking frequency, making connections, comparing, tracking changes over time, grouping, seeing processes, and mapping out paths).	
Provide multiple kinds of data visualization to cater to different preferences and needs.	In total 19 different types of data visualization were compiled and added to the prototype. 
Provide guidance on how different kinds of data visualization can be used.	The categories placed in the top bar separate data visualizations based on needs, guiding the user to choose one based on their research goals. 
Preserving autonomy.	Two things were done to preserve autonomy: explaining how the AI reached the results and providing the user with multiple choices. The choices were present in the form of the multiple data visualization option based on need, which the user can pick based on their goals.

Table 5: correlation between guidelines and design element(s) included in prototype 1 to satisfy requirements.

Chapter 7
Feedback on prototype

The testing of the prototype was done with the same 5 participants from the phase before to continue the user centered design process. They were shown images of the prototype and had the researcher walk them through how they could use it to conduct research. Following that, they were presented with all pages of the prototype on ZOOM Whiteboard and allowed to sketch over as they saw fit to better express their feelings about it or make any changes they desired.

Additionally, they were asked questions about ways they envisioned using the tool to conduct their own research, as well as asked to describe any additional wants, needs, or issues they might have had. This was done as an effort to include user in the process of the design of the data visualizations and ensure user centered design. The protocol for this interview can be found in appendix E.

7.1 Results from testing

The results from the testing were divided into what the participants expressed would facilitate or hinder their research and sense of agency, and their feedback in the form of suggestions for additions or changes to the prototype. The perceptions of agency were specific to what could be fostered through interacting with the interface for the data visualization tool, instead of agency as coming from controlling the inputs and outputs of the AI. First, the facilitators and barriers will be discussed.

7.1.1 Facilitators and barriers

All participants demonstrated interest in using the tool for their analysis. Overall participants displayed awareness of what data visualizations can be used for what, showing prowess for when to choose what kind of visualization. For example, all participants were able to look at the options and guess how they could apply to their research, abstracting the information from their data and anticipating whether the visualization would be useful for them rather quickly.

Additionally, all participants displayed knowledge in what kinds of data visualization should be used for qualitative versus quantitative data through looking at the icons in the toolkit. For example, participant 5 *quoted "and then I think in the frequency counts, they all look like they'd be very useful for quali data. Sorry quanti data,"* when explaining which categories she sees herself using the most.

Some visualizations scored lower in participant understanding, amongst them the statistical time map and the statistical linking diagram were the ones that confused the participants the most. Participant 5 and 2 dismissed statistical linking diagrams as solely quantitative initially, only processing the qualitative aspect of it after some explanation by the researcher. Participant 5 quoted that she *"never thought that that could be done like that. It's really nice,"* upon understanding the applications of a statistical linking diagram for qualitative analysis.

On the other hand, participant 4 had difficulties understanding the statistical time map. She explained she remembered what it was from the last interview, but that she would still consider it a complicated data visualization to be available with no explanation. She quoted *"I think you explained the statistical time map to me. But I feel like if I don't know what it is, and then I press it, and let's say, let's say… vomits out a statistical time map I might not know what it is."*

However, participant 1 expressed her beliefs that the addition of those two data visualizations made her feel more confident in the tool, as they are the types of datavis that are harder to make if you are inexperienced. She explained her need for a statistical linking diagram for her thesis, and both her and her committee's loss on how to make one easily and quickly. She quoted *"Because I've even had that need myself. And I'm not experienced. […] Like an easily accessible need, because even my committee members like 'yeah, when I was a master student it was really hard to even figure out how to make a statistical linking diagram'".*

She highlighted how AI could be very useful in these instances, especially to do frequency related visualizations, since a machine can keep track of number more efficiently. Participant 5 echoed the same sentiment, saying frequency checking is a task she is happy to delegate to AI, freeing more of her time for other parts of her research, such as writing.

Contrarily, when it came to delegating tasks, participant 3 brough up the issue of transparency. She explained that working with machines can sometimes be similar to a "black box" as it gives the user results without disclosing the process. According to her, if a person is not well versed in programming language, and the machine delivers a result they are unsatisfied with, there is little guidance on how to alter the process to arrive at a satisfactory result.

The explanation of how the AI has reached the results in the prototype was mentioned by her as an interesting feature that could combat that lack of transparency by allowing the user to understand how to make changes to reach the desired results. She quoted *"So, presenting them with some kind of clarity on like… why things were done might help to clear up some of that mental process and be able to like point to like one part of the process and be like, OK, like this is where we went wrong."*

Similarly, participant 5 talked about how she enjoyed the transparency elements of the prototype, saying she is interested in checking the results given by the AI and screening them. She compared it to *"checking back math,"* when talking about overseeing how the AI had reached its conclusions and whether she wanted to keep those results. Participant 2 echoed the same feeling, saying the transparency feature in the prototype would be very useful if it could explain to the user in a way that was detailed enough to allow for action. She quoted *"I think that would make a lot of sense as long as it's detailed enough."*

Regarding positive examples of delegation, participants 1, 3, 4, and 5 talked positively about the choices that the tool allowed them. Participant 1 expressed her appreciation of the categories presented by the tool, which provided guidance, without overstepping the boundaries of autonomy. She explained that having various options available, while still asking the researcher to point to their needs, was an excellent way to foster autonomy. This because is because, according to her, it pushes the researcher to understand their data and what they want to get out of an analysis based on what they are trying to accomplish. She expressed that through that sense of autonomy, the human aspect of qualitative research would not get lost.

Participant 4 echoed the same sentiment, sharing that she believes she would have difficulties in choosing an option because there are so many of them, but that she does not

believe that variety to be an issue, as it would push the user to think about their analysis purpose. She shared that *"I might have like a selection challenge. Like, there's so many options, which one should I pick? Which one is most relevant for me? Let me think of my purpose a little bit. But I don't think that's necessarily something bad on your end."*

Additionally, participants 1, 3, 4, and 5 talked about their appreciation for a tool that could help them organize their thoughts in image form easily. Participant 3 shared how sometimes she knows what kind of visualization she wants but has a limited vocabulary to express those needs and a lack of knowledge in how to get to the visual she has in her head. Therefore, having a tool that presents her with visual options facilitates the process of understanding how to bring to paper the image she has in her mind.

Similarly, participant 4 expressed how useful was to have a tool that could remove the burden of creating a visualization from scratch with no guidance. She explained that by having the guidance of the categories it makes it easier for her to make decisions. Participant 5 also shared how she appreciates having a tool that makes her data analysis more efficient and allows her to explore connections more easily. She quoted *"And I feel like a tool like this probably would have made the connections seem easier, like on my face. Which would have probably made my life easier."*

Participant 1 echoed the same sentiment, sharing how, for her quantitative research, she prefers to use a software that has pre-prepared data visualization options, instead of using one that requires her to know what kind of visualization she wants and how to do the coding to get it. She quoted that *"I know it's reliable and good… but it takes a long time and I have to learn how to do everything. […] But! I can't edit it! I can only export it! I can probably edit it in the code, but that's too hard. And then in contrast! I have this software, which… is so user friendly."*

She further explained that by having a tool that felt familiar using it would be less intimidating. To illustrate this point, she shared the experience of having to use Adobe Illustrator for her data visualizations at one point and being intimidated by the unfamiliar interface and abundance of choices. On the other hand, faced with the tool, she said that her limited experience with PowerPoint would make her comfortable enough to use it because of their similarities in interface design. Participants 2, 4, and 5 shared the same

feelings, highlighting how they found the design pleasing and appealing, and had no issues recognizing any of the icons or navigating.

When it came to navigation, all participants reacted positively to the navigation tools. Participant 1 and 2 especially compared the tool to NVIVO and said they preferred the tool. Participant 1 shared she found NVIVO quite complicated and needed help navigating it when she first started.

Participant 2 echoed the feeling and added to it by saying her experience using the CTA also left her frustrated by the lack of guidance and complication of navigation features. Despite not feeling the same frustration with the prototype, she still suggested adding a tutorial to the tool, saying that it might be useful for someone. She shared that *"[the tutorials] might not be something that I would find helpful, but I think someone would find them. And so, I think it's useful to consider."*

Lastly, two of the participants even used the low fidelity prototype of the tool during the interview to figure out a way to possibly represent their data. Participant 5, alongside the researcher, searched through the options and concluded that one of the options given by the tool was better at representing the results she had than the one she had.

What was previously a mind map, now became a group linking diagram, which she expressed fit her needs of showing the connections between several different elements, each belonging to one of two groups. A sketch of this data visualization that was brainstormed during the interview session can be found in Figure 31. The large circles for B and F represent, respectively, the barriers and the facilitators. The lines connecting each showcase which barriers are addressed by which facilitator.

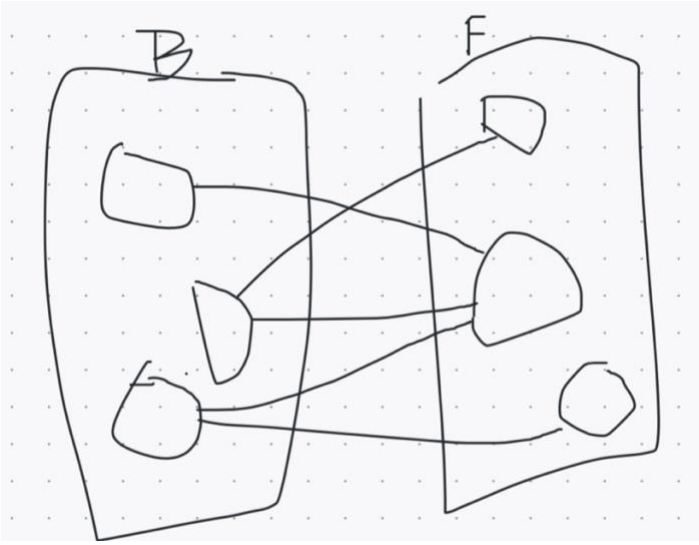


Figure 34: group linking diagram drawn with participant 5.

In the beginning of the process, the participant expressed her frustration in not being able to even sketch out what she wanted to represent, quoting *"honestly, I don't know how to draw that out".* However, as the researcher and participant worked together to brainstorm possible forms of visualizing her data, she began to grow more confident and finally arrived at a representation she was pleased with, quoting *"yeah, it makes more sense to me. Yeah, I like It. I feel like these are the things that makes you, I don't know, gives you ideas because you can actually build connections between two variables in your research more effectively."*

Similarly, participant 2 shared how she usually conducts her process of data analysis, complementing the information she shared during the first interview with another sketch of how she chooses to organize information. Echoing what was said in the first interview, she discussed using what she called a "coding map". This coding map (Figure 32) is a web-like mind map that connects themes to its sub-themes, and then those sub-themes and themes to one another.

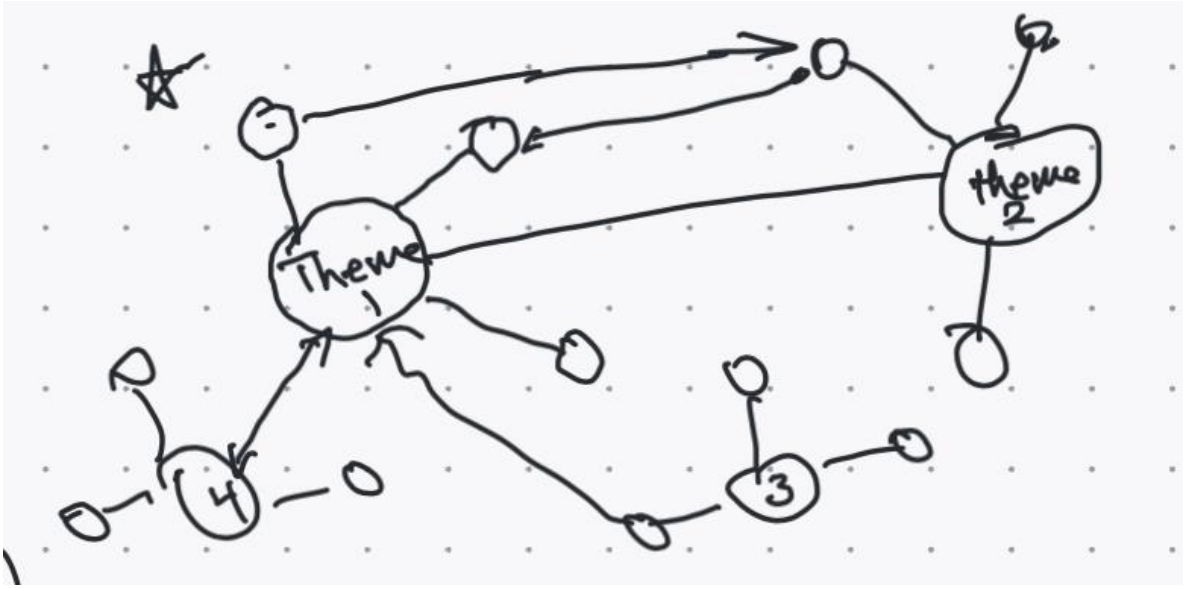


Figure 35: participant 5's coding map.

However, through the interaction with the prototype and the researcher, a new idea was sketched for the same kind of representation but using linking grouping diagrams instead (Figure 33). This idea would make the themes large groups and insert the sub-themes inside of it, still maintaining the connections between the individual elements. The participant voiced how much she preferred the new version of the data visualization to the old one, quoting *"I think yours [link grouping diagram] almost makes more sense because codes like sub codes are within the theme."* She explained that the only reason she opts for using coding maps is because of available space, but that her coding maps always end up disorganized and that she believes the grouping diagram would be a better choice for keeping information organized.

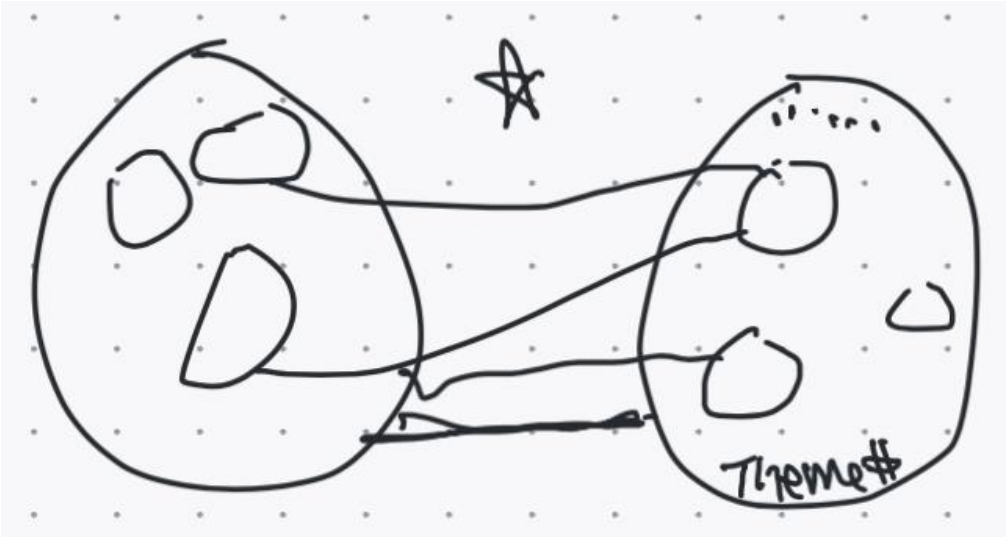


Figure 36: link grouping diagram for participant 5.

7.1.2 Participant feedback

Participant feedback was divided into 3 categories: elements that were asked to be added to the interface, data visualizations the participants felt were missing, and interactions the participants would like the tool to have. Some of those elements were already stated above, so this section will focus on the novel items.

When it came to the elements that were asked to be added to the interface, participant 2 highlighted two different uses she anticipates having for the toolkit: making graphs for presentation and for data analysis. For presentations she explained preferring the ones that related to connections, such as grouping diagrams and mind-maps, while for presentations she anticipates using hierarchies and tables.

Regarding using datavis for presentations, participant 3 expressed a need for what she called “granularity”, which she described as the ability to add or remove details from a data visualization depending on their use context. For example, if she were doing a data visualization for industry partners, she would prefer it to be simpler and only give the overall story behind the data. On the other hand, if she was making a datavis for publication in a scientific paper, she would want to add more details to guide the people reading it, *“focusing more on accuracy as opposed to visual cleanliness”.*

However, changing this granularity was described as quite difficult, since it has to be done manually and is subject to human error. To solve this, a smart scale was suggested to be able to go from very simplified data visualizations to more complex ones without the risk of losing data by doing it manually. She *quoted "because otherwise that aggregation would need to kind of be done manually, and there's a lot more room for error there so… Having a smart slider of some kind would be very helpful."*

Overall, the editing abilities of the tool were overall considered beneficial. Participant 4 expressed her appreciation for a tool that creates the visualization for her, but still allows her the freedom to edit if desired. Likewise, participant 1 explained how she appreciates not having to create a visualization, but instead only edit what the software has given her. She shared *"that's why I appreciate the options, and then the ability to edit if you want. But it's not like you start with the editing and the creation, and then apply it later."*

Additionally, two participants shared their need for not only editing the data visualization, but also altering the data behind it. Participant 1 talked about how frustrating it can be when she had to edit her original data in order to get a software to generate the visualization she wants and suggested a way for the software to be flexible when reading the input. She gave the example of generating data visualizations using Microsoft Excel and having to change the data from vertical to horizontal, quoting that *"it can be time consuming and annoying to transfer data, you know, the way data is stored".*

Participant 4 shared her insights on data alteration as well, saying that she would appreciate being given access to the data behind a visualization if she ever needs to alter or add to it. She gave the example of wanting to add demographic data that was not in the original dataset she input and suggested a pop-up that could open and allow for the user to make quick edits and see their results efficiently. She and participant 1 also emphasized how important it is that the changes made by the user be applied to the visualization relatively fast.

However, two participants brought up possible issues with editing. Participant 3 talked about the interconnectedness of the data visualizations and how moving one element would change the entire visualization. She exemplified this by questioning how a user would be able to click and drag an element of a line graph without it moving the entire visualization or

breaking the graph. As a solution, participant 3 suggested the use of smart visualizations that would adjust as the user moves objects.

Similarly, participant 4 talked about her apprehension towards the editing features, citing her fear of moving an element and accidently changing the meaning of the graph, or misrepresenting the data. She explained that she does not see herself using a lot of the editing features because of this apprehension, and that if she were to add elements to a visualization, she would like a way to change them or undo. She quoted *"I feel like if I'm adding stuff to it, it would be helpful to move things. If not, I worry that I'll be silly and move things by accident."* For that, she suggested the addition of an undo and re-do button, as well as a feature that warns user if they are making changes to the data visualization that might risk misinterpretation.

Additional feedback that was suggested by the participants included: the addition of organization tabs, a text tool, a ruler on top of the drawing space, a grid, a color wheel, HEX code, eye-drop, heading for the categories of data visualization, and categories separating data visualizations into quantitative, qualitative, and mixed methods. All of those fell in the category of interface additions that would either allow for further editing capabilities, or different forms of organization.

Lastly, participant 4 brought up the issue of color-blindness, and suggested the addition of a tool for checking if the colors chosen by the user are accessible. In addition to that, a tool for changing the thickness of lines so that the users have another option for highlighting information other than color.

Six suggestions were made in terms of the data visualizations present in the prototype. Participant 2 asked for a hierarchy to be included in the category of processes, while participant 3 asked for a Gantt chart and participant 5 asked for a word cloud and an exact word count visualization.

Participant 1 asked for a tabular grouping process and a nesting group diagram to be added and sketched out her idea for what the former could look like (Figure 34). The table would be separated into sections, the smaller sections would be inside of the large sections on top, but the arrows connecting them would imply a sequentiality, as well as relationship between the elements.

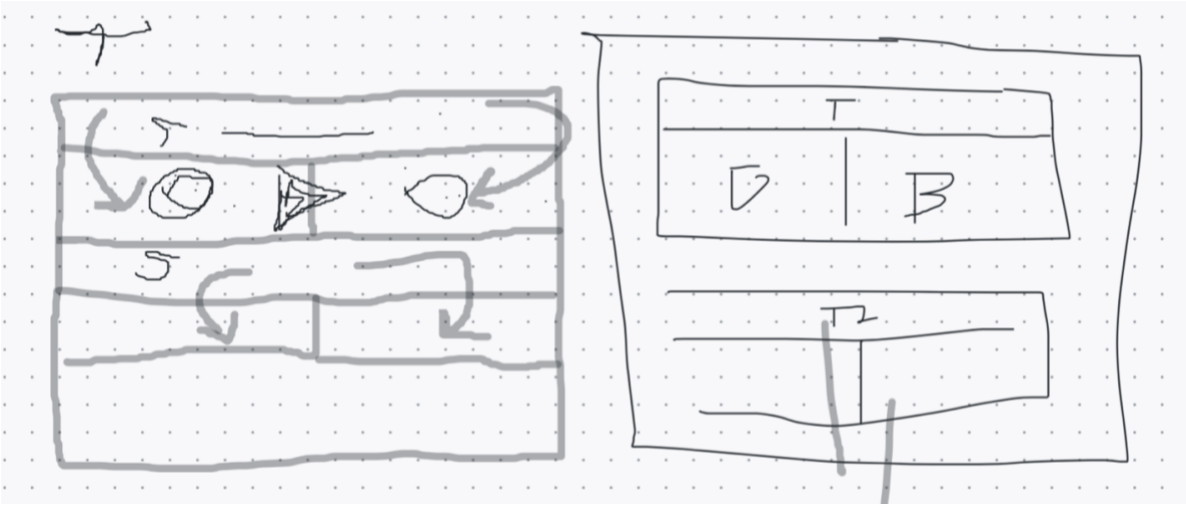


Figure 37: tabular grouping process, by participant 1.

The last category of feedback shared by the participant were the types of interaction they would like to have with the tool. Firstly, relating to saving the visualizations, three participants mentioned wanting the ability to save all visualizations together as a project file which would be compatible with the software of the tool – retaining editing capabilities –, and one that would save only the images – with no editing capabilities –, which would be used for presentations and papers.

Additionally, participant 3 asked for a tool that could alter the opacity of the objects, and for the 'all options' window to open in a separated pop-up, as well as adjust to user behavior, keeping record of the latest and most used kinds of data visualization. Lastly, participant 5 shared her desires for the 'all options' window to have a horizontal scroll enabled by scrolling with two fingers on the trackpad.

7.2 Final observations from testing

Overall, the participants demonstrated interest in the tool and pointed out many elements that were proposed in the guidelines as useful for their research. Feedback from the participants was organized into a list of alterations to be applied to the second iteration of the prototype. Those were organized into: interface, data visualizations, and interaction and presented below:

- Add to interface:

    - Tabs.
    - Text tool for taking notes and adding text.
    - Group for datavis good for qualitative, quantitative, and mixed methods.
    - Heading for the side of the top menu that is separated into categories.
    - Ruler.
    - Grid.
    - Undo arrows.
    - Eye-drop for colors.
    - HEX code for colors.
    - Color wheel.
    - Tool that allows for checking the design for color-blindness.
- Addition of data visualizations:
    - Tabular grouping process.
    - Nesting group diagram.
    - Adding hierarchy into the possible datavis for processes.
    - Gantt chart.
    - Word cloud.
    - Precise word count datavis.
- Interactions with the tool:
    - Save multiple data visualizations as a single project file.
    - Save data visualizations as images.
    - Add a tutorial.
    - Warn the user for any changes that might alter the visualization to the point of misinterpreting the results.

- Employ smart editing that will change the visualization when the user tries to move elements that are connected (e.g., change the entire graph if one of the data points is moved).
- Add a smart slider to control the density of information presented.
- Ability to change the thickness of lines.
- Ability to change the opacity of elements.
- Responsive window for ‘all options’ that shows the newest and most used options by the user.
- Make the window for ‘all options’ a pop-up.
- Double finger trackpad scroll functionality in the window for ‘all options.’
- Ability to edit the data that was input into the visualization (e.g., excluding specific data points, adding data that was not originally there, etc.).
- Fast changes to the visualization when something is altered.

Chapter 8
Adjustments to prototype

The prototype was updated based on the results and suggestions from the second interview. Twenty-two pages were designed following the recommendations provided by the participants, which can be found in appendix F.

An example of the main screen can be found in Figure 39. Like the first version, the prototype is divided into sections, each with its own functionality. Many of the elements of prototype 1 were preserved, with only some changes to editing capabilities, types of interaction, and available data visualization types.

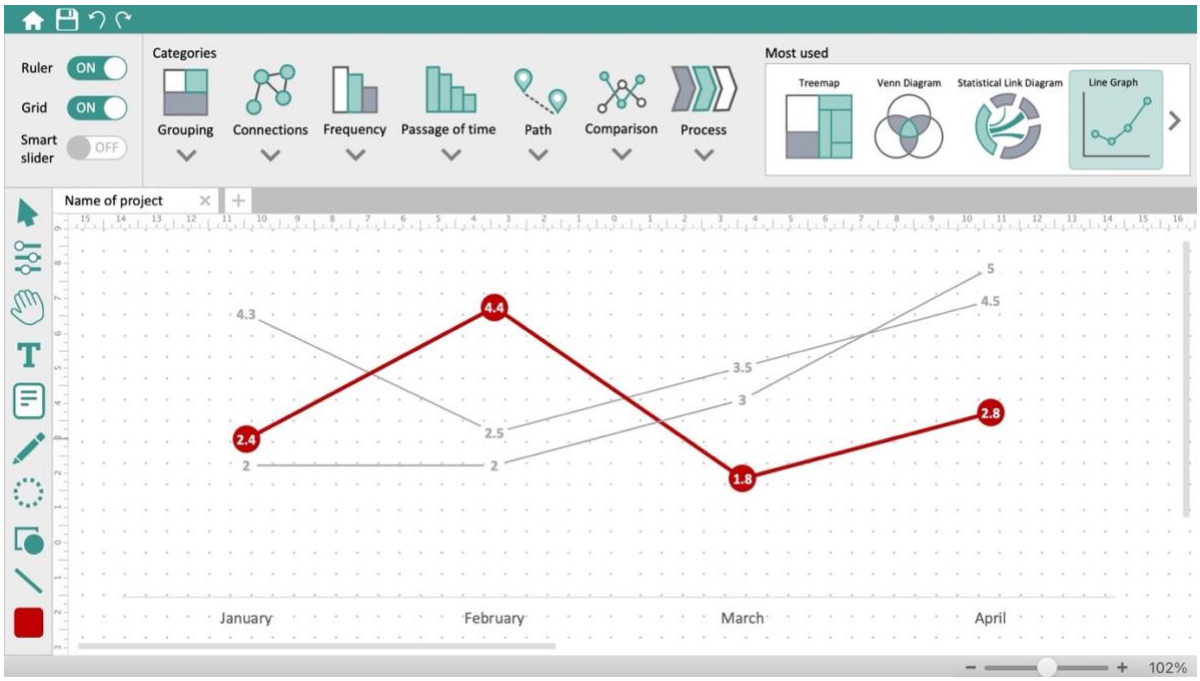


Figure 38: updated prototype, main screen.

Above the top bar, a smaller bar was added for functions such returning to a main menu, saving the project, or saving the datavis as images, as well as re-do and un-do buttons. The top bar has mostly remained the same, with the addition of a section to the right that enables some of the suggestions of the participants (i.e., ruler, grid, and smart slider). The smart slider was the suggestion of participant 3 and would enable the data visualization to change the amount of detail presented.

Other additional changes were tabs added so the user can make more data visualizations and go back and forth between them, and 3 more options in the side bar (i.e., an option for editing the data input in the datavis, adding text, and making notes) (Figure 40).

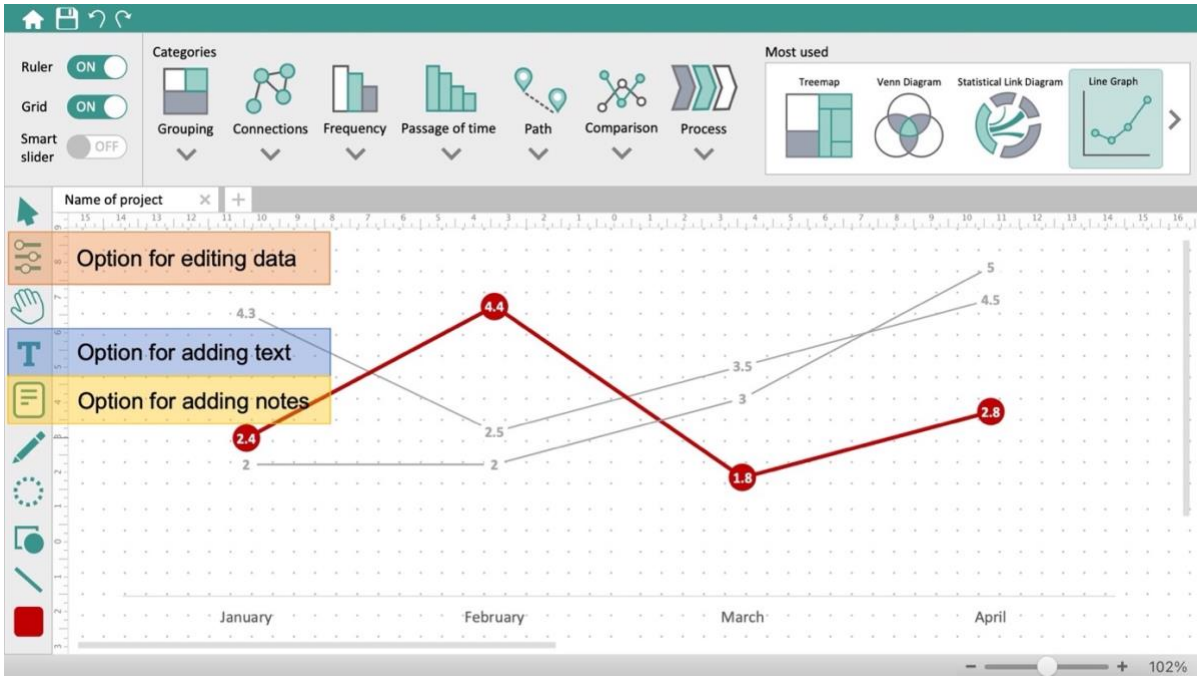


Figure 39: updated prototype, editing options.

Another element that was added to the updated prototype was the ability to edit the data that gets input into the datavis and safeguard it against accidental mistakes. For that purpose, an AI assistant was added (Figure 41), both to communicate how the result was reached by the AI, as well as to allow the user to edit the data to get different results. To keep the user from accidentally altering the data visualization in a way that misinterprets data, a warning was added to the prototype (Figure 42).

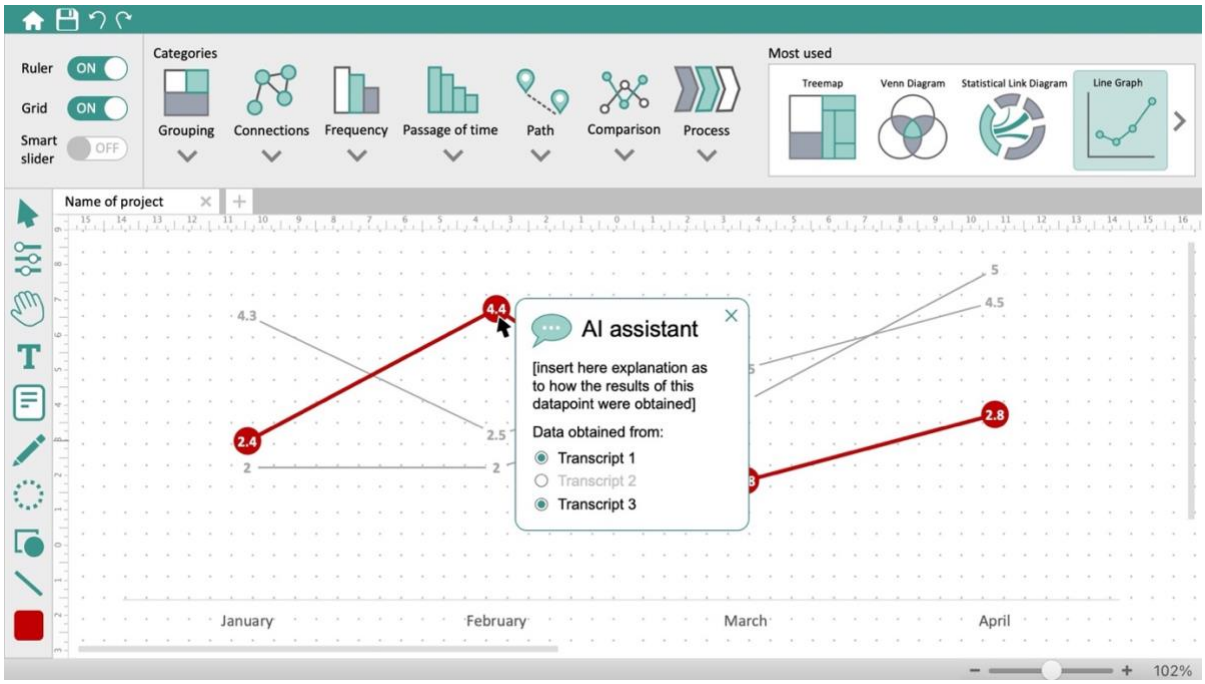


Figure 40: updated prototype, AI assistant.

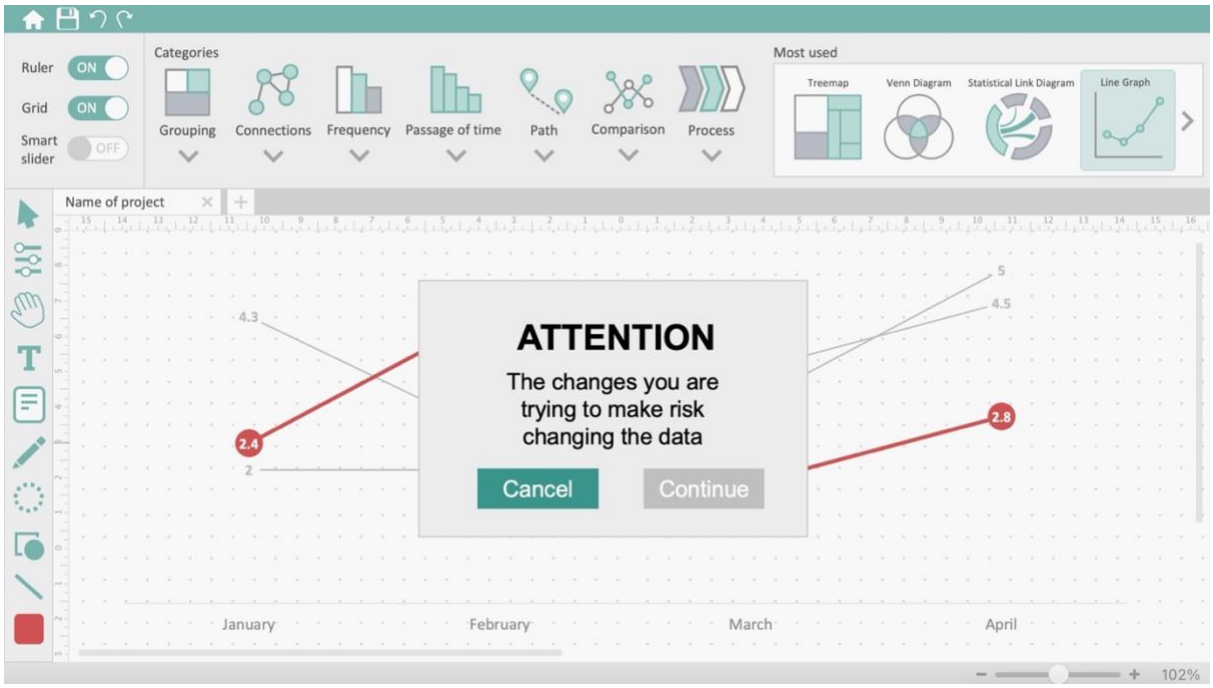


Figure 41: updated prototype, mistake reducing features.

Lastly, five additional datavis were added, all based on the needs the participants expressed. Those were: gantt chart, nesting group diagram, tabular group process, word count, and word cloud. The window that displayed all the datavis options in the first prototype was also altered following participant 2's suggestion and separated into what would be best suited for qualitative, quantitative and mixed methods (Figure 43).

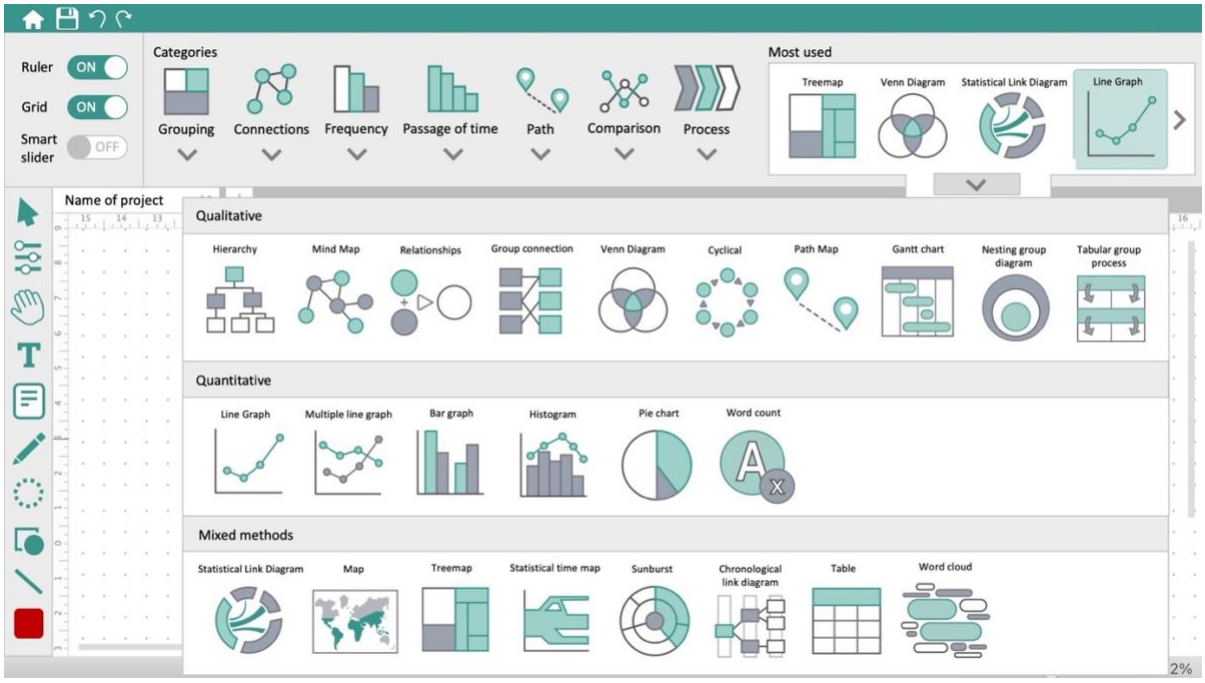


Figure 42: updated prototype, all data visualization options.

There were some interactions suggested by the participants that could not be applied to the prototype since it is not functional. However, all requests were noted for the future. They were the accessibility needs for color-blind people, the addition of a tutorial, smart editing of the visualization when one element is moved, double finger side scroll for the 'most used' window, and fast changes to the datavis when input data is modified.

A correlation between the changes made to the prototype and the requirements derived from the second interview are in appendix L. This phase satisfies objective five of developing a template for using data visualizations as knowledge translation tools for qualitative researchers performing interview data analysis with the assistance of automated tools.

Chapter 9
Discussion

The discussion was framed by the four ways media can foster agency for the user: choice, action, narrative, and space and how the participants perceived them during the research process. Fostered agency is discussed here through the lenses of how people interact with the data visualization tool, as opposed to the relationship between participant and AI.

9.1 Action

Agency can be fostered through the mastery of action when people perform actions with their physical bodies and see them mirrored in a digital space [24]. The biggest barrier when it came to fostering agency through action was delegation. Overall, the participants were open to some delegation, but were firm in maintaining their sense of autonomy and control. The tasks they were willing to delegate were the ones considered to be time consuming and robotic, especially counting, or checking for simple themes. Automation can be beneficial for users when it lessens the mental workload and leads to healthier workers [34]. However, participants were adamant on maintaining their autonomy on two fronts: checking the work of the AI and making choices for their own research.

According to the model for types and levels of human interaction with automation [34], these two characteristics are important to be kept for AI delegation, since it keeps people from falling into complacency and neglecting checking the results AI provides, as well as keeping the person's skills from degrading over time.

Participants 4 and 5 voiced their belief that qualitative research requires creativity and the humanity of a real person behind the results, which is in line with the opinions found in previous research [4, 7, 8]. However, they were receptive to delegation if that humanity did not disappear.

It is important that people not stop being at the front of the work, instead opting for using AI as an assistant [29, 34]. This sentiment was echoed by the participants during the interviews, which agreed that AI should be used as a tool for assistance and not replacement. Delegation heavily related to action, since the more tasks are delegated to

machines, the less input comes from the participants and the less opportunities for fostering agency happen.

However, an issue raised by the participants when it came to assistance was the lack of trust, they felt towards working with machines. There were two sides to this mistrust, either participants did not believe AI would be competent enough to get useful results, or they feared AI would be too good and replace them.

As a response to this lack of trust, and element that was appreciated in the prototype were the explanations. Participants reacted positively to having a tool that would allow them to work alongside AI by providing explanations about how the results were reached, commenting on how they felt the transparency made them feel more in charge.

The explanations allowed them to either check back on what the AI had done to ensure the results were up to par with the needs of the researcher (e.g., participant 5 comparing it to checking back math), or gave them confidence to believe they were still in control of the research (e.g., participant 4 expressing her desires to remain the one to make choices and not have the AI take that autonomy from her).

Explaining results is important for establishing trust between people and AI [29]. One of the four aspects of trust described by Lubars and Tan (2022) was that, for people to trust AI, they need to know the intent behind the automation and how it aligns with their own personal goals. For that to work, the machine needs to provide enough explanation of its internal processes for the humans to understand it [29]. Participant 2 talked about that same feeling of wanting transparency, explaining that it would be ideal for her if the AI could explain what it had done with enough detail to allow her to make any changes if she wanted. Therefore, if the explanation is presented in a way that can allow for the user to understand the AI, then they can feel confident in acting upon those results, fostering agency.

During the sample analysis, issues with fostering agency through action were found for the software that did not provide transparency, which meant the users had to spend significant time making sense of the results given to them by the AI. Participant 3 echoed this feeling when she shared how frustrating it could get to work with AI as a “black box” and have no idea where the results came from or how to change the query to get different ones. Because of this uncertainty, acting can become difficult and hinder agency.

Scholastic was the outlier when it came to transparency, employing some features that allowed the user to understand the machine. The most important of them was showing the multiple different rounds of AI work and letting the user go back and forth between them and see what the difference was.

Another element that assisted in transparency were the categories presented to the researchers. The options gave the participants the choice to act if desired, and the possibility of acting is just as powerful for fostering agency as the action itself [24]. Participants shared how having multiple options for different visualizations available improved their sense of autonomy.

The use of datavis for research was similarly found to be positive in terms of fostering agency through action. Four out of five participants described using data visualizations as cognitive assistants when conducting research, which helped them organize, explore, and ultimately take action of their own research [11, 16, 24, 45].

The participants described using datavis to conduct analysis as tools for exploration, organization, and for searching to patterns or stories in data, which is similar to what previous research has showed about the benefits and uses of datavis for qualitative research [3, 4, 6].

Lastly, through the editing tools provided in the till, agency was fostered thorough the researchers' interactions with the datavis [43, 55]. This feeling was being fostered through allowing the for the researchers' opportunities to act by changing the datavis and seeing the consequences of their actions as the datavis changes, and through that feedback loop perceiving agency [24, 43, 54]. Participants reported noticing this by showing appreciating for the editing tools available.

9.2 Choice

Agency can be fostered through the mastery of choice when the user recognizes they have a possibility to make a choice. When the choice is made, then agency is perceived regardless of the consequences [24].

One of the biggest aspects of fostering agency through choice was trust. Like the research done before [2, 7, 29] our participants expressed being open to working with

machines as an assistant, but against full automation, citing they did not want their choices taken from them. The fact that the tool suggested possible results but ultimately left the choice up to the researchers was an element that participants commented improved their sense of autonomy.

Both the results from the interview and the analysis of the sample showed that another element that influences trust is receiving explanations for the actions of the AI. During the sample analysis, not providing transparency was considered to be an element that hindered the perception of agency. Similar to action, when users do not understand how the AI reached the results, they do not feel confident in making choices, hindering their perception of agency. The CTA and ATLAS.ti had issues with this by only showing the participant the result with no context behind it, which ultimately can hinder the ability of the user to feel agency since all their choices are being removed [24].

Elements that were considered to the positive were the ones that allowed participants to choose freely (i.e., editing tools and the categories of datavis). Chiefly amongst the elements that participants mentioned fostering agency were the categories and the ability to choose which datavis best suit their goals. Additionally, there was an emphasis on wanting the effects of their choices to show up quickly. Participant 1 and 4 voiced that they would like the tool to adapt quakily when they made changes to the data. All these elements contribute to the idea that participants perceive agency when they make choices and see their effect on the world around them [54]. Because of the multiple categories participants felt comfortable choosing the one that best suit them and quoted that possibility of choosing as a positive.

During the sample analysis, an element that was considered a barrier for choice was not allowing a user to make changes or input new elements. CTA and ATLAS.ti both failed in that aspect, restricting the users from creating new codes or editing many of the data visualizations. These features are the opposite of what participant 1 and 2 talked about wanting in a tool for assisting with research. Both shared enjoying working with AI the most when it provided suggestions and allowed them the option to accept or reject, providing opportunities for fostering agency through choice [24]. Furthermore, participants showed different preferences in the ways they used visualizations to process information. Because

of these differences, they were receptive to a tool that catered to their many different needs during the second interview.

However, despite participants enjoying the freedom of choices, they also expressed their appreciation for a tool that provides guidance. Participants quoted liking not having to come up with the visualization, instead delegating that task to the machine and only editing the results given. Additionally, the categories were also appreciated as they provided them enough guidance to make a choice without completely removing that autonomy. This guidance and restrictions contributed to fostering agency because in order to perceive agency through making choices restrictions are necessary [24, 55].

Participant 1 illustrated this point when she mentioned preferring a software that gives her guidance, instead of the one with complete freedom, but no assistance. She mentioned how she would rather use the one that was more restricting, than struggling through figuring out what she wants from scratch with no help from the system.

9.3 Narrative

Mastery of narrative happens when people detect the genre of the media they are interacting with and can accurately predict what will happen based on previous experiences [24]. During the interviews, participants mentioned using data visualization for narrative purposes, both for finding the story in their own data, or using the datavis to weave a story as an explanation as to why their project was important.

Participant 3 highlighted the importance the datavis when communicating information and the need to focus on narrative to garner tangible change. Using datavis as a way to engage people and promote change has been discussed before, as it is a type of media that can increase intrinsic motivation and empathy [11, 14, 44].

However, additional to the type of mastery of narrative described by Eichner (2014), the data visualizations used by the participants fostered agency through narrative by mutual creation [43]. Participants expressed liking the editing tools and flexibility of a tool that allows them to craft the datavis to best fit their desired narrative [43, 52].

Another way to look at the positive reaction editing tools got from the participants is through a commitment to meaning [55] and the crafting of said meaning [43]. Participant 3

mentioned this when she talked about using datavis to tell a story, crafting the narrative alongside the system. This means that through interacting with the tool, the researcher is creating meaning, and committing to that meaning, by making alterations [43, 55].

Especially when it comes to crafting a narrative, participants talked about liking the use of data visualizations, sharing how they prefer to let the data lead the narrative and create a story that makes sense using visual elements. Participant 5 shared how she prefers to visualize data before crafting a story, and that is the way she found makes the most sense for her way of thinking.

The types of data visualization most appreciated for narrative purposes were the ones that showed relationships and connections. This was seen both in the analysis of the sample, as well as the interviews with the participants. In the sample analysis, CTA and NVIVO were the two software that focused the most on connections, offering data visualizations for that purpose. This was considered positive during the analysis since it was a type of data visualization that allows for elements of narrative [13, 52]. Similarly, the participants commented that through mapping out connections they had an easier time following the thread of a story, which is in line with Eichner (2014) and the idea that by recognizing similarities in media, a user can follow a narrative and experience agency [24].

Additionally, by providing multiple different types of data visualization, the tool allowed participants to choose the one that best suits their narrative needs and thus better supported their way of visualizing stories. For example, participant 5 talked about how she understood stories best when they were being presented in mind maps, because that was the kind of media that she recognized and could understand, which supports the idea that through recognition of patterns in media, agency can be fostered through a narrative [24].

9.4 Space

Agency can be fostered by the mastery of space every time a person moves through a space and feels in control of the movement [24, 54]. Overall, participants enjoyed using data visualizations to conduct data analysis, quoting that the act of visualizing things made it easier for cognitive processes. However, a caveat of those visualizations was the need for freedom of movement within that visualization.

Participants reported feeling that expressing their frustration with tools that had complicated navigations and did not allow them to move freely. The less intuitive it is to move through a digital space, the less agency through space is fostered [28, 54, 57]. Additionally, participant 4 also commented preferring to use whiteboards, paper, or her iPad when making data visualizations for organizing her thoughts, as they gave her the movement freedom for being messy.

Similarly, participant 2 and 3 also described the act of physically moving elements around as a positive and appreciated it during research. All these actions of moving through digital space work for fostering agency through the mastery of space [54].

During the sample analysis, similar results were found, in which the software that allowed for movement were considered better fit for fostering agency through movement. For example, NVIVO's ability to click on elements and move them around freely mimics the experience participant 4 described when she talked about being able to pick elements up and move them around during data exploration. Another example of mastery of space interfering with perceptions of agency comes from participant 3, who discusses her frustration in trying to navigate word files as they had no features to facilitate data analysis.

Specific elements that were mentioned to be helpful for navigating were the use of color, clustering, and guidance. Participants expressed preferences for color-coding, in particular participant 4 who explained that it helped her organize and make sense of her data. She described it as making things messy on paper so they could be organized in her brain.

Clustering of similar elements was talked about when participants mentioned that the sections of the tool seemed intuitive and easy to follow, reporting no issues with getting their desired datavis based on the categories proposed. NVIVO and CTA also employed clustering to organize their interface, which was found during the sample analysis to be a positive trait.

Another type of clustering that was found to be good was nesting windows. Participant 5 talked about being appreciating seeing options without them crowding the visualization space. Similarly, the sample analysis showed that having nested windows would help with navigation thus making moving through space more seamless. These details assist in

navigation are important for fostering agency, since allowing the user to get lost in the software causes frustration instead of agency [24, 54].

Lastly, guidance was regarded as an important part of navigation during the sample analysis. For example, NVIVO offers options for further analysis, but they are hidden under right-clicks that the user either has to discover or already know. This lack of guidance can be considered a barrier since it makes navigation confusing and frustrating [24, 28, 54].

Participant 3 reported a similar frustration with her use of the CTA, quoting that the lack of guidance made navigating the software difficult. Both the participants and the sample analysis showed that it is important to have proper guidance when trying to foster agency through the mastery of space, since the more frustration navigation causes the less agency is perceived [24, 57].

Chapter 10
Final Considerations

This chapter discusses the main considerations of this thesis. Firstly, the main conclusions will be presented, following by the objectives and how they were reached, and lastly a discussion of the limitations and considerations for future work.

10.1 Main conclusions

The use of AI as a research assistant was considered to be of interest by the qualitative researchers interviews for this thesis. However, they highlighted the importance of delegating tasks to the machine in a way that would not overstep their own sense of agency.

For a tool to foster agency for researchers it needs to follow their specific wants and needs. Through the use of Design Science Research and user centered design, a prototype was developed to map out those needs and extract reasonable facilitators for the barriers pointed out by the participants.

Following the concept of fostering agency through media proposed by Eichner (2004) we have divided the barriers discussed during the interviews into four categories, those that impede actions, choice, narrative, and navigation through space. These barriers relate to the agency that could be fostered through interactions with the data visualization tool.

The main barrier in action was through delegating work that researchers did not want to delegate to the machine and thus felt loss of agency. Participants were open to entrusting AI to do simple coding, frequency checks, counting, and providing suggestions for possible results they might have missed without imposing them. However, they expressed wanting to be in control of their own research process and ultimately deciding whether or not to use what the AI had provided.

For this, the facilitator designed with the help of the participants was providing them with choices before they used the AI, so they could understand what they were asking the machine to do, as well as giving researcher tools for co-creating and altering the results delivered.

A barrier that was partially attributed to action and partially to choice, was the lack of transparency with working with AI. Participants expressed feeling uncertain when AI gives results with no clarity as to how they were reached. To combat that, a feature of the prototype included was an interface to facilitate communication between the AI and the researcher.

Regarding choice, two opposing issues were found, the lack of choice in how the results will be achieved, and excessive freedom with no guidance. The lack of choice was perceived by the participants when they felt their actions were meaningless when interacting with AI and they had no freedom to impact the research. On the other hand, if a tool provided too much freedom, they became uncertain on how to begin and faltered in their decisions.

To counter-act those barriers, it was suggested that the prototype include explanations to the AI process so the researcher could alter them if needed (increasing freedom), as well as guided them with categories to the data visualizations so they did not have to design from zero (providing guidance).

In terms of narrative, the barrier observed was the lack of ability to alter the results given to the researcher and this the inability to cater the data to a narrative that they had found during they data analysis. As a solution, editing tools were included in the prototype to ensure customization for the researcher to explore their narratives as they saw fit.

Lastly, for space, two barriers were identified. First was the lack of guidance as to how a tool works. Many of the software in the market have complicated interfaces that lack explanations as to how to navigate and operate them. This lack of explanation was reported to cause frustration by the participants. As a possible solution, it was suggested the prototype include a familiar design, with tutorials and organization based on color-coding, nesting windows, and grouping of similar elements.

The second barrier to fostering agency through mastery of space was found to be the limited movement that the software currently in the market provide. Participants expressed wanting freedom to move elements and the screen when they are manipulating datavis for research, and that doing so assisted with their thought organization and explorations. However, a caveat with this freedom was the fear of manipulating data wrongly, which was

solved using either a smart tool that adjusts the visualization as elements are moved, and a warning for user so they do not misinterpret data. A summary of the barriers and facilitators can be found in Figure 44.

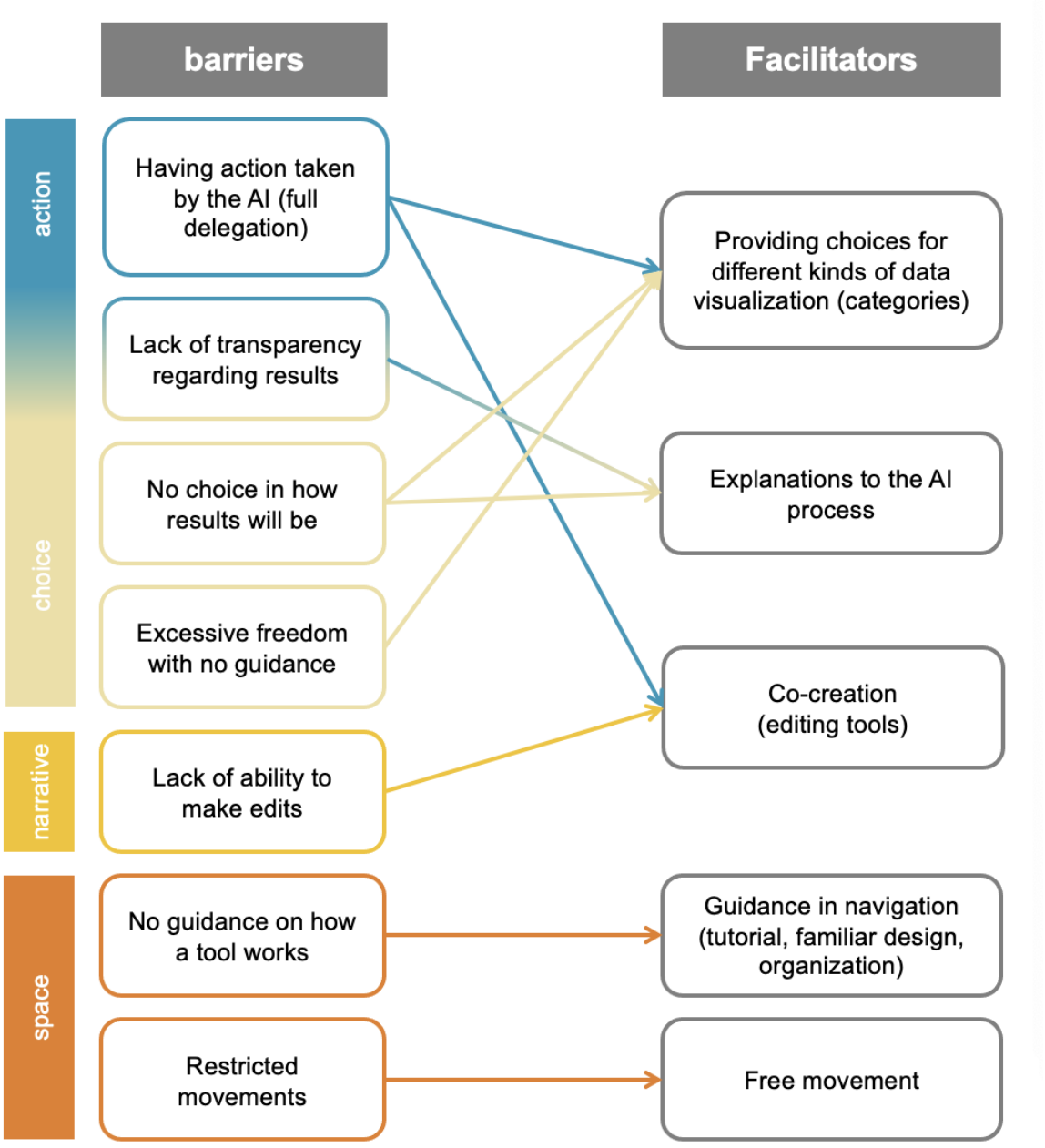


Figure 43: barriers and facilitators.

10.2 Study limitations and future developments

For future research, further improvements to the design of the interface should be applied to the prototype, following the feedback from the second interview. After these changes have been implemented, future work should seek to test the prototype with a new set of participants to assess its usability and efficacy. Even with a small sample we were able to map out many differing specific needs for qualitative health researchers, therefore, we recommend testing the tool with a different sample to ensure all the needs of the community are being met.

Additionally, the largest unexplored potential of this thesis was the ways for researchers to interact with AI and how to improve that communication. The participants voiced their need for transparency; however, this thesis has only managed to explore this briefly. Because of the interactive aspect of transparency, the prototype could not fully dive into how to improve communication between people and AI, since it was non-functional. We recommend that future studies seek to make functional prototypes to understand how researchers would prefer to have that transparency communicated to them.

Lastly, aspects of accessibility need to be included for future prototypes. A feature for designing for color-blindness was brought up during the interviews, but future work should aspire to include accessibility planning and consultation with people with disabilities to make the tool more accessible for different kinds of people with disabilities.

References

[1] A. L. Chapman, M. Hadfield, and C. J. Chapman, “Qualitative research in healthcare: An introduction to grounded theory using thematic analysis,” Journal of the Royal College of Physicians of Edinburgh, vol. 45, no. 3, pp. 201–205, 2015, doi: 10.4997/JRCPE.2015.305.

[2] J. L. Feuston and J. R. Brubaker, “Putting Tools in Their Place: The Role of Time and Perspective in Human-AI Collaboration for Qualitative Analysis,” Proc ACM Hum Comput Interact, vol. 5, no. CSCW2, Oct. 2021, doi: 10.1145/3479856.

[3] P. L. Rice, “In Quality We Trust! The Role of Qualitative Data in Health Care,” 1996.

[4] Z. Q. Al-Busaidi, “Qualitative Research and its Uses in Health Care,” Sultan Qaboos Univ Med J, vol. 8, no. 1, pp. 11–19, 2008.

[5] R. Fitzpatrick and M. Boulton, “Qualitative methods for assessing health care The evaluation of health care and efforts to maintain and improve quality in health care have very largely drawn,” 1994.

[6] C. Pope, P. Van Royen, and R. Baker, “Qualitative methods in research on healthcare quality,” Qual Saf Health Care, vol. 11, no. 2, pp. 148–152, 2002, doi: 10.1136/qhc.11.2.148.

[7] J. A. Jiang, K. Wade, C. Fiesler, and J. R. Brubaker, “Supporting Serendipity: Opportunities and Challenges for Human-AI Collaboration in Qualitative Analysis,” Proc ACM Hum Comput Interact, vol. 5, no. CSCW1, Apr. 2021, doi: 10.1145/3449168.

[8] C. Pope, S. Ziebland, and N. Mays, “Qualitative research in health care Analysing qualitative data,” BMJ, vol. 320, pp. 114–116, 2000, [Online]. Available: www.bmj.com

[9] M. H. Hong, L. A. Marsh, J. L. Feuston, J. Ruppert, J. R. Brubaker, and D. A. Szafir, “Scholastic: Graphical Human-AI Collaboration for Inductive and Interpretive Text Analysis,” in UIST 2022 - Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, Association for Computing Machinery, Inc, Oct. 2022. doi: 10.1145/3526113.3545681.

[10] R. P. Gauthier and J. R. Wallace, “The Computational Thematic Analysis Toolkit,” in Proceedings of the ACM on Human-Computer Interaction, Association for Computing Machinery, Jan. 2022. doi: 10.1145/3492844.

[11] J. V. Braga, T. B. Pontes e Silva, and V. T. Souto, “Manipulações estatísticas e anomalias visuais: design de visualização de dados e reconhecimento de vieses estatísticos,” InfoDesign - Revista Brasileira de Design da Informação, vol. 17, no. 2, pp. 145–162, 2020.

[12] J. C. Gomes, R. M. Pimenta, and M. Schneider, “MINERAÇÃO DE DADOS NA PESQUISA EM CIÊNCIA DA INFORMAÇÃO: DESAFIOS E OPORTUNIDADES,” in GT-1 – Estudos históricos e epistemológicos da Ciência da Informação, 2019.

[13] M. T. Rodríguez, S. Nunes, and T. Devezas, “Telling stories with data visualization,” in NHT 2015 - Proceedings of the 2015 Workshop on Narrative and Hypertext - co-located with HT 2015, 2015, pp. 7–11. doi: 10.1145/2804565.2804567.

[14] J. Gray, “The data epic: Visualization practices for narrating life and death at a distance,” in Data Visualization in Society, 2020, pp. 313–328. doi: 10.5117/9789463722902.

[15] F. Cabitza, A. Locoro, D. Fogli, and M. Giacomin, “Valuable Visualization of Healthcare Information : From the Quantified Self Data to Conversations,” pp. 376–380, 2016, doi: http://dx.doi.org/10.1145/2909132.2927474.

[16] H. Kennedy and M. Engebretsen, Data Visualization in Society. 2020. doi: 10.5117/9789463722902.

[17] T. U. Nærland, “The political significance of data visualization: Four key perspectives,” in Data Visualization in Society, 2020, pp. 63–74. doi: 10.5117/9789463722902.

[18] E. Beauxis-Aussalet et al., “The Role of Interactive Visualization in Fostering Trust in AI,” IEEE Comput Graph Appl, vol. 41, no. 6, pp. 7–12, 2021, doi: 10.1109/MCG.2021.3107875.

[19] B. J. J. Otten, K. Cheng, and A. Drewnowski, “Infographics And Public Policy: Using Data Visualization To Convey Complex Information,” Health Aff, vol. 34, no. 11, pp. 1901–1907, 2015, doi: 10.1377/hlthaff.2015.0642.

[20] J. W. Rettberg, “Ways of knowing with data visualizations,” in Data Visualization in Society, 2020, pp. 35–48. doi: 10.5117/9789463722902.

[21] K. Hepworth, “Editorial Big Data Visualization : Promises & Pitfalls,” Communication Design Quarterly, vol. 4, no. 4, pp. 7–19, 2016, doi: 10.1145/3071088.3071090.

[22] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini, “The persuasive power of data visualization,” IEEE Trans Vis Comput Graph, vol. 20, no. 12, pp. 2211–2220, 2014, doi: 10.1109/TVCG.2014.2346419.

[23] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini, “How Deceptive are Deceptive Visualizations?: An Empirical Analysis of Common Distortion Techniques,” in CHI 2015, Association for Computing Machinery, 2015, pp. 1469–1478. doi: 10.1145/2702123.2702608.

[24] S. Eichner, Agency and Media Reception Experiencing Video Games, Film, and Television, 1st ed. Springer VS, Wiesbaden, 2014. doi: https://doi.org/10.1007/978-3-658-04673-6.

[25] T. Medeiros and V. T. Souto, “Design Space para Jornalismo de Dados,” pp. 122–130, 2019, doi: 10.5151/9cidi-congic-1.0095.

[26] F. B. Viégas and M. Wattenberg, “Artistic data visualization: Beyond visual analytics,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2007, pp. 182–191. doi: 10.1007/978-3-540-73257-0_21.

[27] E. S. Tønnessen, “What is visual-numeric literacy, and how does it work?,” in Data Visualization in Society, 2020, pp. 189–206. doi: 10.5117/9789463722902.

[28] L. Ugaya Mazza, “DIGITAL GRAVEYARDS: A STUDY ABOUT DATA VISUALIZATION PORTRAYING DEATH DURING THE COVID-19 PANDEMIC,” Universidade Federal do Paraná, Curitiba, 2022. Accessed: Oct. 23, 2022. [Online]. Available: https://hdl.handle.net/1884/77435

[29] B. Lubars and C. Tan, “Ask Not What AI Can Do, But What AI Should Do: Towards a Framework of Task Delegability,” ACM, Feb. 2022, doi: https://doi.org/10.1145/nnnnnnn.nnnnnnn.

[30] A. Fügener, J. Grahl, A. Gupta, and W. Ketter, “Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation,” Information Systems Research, vol. 33, no. 2, pp. 678–696, Jun. 2022, doi: 10.1287/isre.2021.1079.

[31] D. Anderson, “GLAAD and HRC call on Stanford University & responsible media to debunk dangerous & flawed report claiming to identify LGBTQ people through facial recognition technology,” GLAAD, Sep. 08, 2017. https://www.glaad.org/blog/glaad-and-hrc-call-stanford-university-responsible-media-debunk-dangerous-flawed-report (accessed Apr. 06, 2023).

[32] S. Shaffi, “‘It’s the opposite of art’ why illustrators are furious about AI Art,” The Guardian, Jan. 23, 2023. https://www.theguardian.com/artanddesign/2023/jan/23/its-the-opposite-of-art-why-illustrators-are-furious-about-ai (accessed Apr. 06, 2023).

[33] B. Edwards, “Netflix stirs fears by using AI-assisted background art in short anime film Ars Technica,” ARS Technica, Feb. 01, 2023. https://arstechnica.com/information-technology/2023/02/netflix-taps-ai-image-synthesis-for-background-art-in-the-dog-and-the-boy/ (accessed Apr. 06, 2023).

[34] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A model for types and levels of human interaction with automation,” IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans., vol. 30, no. 3, pp. 286–297, 2000, doi: 10.1109/3468.844354.

[35] R. M. Ryan and E. L. Deci, “Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being,” American Psychologist, vol. 55, no. 1, pp. 68–78, 2000, doi: 10.1037110003-066X.55.1.68.

[36] M. Csikszentmihalyi, Flow – The Psychology of optimal experience. 1990. [Online]. Available: https://www.researchgate.net/publication/224927532

[37] A. Cubillo, R. Halari, A. Smith, E. Taylor, and K. Rubia, “A review of fronto-striatal and fronto-cortical brain abnormalities in children and adults with Attention Deficit

Hyperactivity Disorder (ADHD) and new evidence for dysfunction in adults with ADHD during motivation and attention," Cortex, vol. 48, no. 2. pp. 194–215, Feb. 2012. doi: 10.1016/j.cortex.2011.04.007.

[38] A. Bandura, "Toward a Psychology of Human Agency," 2006.

[39] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," Hum Factors, vol. 46, no. 1, pp. 50–80, 2004.

[40] E. T. Kazmierczak, "Design as Meaning Making: From Making Things to the Design of Thinking," Design Issues, vol. 19, no. 2, pp. 45–59, 2003, doi: 10.1162/074793603765201406.

[41] B. Hjørland, "Data: With big data and database semantics," ISKO Encyclopedia of Knowledge Organization home, 2018. https://www.isko.org/cyclo/domain_analysis#top (accessed May 31, 2021).

[42] R. R. S. Oliveira, T. A. Alquete, S. G. Coutinho, and S. B. Campello, "O Design da Informação, Instrucional e de Interação sob uma perspectiva de uso em Artefatos Digitais de Aprendizagem," in 6o Congresso Internacional de Design da Informação, 5o InfoDesign Brasil 6o Congic, 2013.

[43] J. D. Rawlins and G. D. Wilson, "Agency and interactive data displays: Internet graphics as Co-Created rhetorical spaces," Technical Communication Quarterly, vol. 23, no. 4, pp. 303–322, Oct. 2014, doi: 10.1080/10572252.2014.942468.

[44] B. Cooley, "Data visualization as Grief," Nightingale, 2020. https://medium.com/nightingale/data-visualization-as-grief-599dc6536e6f (accessed May 29, 2021).

[45] F. Cabitza, A. Locoro, D. Fogli, and M. Giacomin, "Valuable visualization of healthcare information: From the quantified self data to conversations," in Proceedings of the Workshop on Advanced Visual Interfaces AVI, Association for Computing Machinery, Jun. 2016, pp. 376–380. doi: 10.1145/2909132.2927474.

[46] V. E. Lechner, "What a line can say: Investigating the semiotic potential of the connecting line in data visualizations," in Data Visualization in Society, 2020, pp. 329–346. doi: 10.5117/9789463722902.

[47] Y. von Engelhardt, The language of graphics. 2002. [Online]. Available: http://dare.uva.nl

[48] R. E. Horn, Visual Language: Global Communication for the 21st Century. 1998.

[49] K. Börner, A. Bueckle, and M. Ginda, “Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments,” Proc Natl Acad Sci U S A, vol. 116, no. 6, pp. 1857–1864, 2019, doi: 10.1073/pnas.1807180116.

[50] S. Doan, “Misrepresenting COVID-19 : Lying With Charts During the Second Golden Age of Data Design,” J Bus Tech Commun, pp. 1–7, 2020, doi: 10.1177/1050651920958392.

[51] W. Weber, “Exploring narrativity in data visualization in journalism,” in Data Visualization in Society, 2020, pp. 295–312. doi: 10.5117/9789463722902.

[52] M. Ryan, “Narrative, Media and Modes,” in Avatars of Story, 2006.

[53] S. Brinch, “What we talk about when we talk about beautiful data visualizations,” in Data Visualization in Society, 2020, pp. 259–276. doi: 10.5117/9789463722902.

[54] J. Murray, “Agency,” in Hamlet on the Holodeck: the future of narrative in cyberspace, 1997, pp. 126–153.

[55] K. Tanenbaum and J. Tanenbaum, “Commitment to Meaning. A Reframing of Agency in Games,” After Media — Embodiment and Context, Digital Arts and Culture 2009, Arts Computation Engineering, UC Irvine, vol. 3, no. 2, p. 10, 2009, [Online]. Available: https://escholarship.org/uc/item/6f49r74n

[56] J. Bizzocchi, “Run, Lola, Run - Film as Narrative Database,” fourth Media in Transition conference, pp. 1–14, 2005, [Online]. Available: http://public.csusm.edu/diekman/vsar309f09/bizzocchi.pdf

[57] J. Bizzocchi, “Ceremony of Innocence and the Subversion of Interface: Cursor Transformation as a Narrative Device,” MelbourneDAC, the 5th International Digital Arts and Culture Conference, 2003, [Online]. Available: http://hypertext.rmit.edu.au/dac/papers/Bizzocchi.pdf

[58] M. Maguire, “Methods to support human-centred design,” International Journal of Human Computer Studies, vol. 55, no. 4, pp. 587–634, 2001, doi: 10.1006/ijhc.2001.0503.

[59] M. Leyer and S. Schneider, “ME, YOU OR AI? HOW DO WE FEEL ABOUT DELEGATION,” in 27th European Conference on Information Systems (ECIS), 2019. [Online]. Available: https://aisel.aisnet.org/ecis2019_rp/36

[60] A. Dos Santos, Seleção do Método de Pesquisa. 2018. [Online]. Available: https://editorainsight.com.br/wp-content/uploads/2019/04/af_Metodos-de-pesquisa_web.pdf

[61] C. Abras, D. Maloney-Krichmar, and J. Preece, “User-Centered Design,” Encyclopedia of Human-Computer Interaction, 2004.

[62] A. D. V. Dabbs et al., “User-centered design and interactive health technologies for patientsi,” CIN - Computers Informatics Nursing, vol. 27, no. 3, pp. 175–183, May 2009, doi: 10.1097/NCN.0b013e31819f7c7c.

[63] L. M. Fadel, “O método close reading aplicado à análise de mídias interativas.” Universidade Federal de Santa Catarina. [S.l.]., 2020.

[64] J. Bizzocchi and J. Tanenbaum, “Well Read : Applying Close Reading Techniques to Gameplay Experiences,” Well Played, vol. 3, no. January, pp. 1–16, 2011.

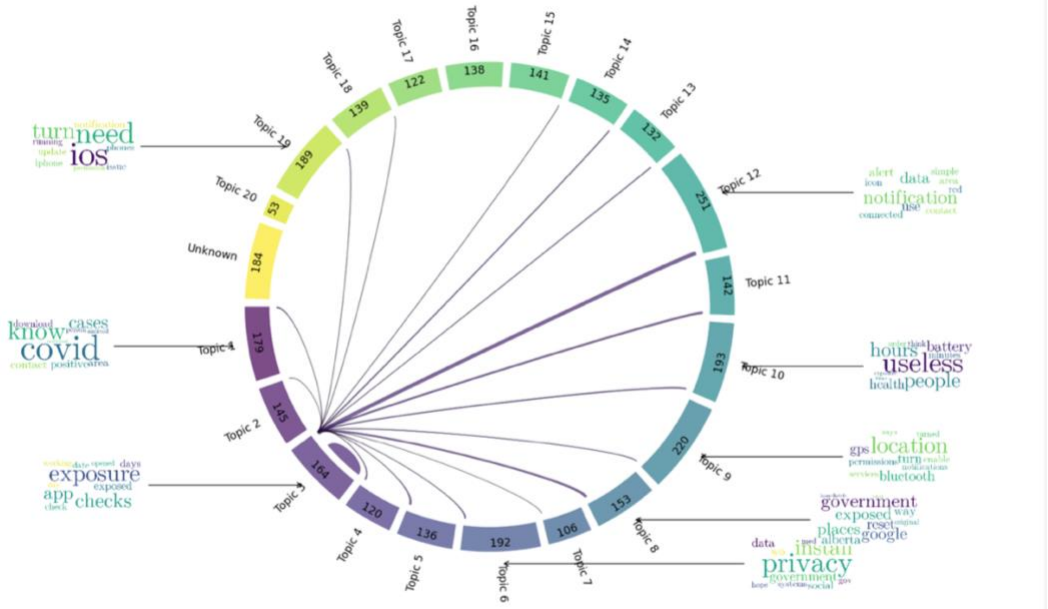
[65] K. Malterud, V. D. Siersma, and A. D. Guassora, “Sample Size in Qualitative Interview Studies: Guided by Information Power,” Qual Health Res, vol. 26, no. 13, pp. 1753–1760, Nov. 2016, doi: 10.1177/1049732315617444.

[66] S. H. Stephens, “Designer perceptions of user agency during the development of environmental risk visualization tools,” in SIGDOC 2017 - 35th ACM International Conference on the Design of Communication, Association for Computing Machinery, Inc, Aug. 2017. doi: 10.1145/3121113.3121223.

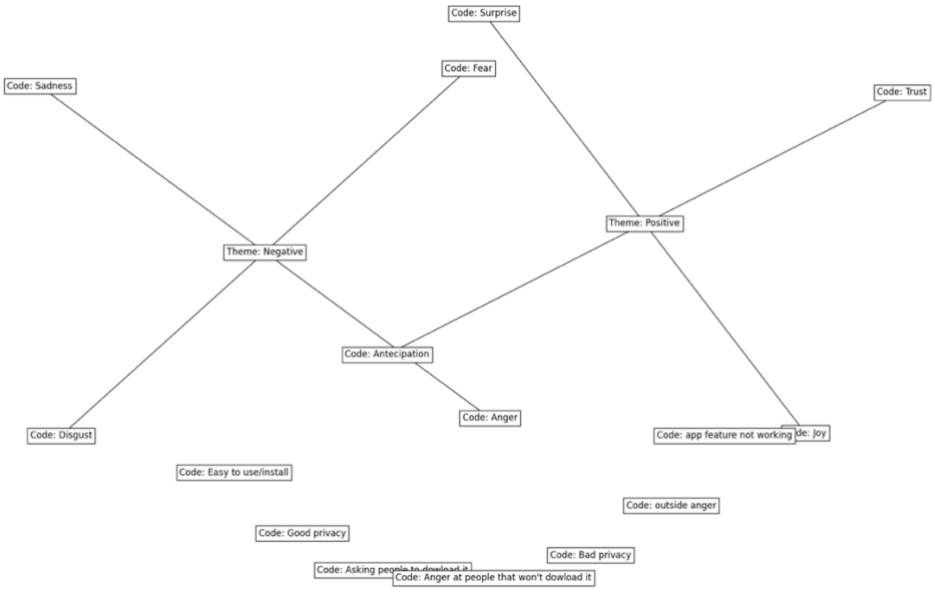
Appendix A

Data visualizations included in the sample analysis

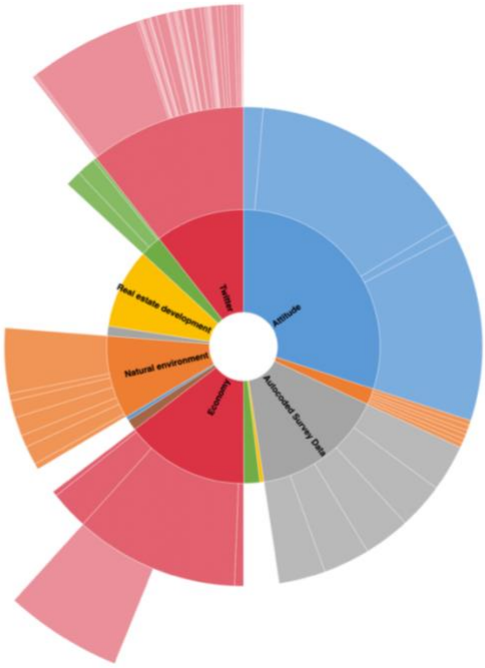
CTA statistical linking diagram.



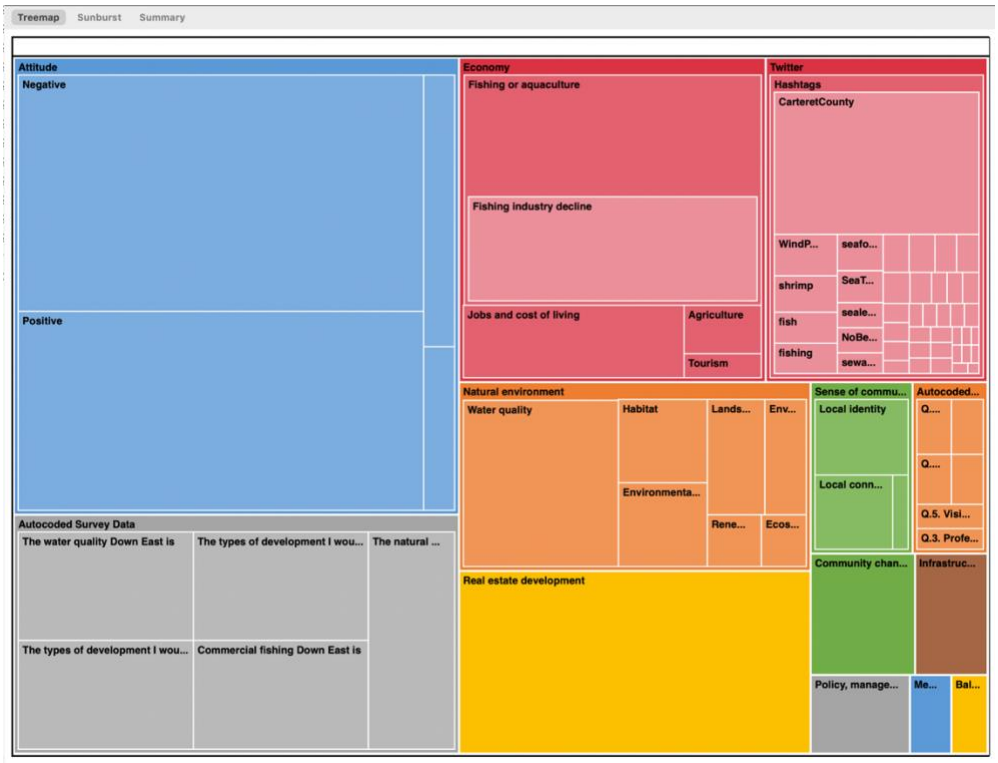
CTA linking diagram.



NVIVO pie chart.



NVIVO treemap.



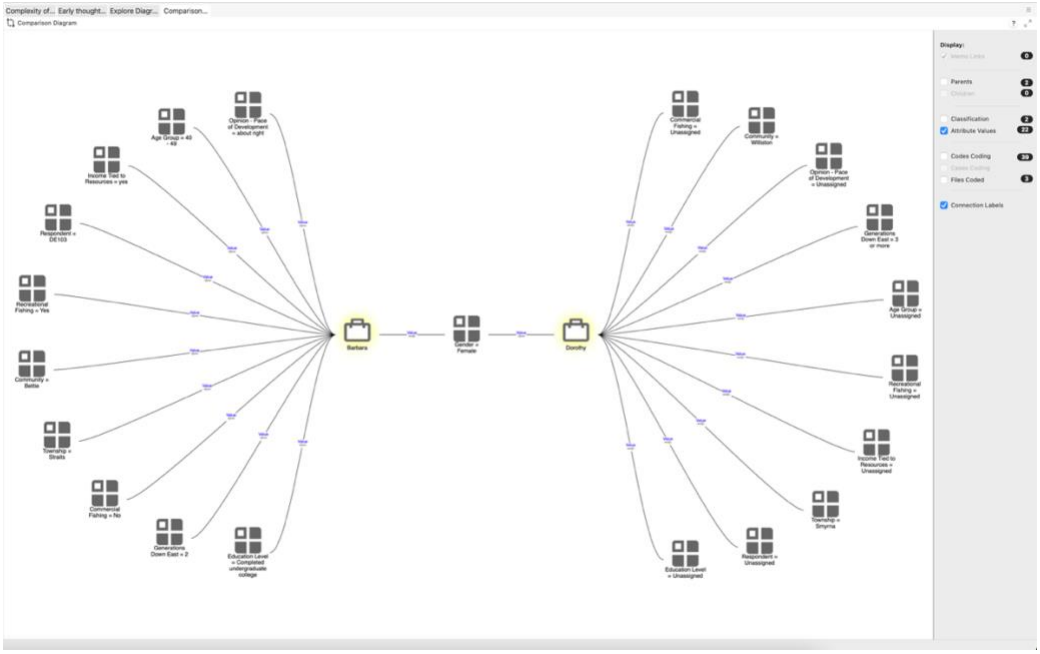
NVIVO table.

Coding Matrix

	A : Agriculture	B : Fishing or aquaculture	C : Jobs and cost of living	D : Tourism
1 : Mixed	0	7	4	0
2 : Negative	5	97	20	1
3 : Neutral	0	7	4	0
4 : Positive	0	31	20	1

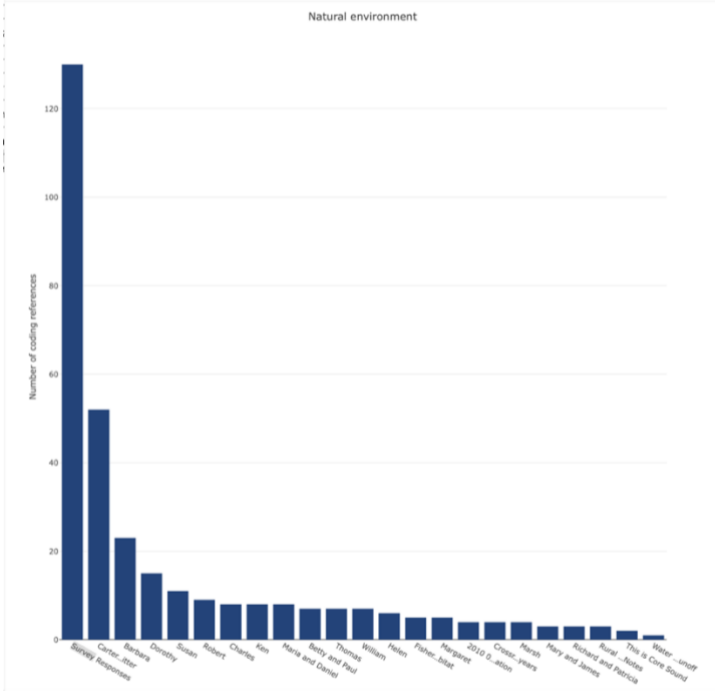
NVIVO word cloud.

NVIVO linking diagram.

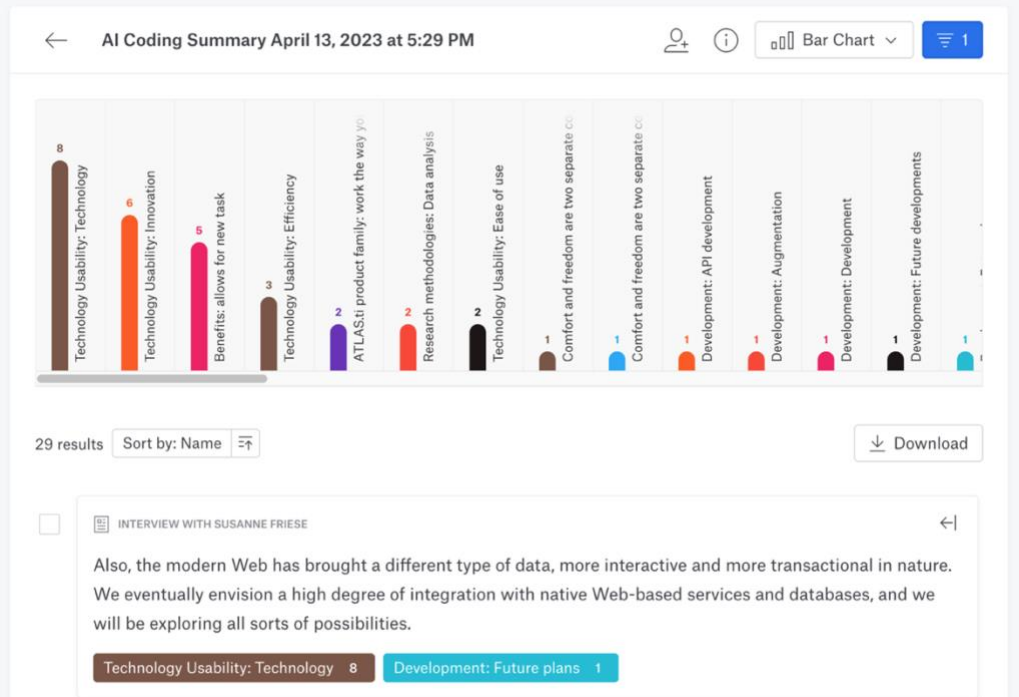


NVIVO word linking diagram.

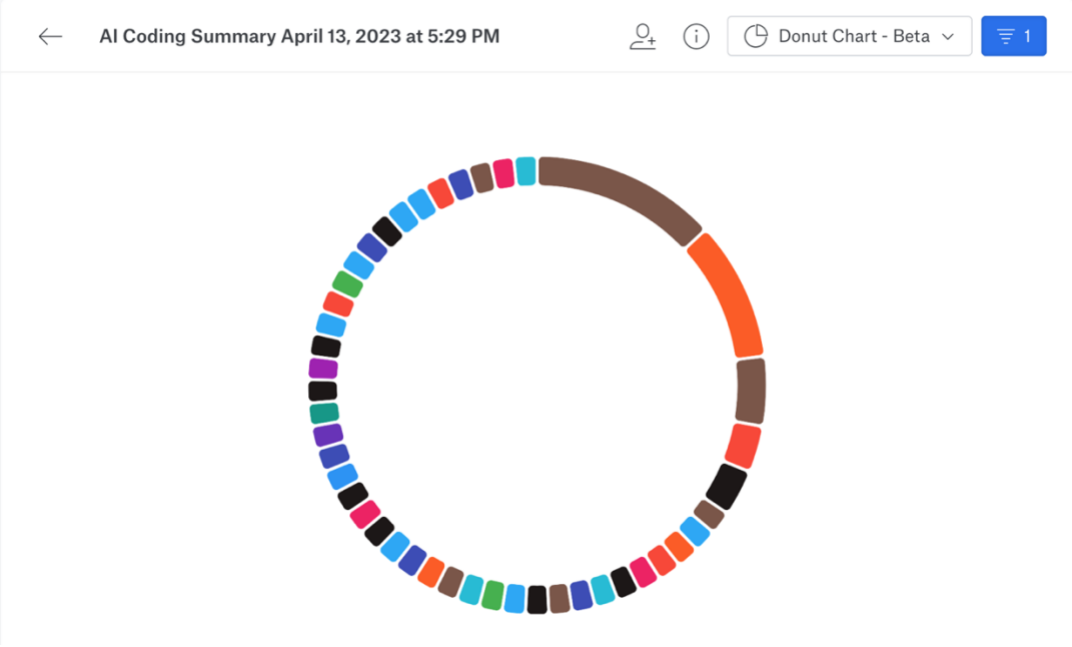
NVIVO statistical bar chart.



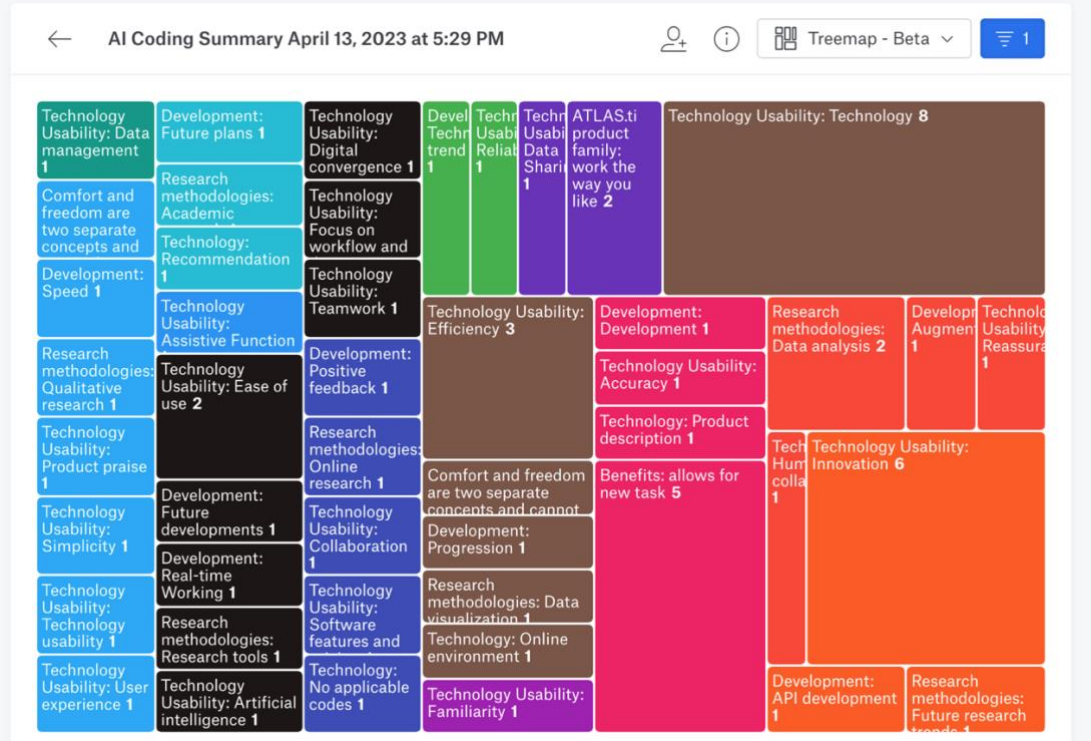
ATLAS.ti Statistical Bar Chart.



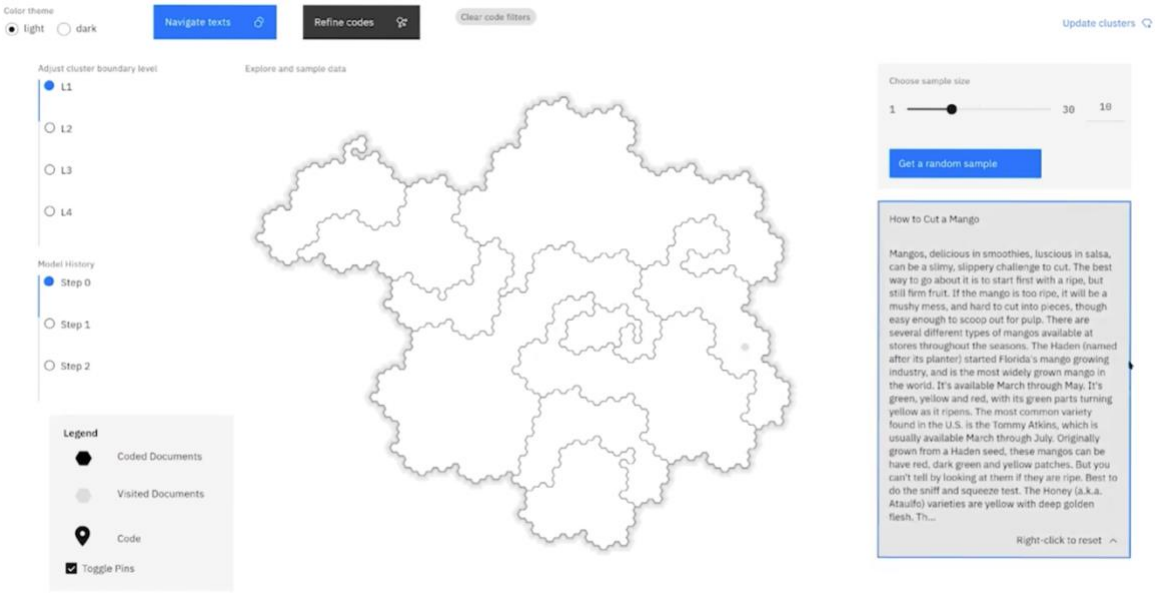
ATLAS.ti Statistical Pie Chart.



ATLAS.ti statistical treemap



Scholastic Statistical Map.



Appendix B

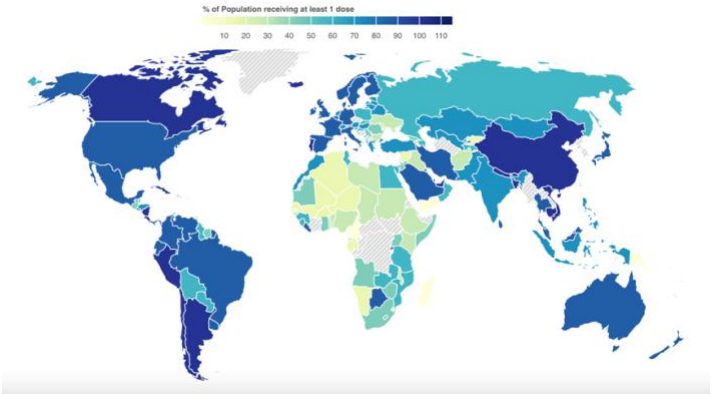
Analysis protocol developed for the Close Reading

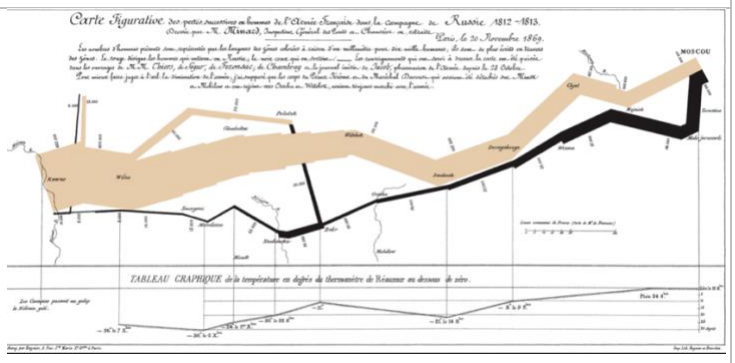
		Sample of software			
Agency	**Elements that foster agency**	CTA	NVIVO	ATLAS.ti	Scholastic
Narrative	Proximity to establish connections.				
	Focus on showing connections.				
	Use of frequency as a storytelling tool.				
Choice	Editing capabilities.				
	Explanation of how the AI derived results from the data (processes).				
	Guidance.				
Action	Filtering options.				
	Clickable options that reveal more data.				
Space	Color coding.				
	Nesting of windows on top of each other.				
	Breadcrumbs to let you know where you came from and how to return to it.				
	Clustering of similar elements.				

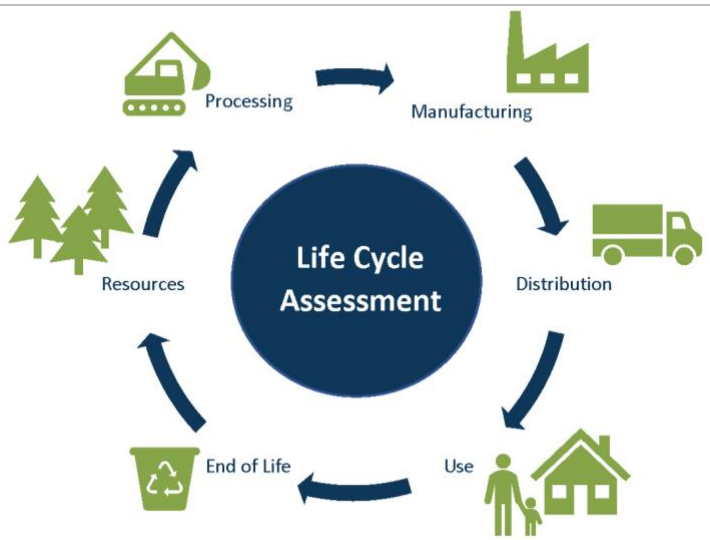
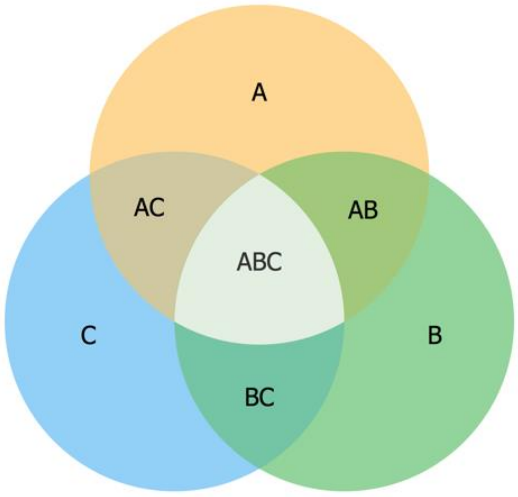
Appendix C

Interview guide phase 1, understanding how qualitative researchers use data visualizations to assist in data analysis

1. What is your main research area?
2. What is your usual process of conducting data analysis?
    a. Prompt for any use of visualizations they might use.
3. Could you walk me through your usual process of conducting data analysis?
    a. Prompt for how they use data visualization in the middle of the process.
    b. Prompt for what they would like data visualizations to do for them in their research.
4. These are the most common types of data visualization; do you use any of them when conducting data analysis?
    a. Present a list with images of data visualizations.

Statistical Maps	 Source: Vaccination progress across the world (regarding COVID-19 vaccines) John Hopkins University retrieved March 16th 2023, 12:47. https://coronavirus.jhu.edu/vaccines/international

Path Maps	 Source: tracking shark – You can now track sharks off the East Cost in real time. The Verge retrieved March 16, 2023 12:54 https://www.theverge.com/2013/8/29/4671128/shark-tracking-in-real-time-ocearch-global-tracker
Statistical Maps	 Source: Charles Joseph Minard Issues One of the Best Statistical Graphics Ever Drawn. History of Information. Retrieved March 16, 2023 12:57 https://www.historyofinformation.com/detail.php?entryid=3204

Link diagram	 *Source:* Life Cycle Assessment. NCASI Technical Studies Program. March 16, 2023 13:25 https://www.ncasi.org/technical-studies/sustainable-manufacturing/life-cycle-assessment/
Grouping diagram	 *Source:* Venn Diagrams. In: Concept Draw. Retrieved from https://www.conceptdraw.com/solution-park/diagram-venn March 16th 2023 13:30

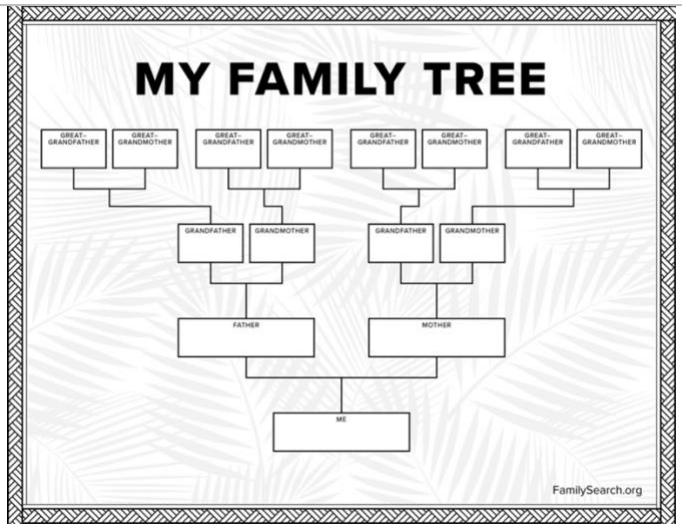
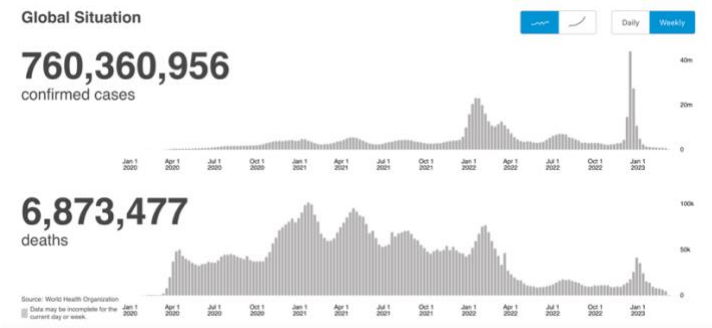
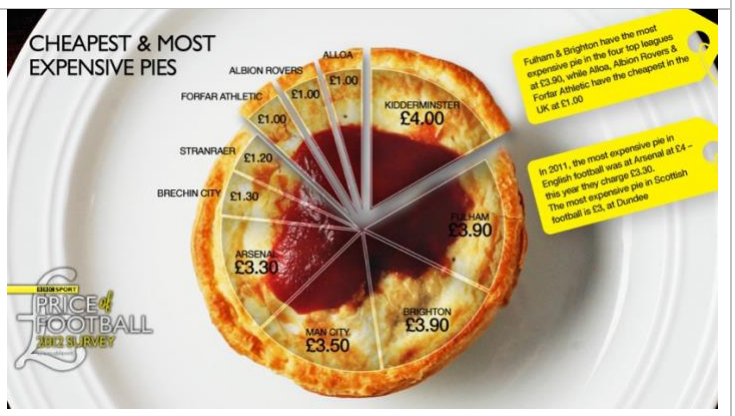
Chronological link diagram	 *Source:* Free Printable Family Tree Templates and Online Family Tree Ideas. IN: Family Search Blog. Retrieved 16 March 2023 13:33 https://www.familysearch.org/en/blog/family-tree-templates-family-tree-make
Statistical link diagram	 *Source:* population dynamics. In: JimPintoBlog, retrieve 16 March 2023 13:36 https://jimpintoblog.blogspot.com/2015/09/population-dynamics_9.html

Table	 *Source:* Table with Column PowerPoint Template and Keynote Slide. In: Slide Bazaar. 16 March 2023 13:39. https://slidebazaar.com/items/table-with-column-powerpoint-template/
Time chart	 *Source:* Global Situation of COVID-19 pandemic. In: World Health Organization (WHO). Retrieved March 16, 2023 13:44 https://covid19.who.int/
Statistical chart	 *Source:* Frontiers in pie-charts. In: StatsChat. Retrieved march 16 2023, 13:47 https://www.statschat.org.nz/2013/05/06/frontiers-in-pie-charts/

b. (Optional) How do you use these data vis during your data analysis.

c. (Optional) Do you use any other kinds of data visualization?

5. Is there a tool that you use for organizing data?

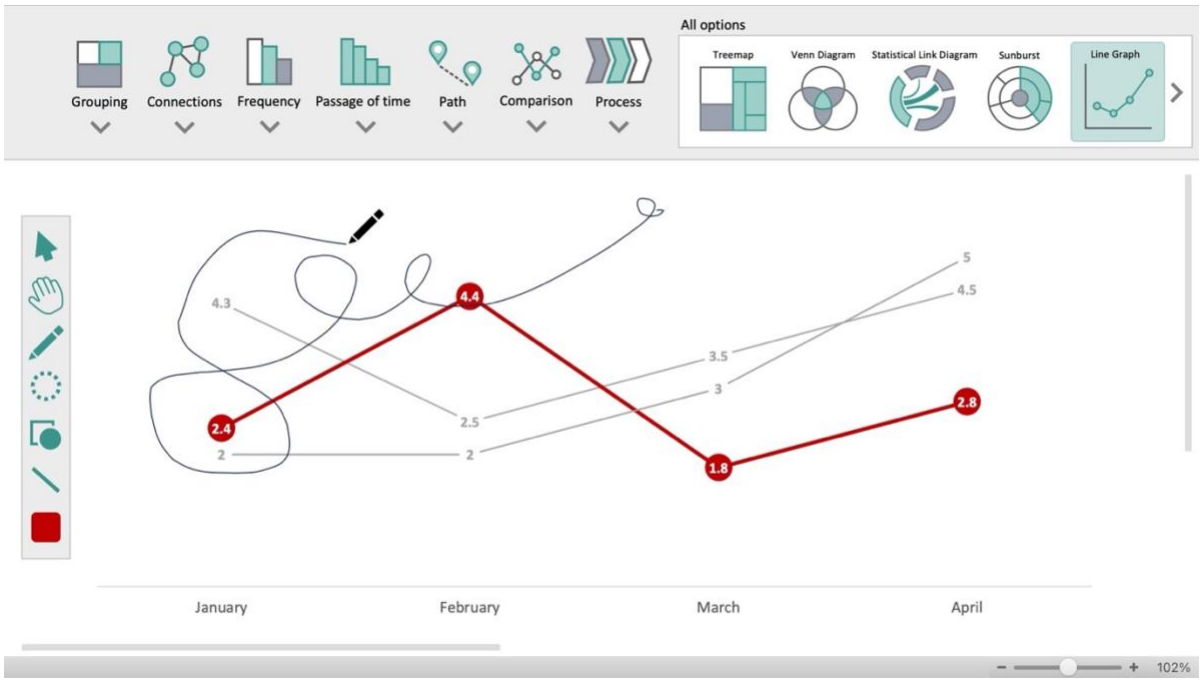
   a. (Optional) How do you like to use the tool?

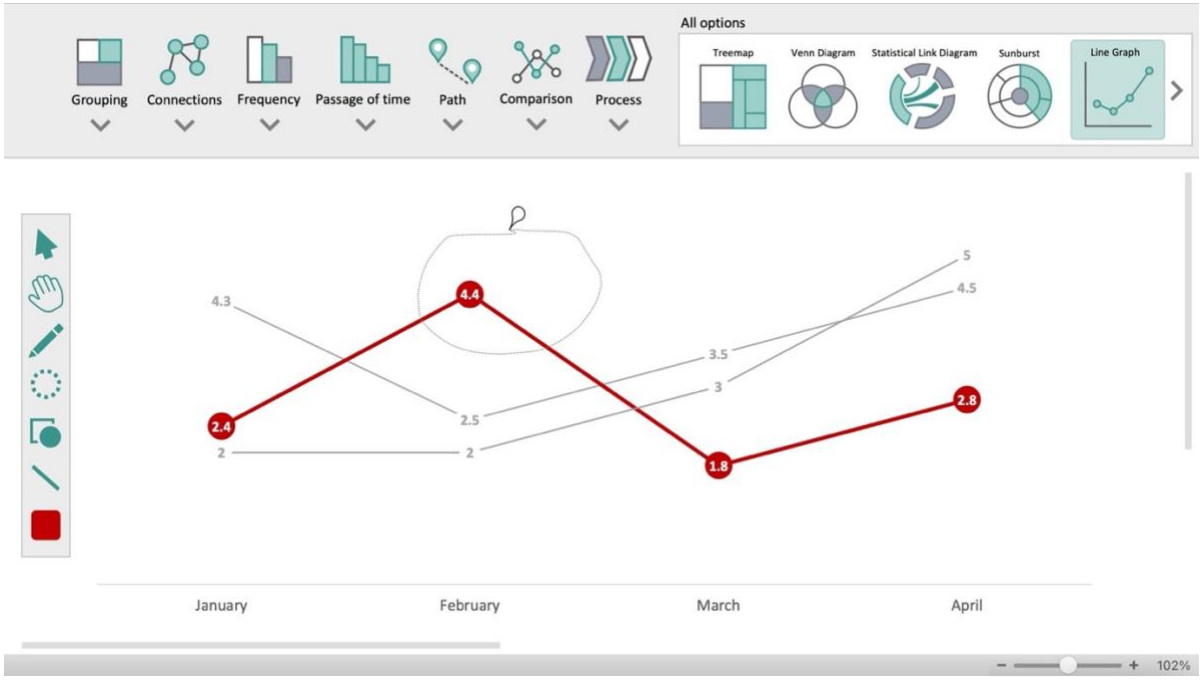
6. Is there any way you like to visualize data?

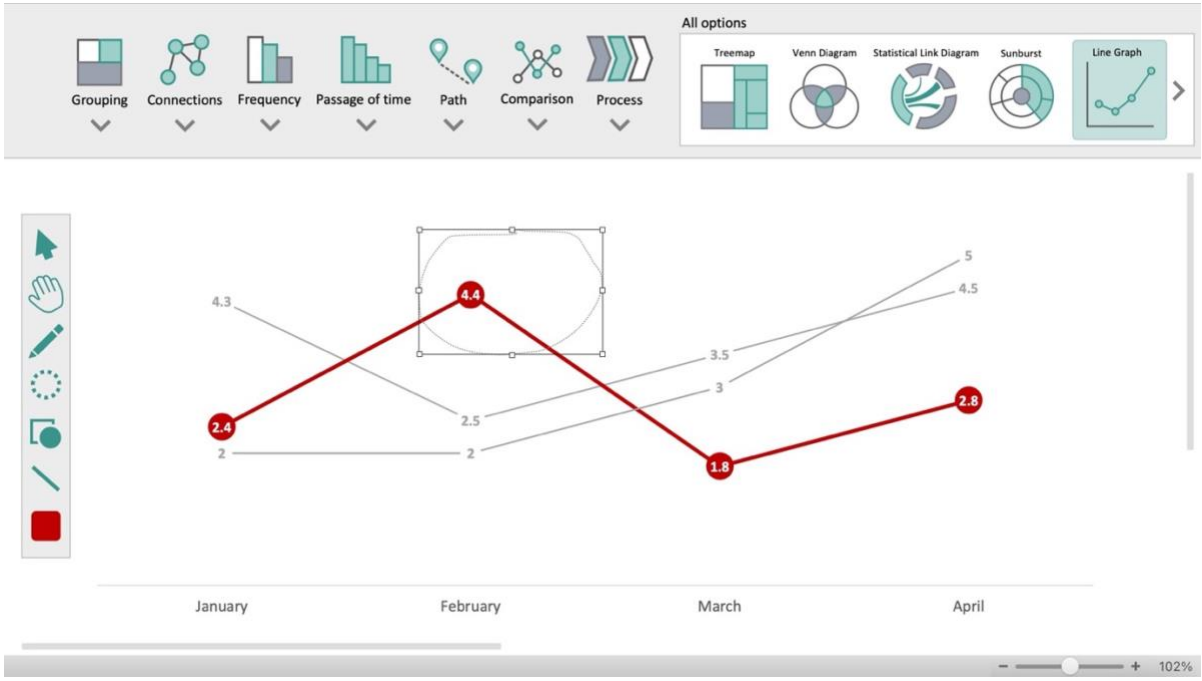
7. Is there anything else you'd like to add/discuss?

Appendix D

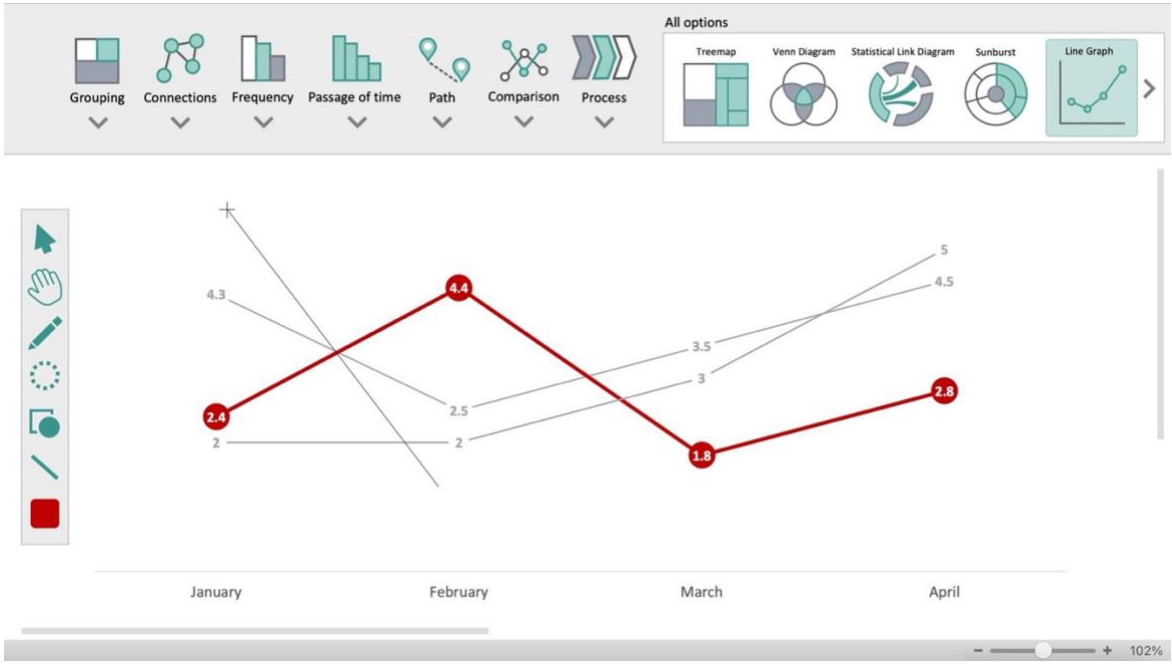
First Prototype




Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
4.3
4.4
2.5
2.4
2
2
3.5
3
5
4.5
2.8
1.8
January
February
March
April
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
4.3
4.4
2.5
2.4
2
2
3.5
3
5
4.5
2.8
1.8
January
February
March
April
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
4.3
4.4
2.5
2.4
2
2
3.5
3
5
4.5
2.8
1.8
January
February
March
April
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
4.3
4.4
2.5
2.4
2
2
3.5
3
5
4.5
2.8
1.8
January
February
March
April
102%

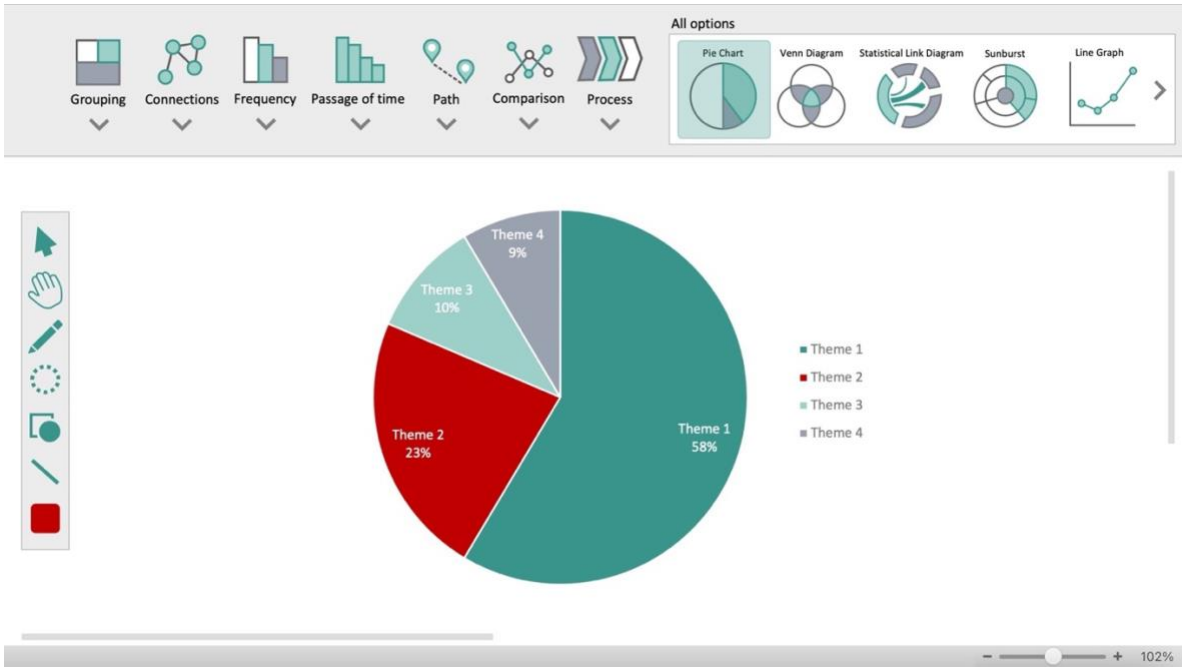

Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
4.3
4.4
2.5
3.5
3
2
5
4.5
2.8
1.8
January
February
March
April
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Venn Diagram
Treemap
Statistical Link Diagram
Sunburst
Leaf 5
Leaf 8
Stem 3
Leaf 9
Leaf 11
Stem 4
Branch 2
Stem 2
Branch 1
Leaf 4
Leaf 16
Leaf 12
Stem 5
Leaf 13
Leaf 15
Branch 3
Stem 6
Leaf 14
Stem 1
Leaf 1
Leaf 6
Leaf 3
Leaf 2
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Mind Map
Group connection
Statistical Link Diagram
Hierarchy
Path Map
Chronological link diagram
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Map
Histogram
Statistical Link Diagram
Treemap
Sunburst
Pie chart
Bar graph
Line graph
Statistical time map
Table
102%

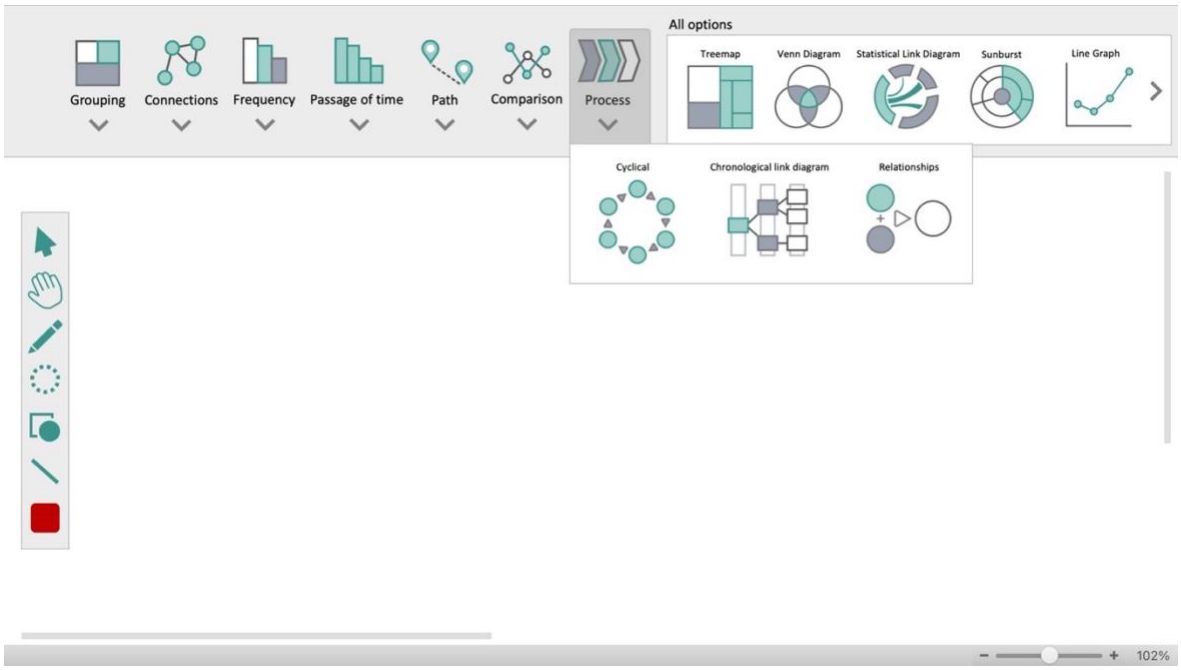

All options
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Map
Histogram
Statistical Link Diagram
Treemap
Sunburst
Pie chart
Bar graph
Line graph
Statistical time map
Table
102%


All options
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Pie Chart
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Theme 4
9%
Theme 3
10%
Theme 2
23%
Theme 1
58%
Theme 1
Theme 2
Theme 3
Theme 4
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Histogram
Chronological link diagram
Line graph
Statistical time map
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Path Map
Statistical time map
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Map
Table
Multiple line graph
Pie chart
Bar graph
Histogram
Statistical Link Diagram
102%


Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
All options
Treemap
Venn Diagram
Statistical Link Diagram
Sunburst
Line Graph
Cyclical
Chronological link diagram
Relationships
102%

Appendix E

Interview guide phase 2 (user needs and wants)

Introduction:

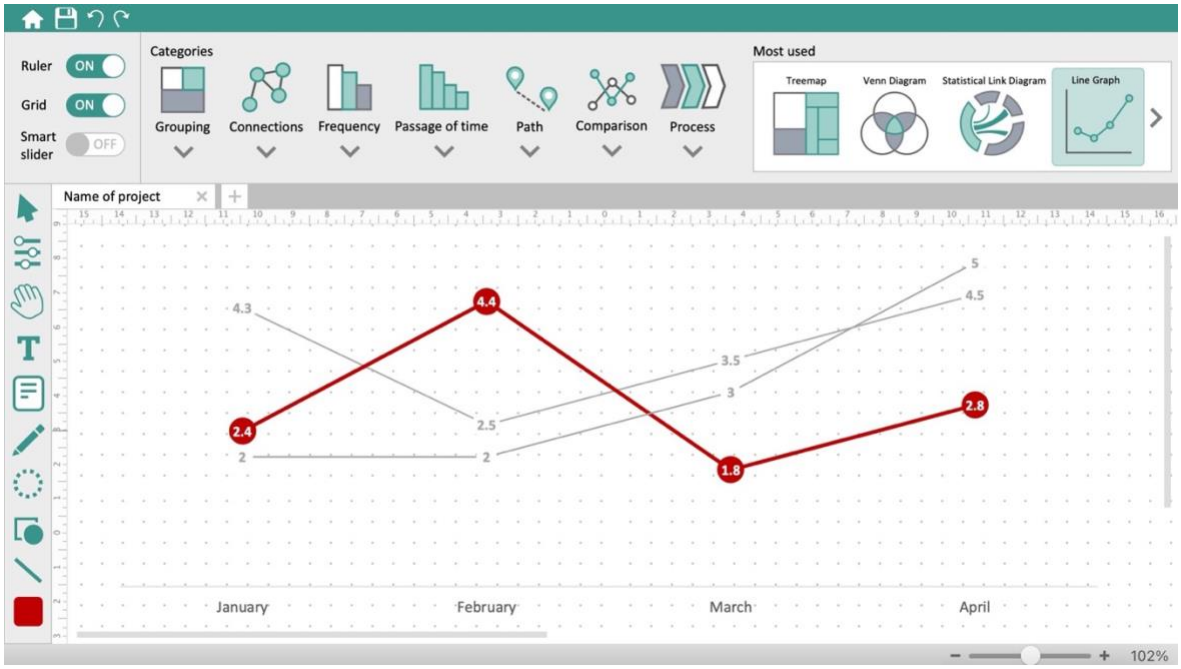
1. Welcome the participant back.
2. Share the results from interview 1 and the template for data visualizations that was developed following their insights.
    i. The template will be a series of images on a ZOOM whiteboard so the participant can sketch over them and move elements around as they desire.
3. Walk the participant to the template so they can familiarize themselves with the design and understand how the data visualization was made.

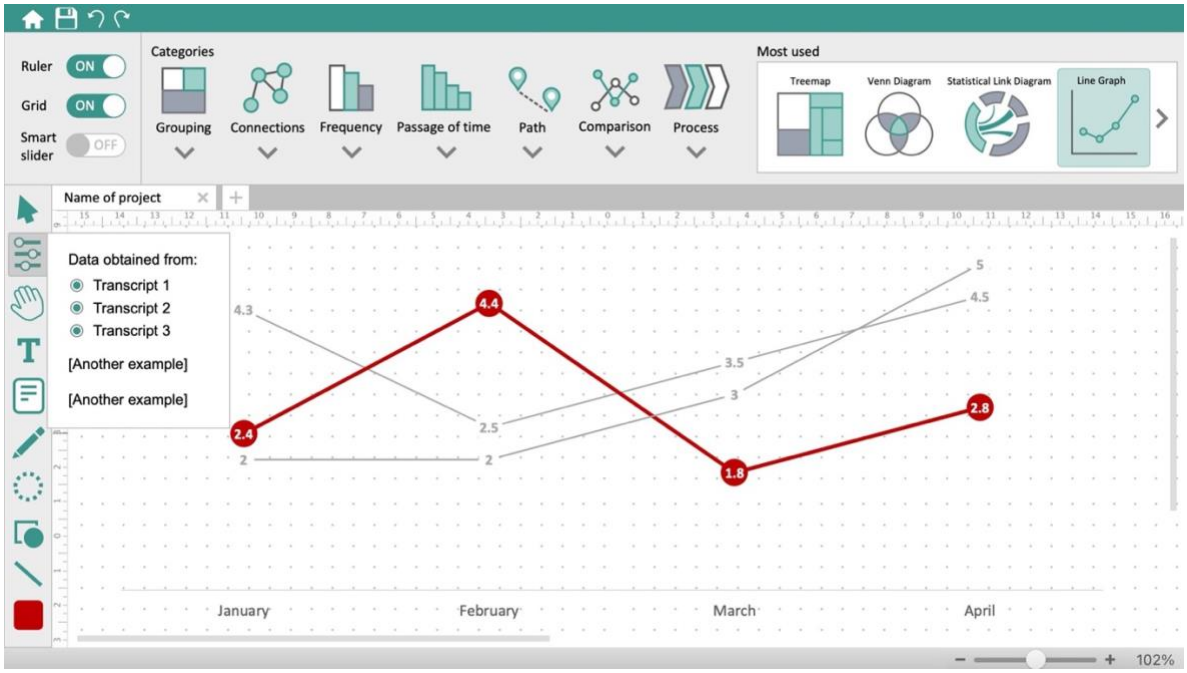
Start the interview process.

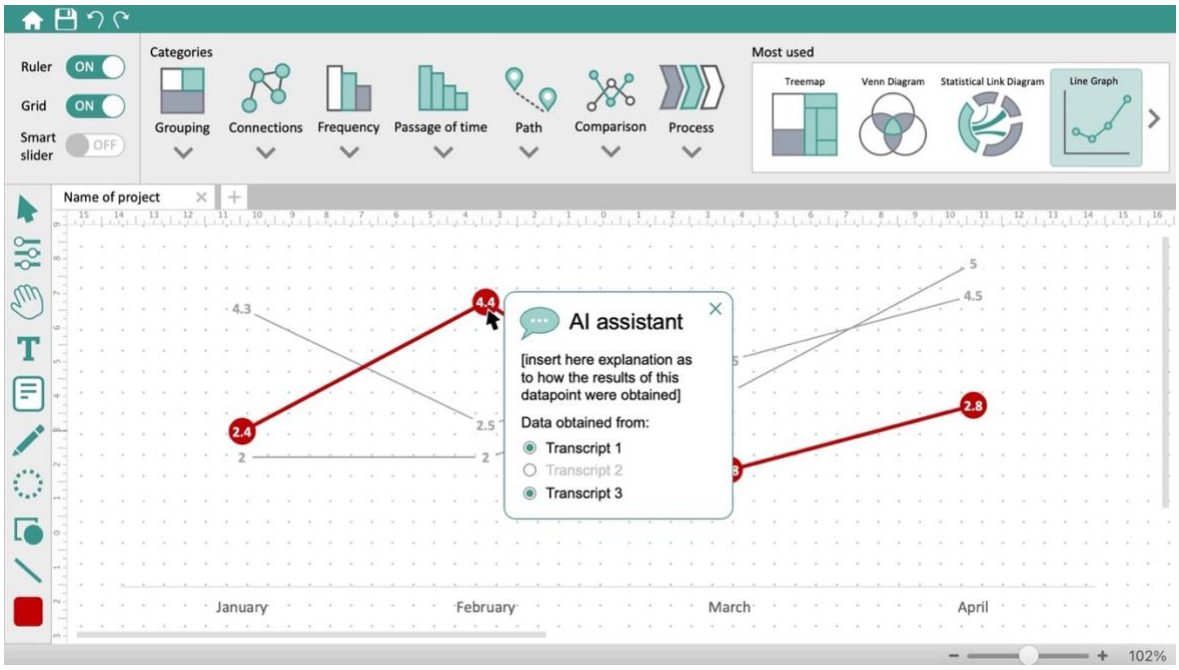
1. Ask the participant:
    i. What do you think will be able to be done with this?
        1. (Optional) Prompt for specifics on how they would use specific design elements they show interest in.
        2. (Optional) Prompt for any elements that they might be unsatisfied with.
    ii. How would you use this to assist in data analysis?
    iii. Is there anything else you would like to add or change?

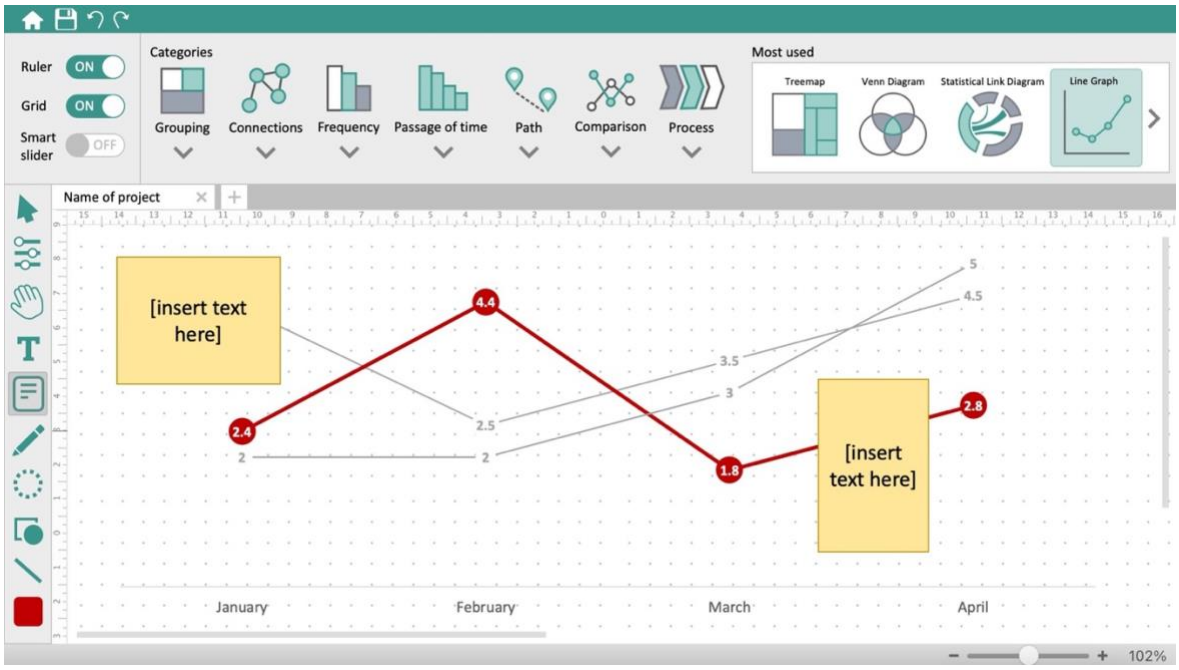
Appendix F

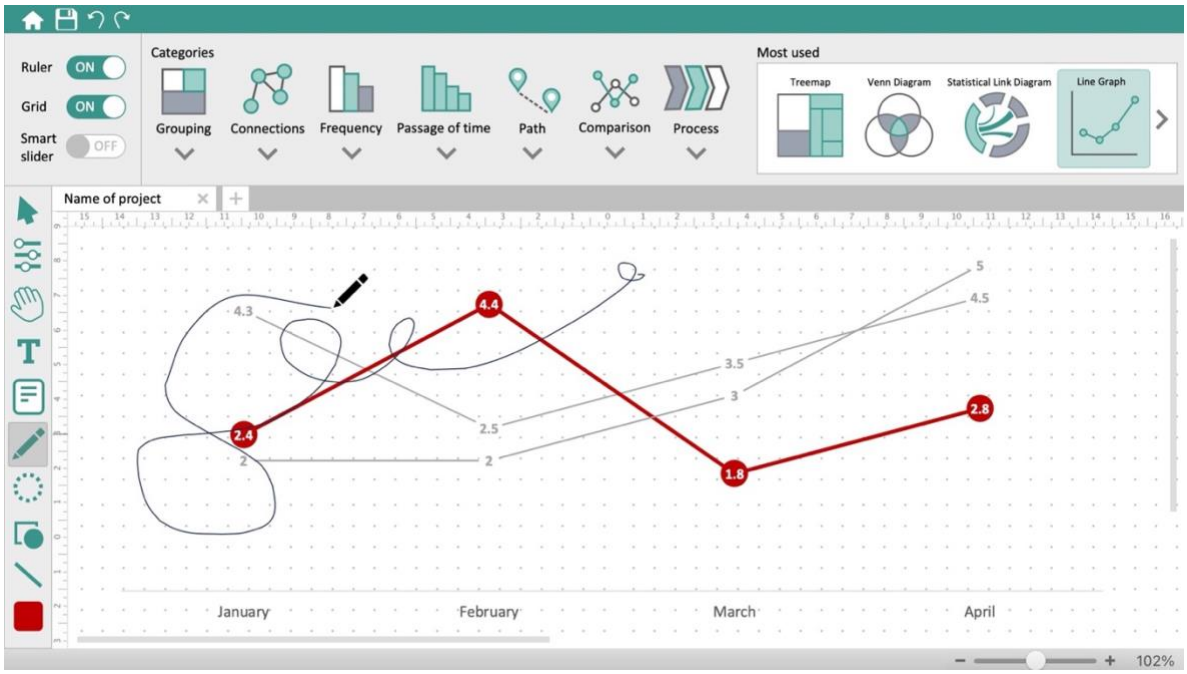
Updated Prototype

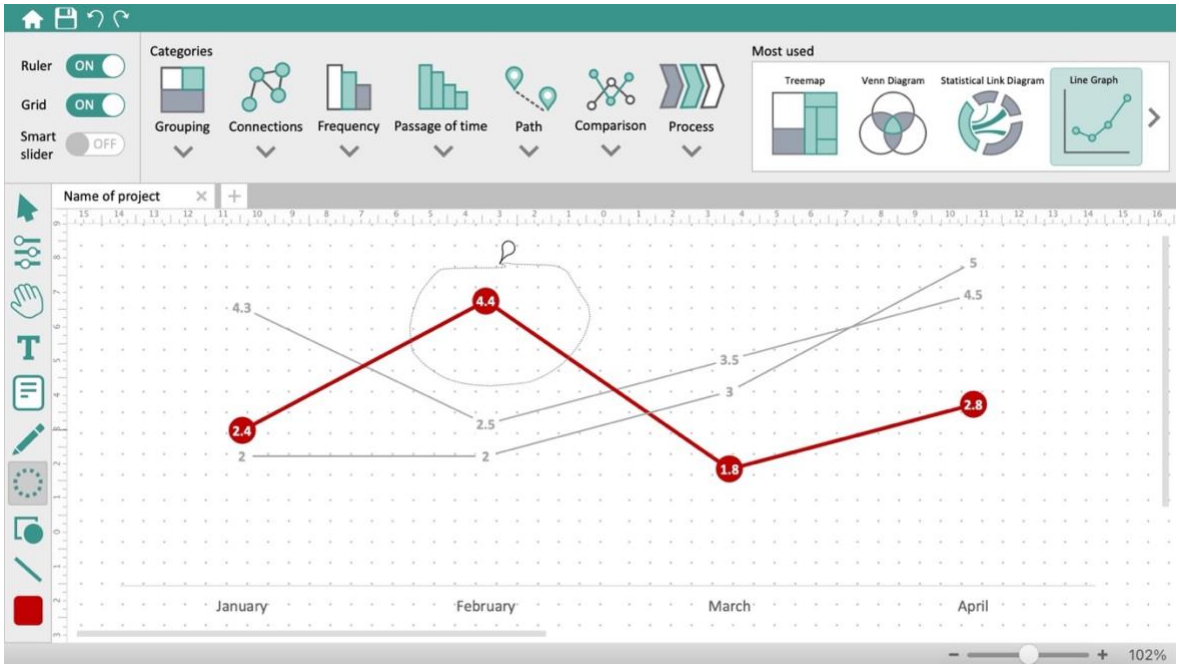


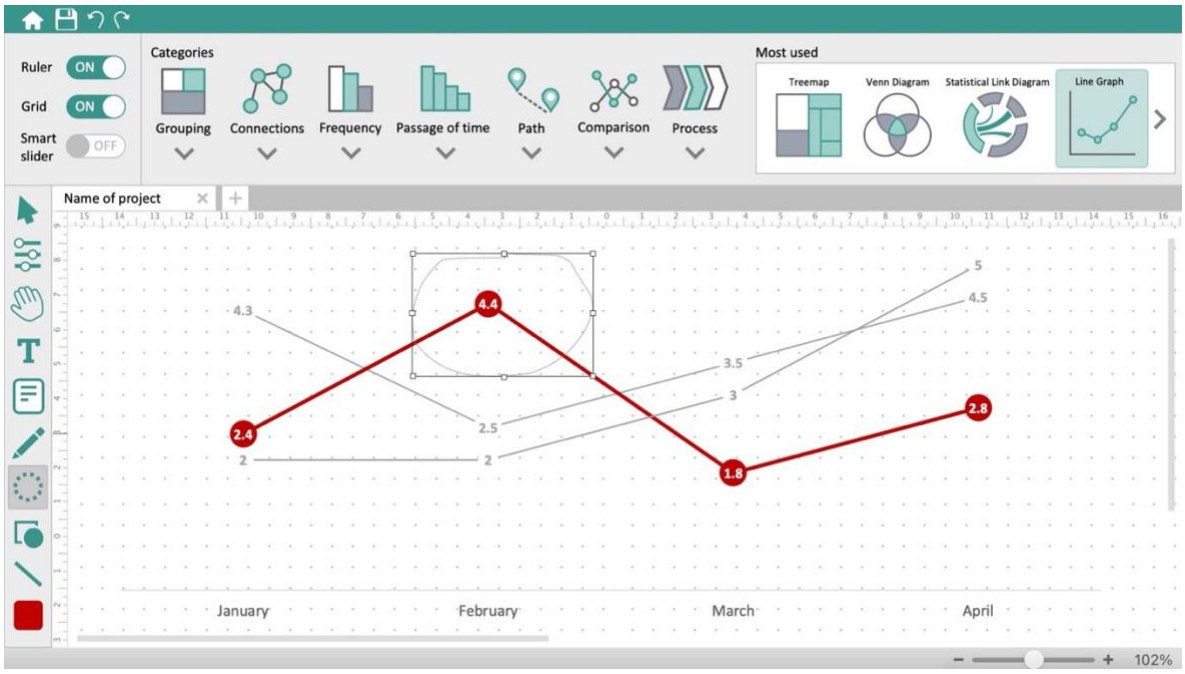


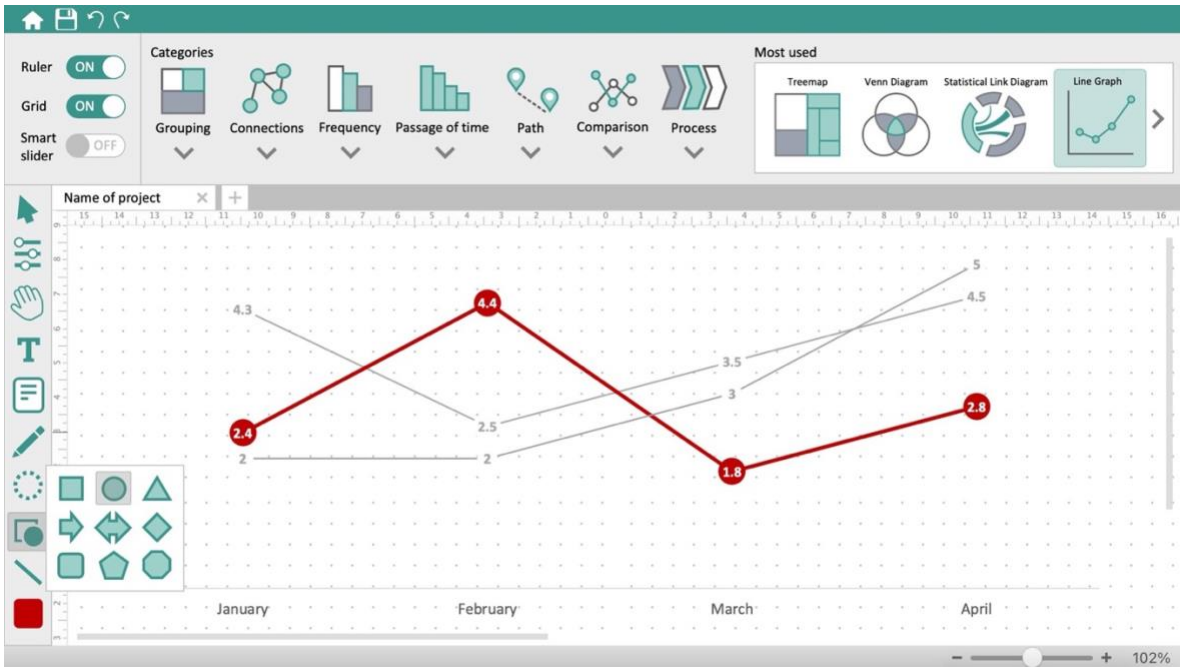

Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
AI assistant
[insert here explanation as to how the results of this datapoint were obtained]
Data obtained from:
Transcript 1
Transcript 2
Transcript 3
January
February
March
April
102%

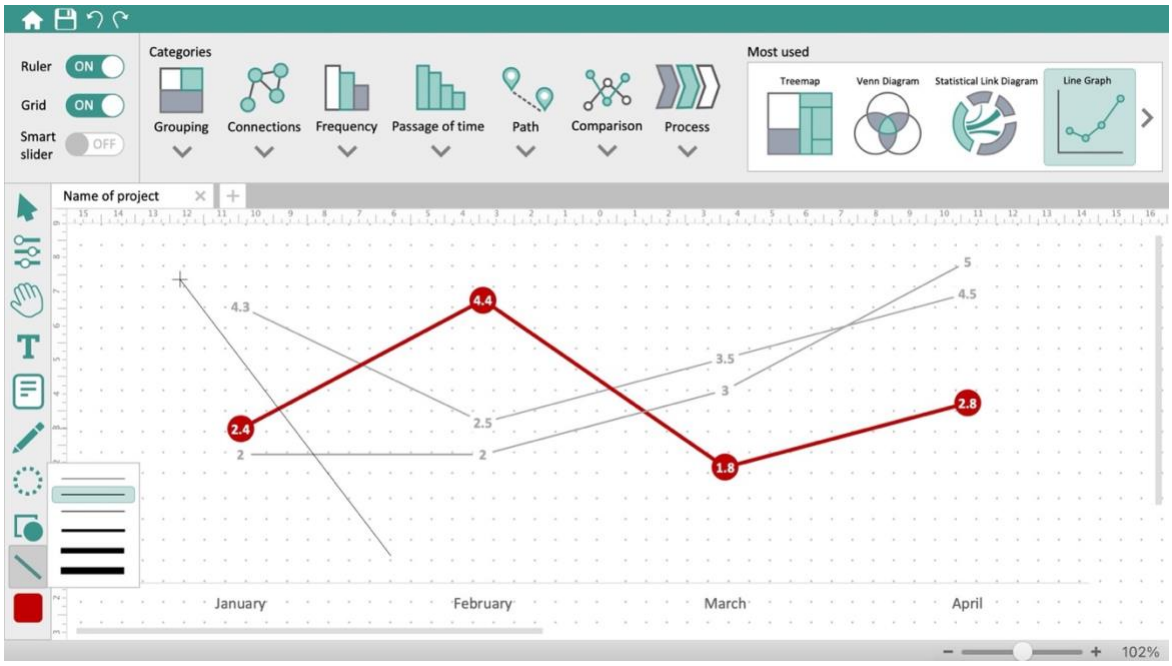

Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
[insert text here]
[insert text here]
January
February
March
April
102%

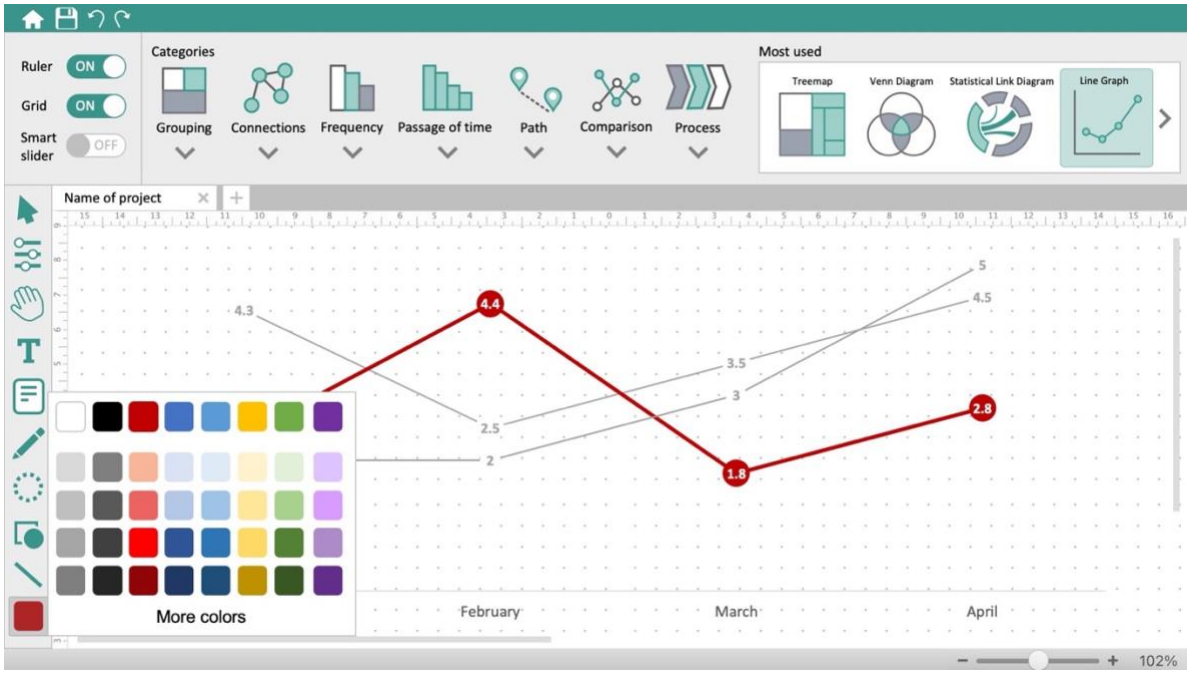

Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
4.3
4.4
2.5
3.5
3
5
4.5
2.4
2
2
1.8
2.8
January
February
March
April
102%

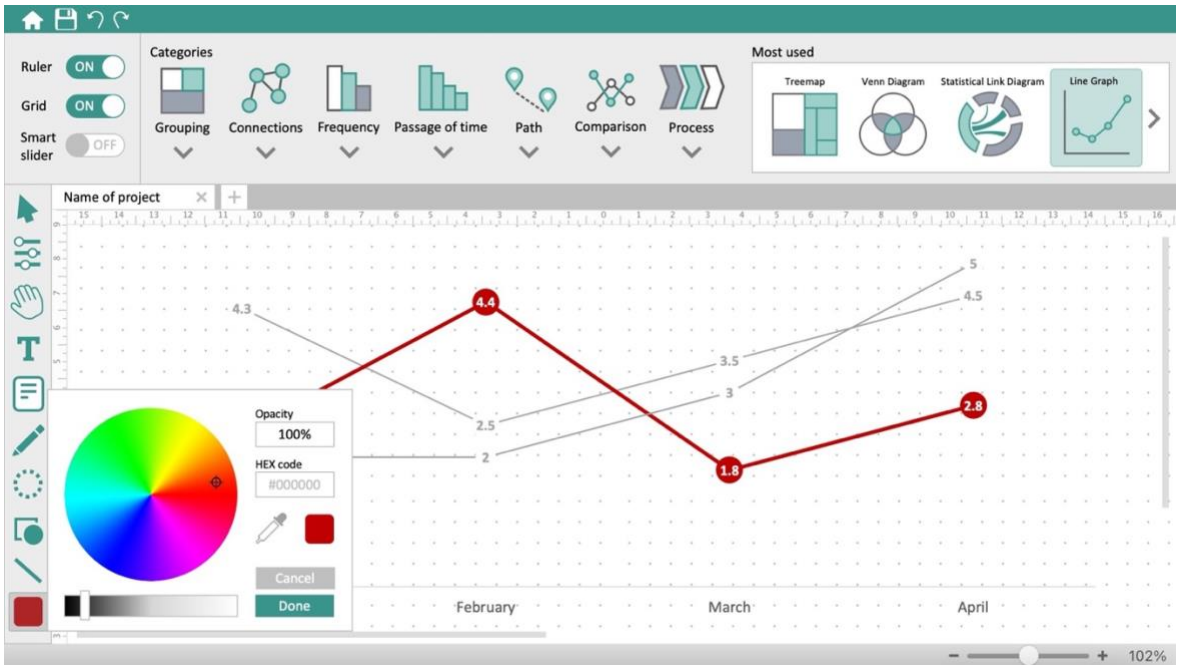

Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
4.3
4.4
2.5
3.5
3
5
4.5
2.4
2
2
1.8
2.8
January
February
March
April
102%

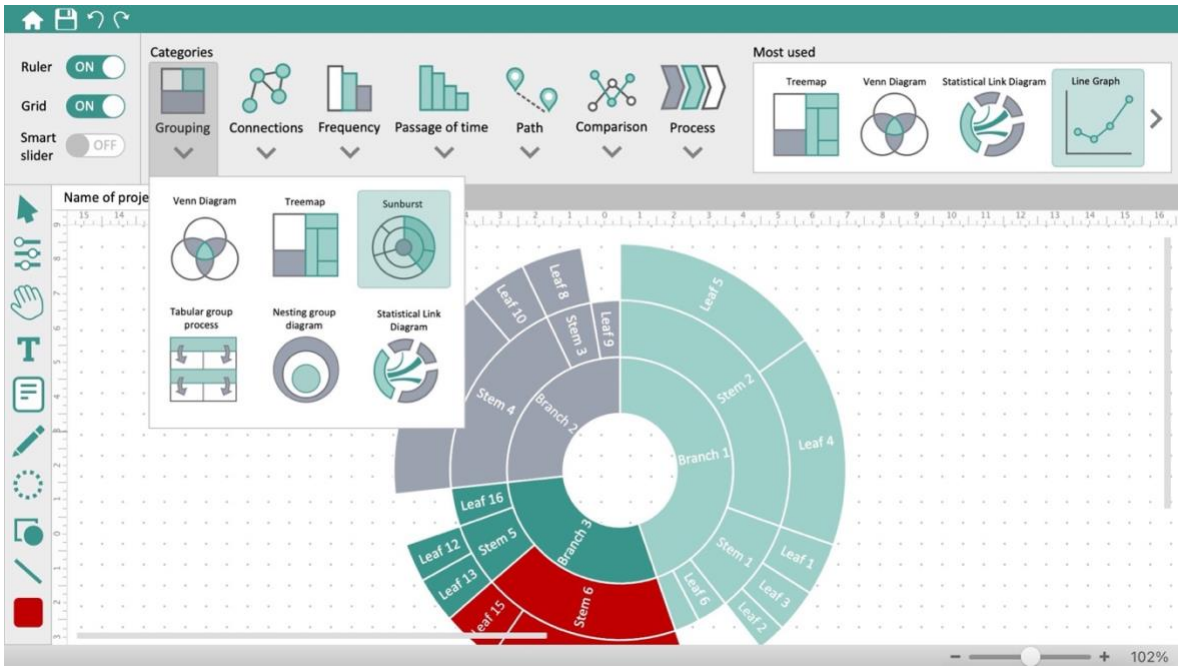

Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
4.3
4.4
2.5
3.5
3
5
4.5
2.4
2
2
1.8
2.8
January
February
March
April
102%

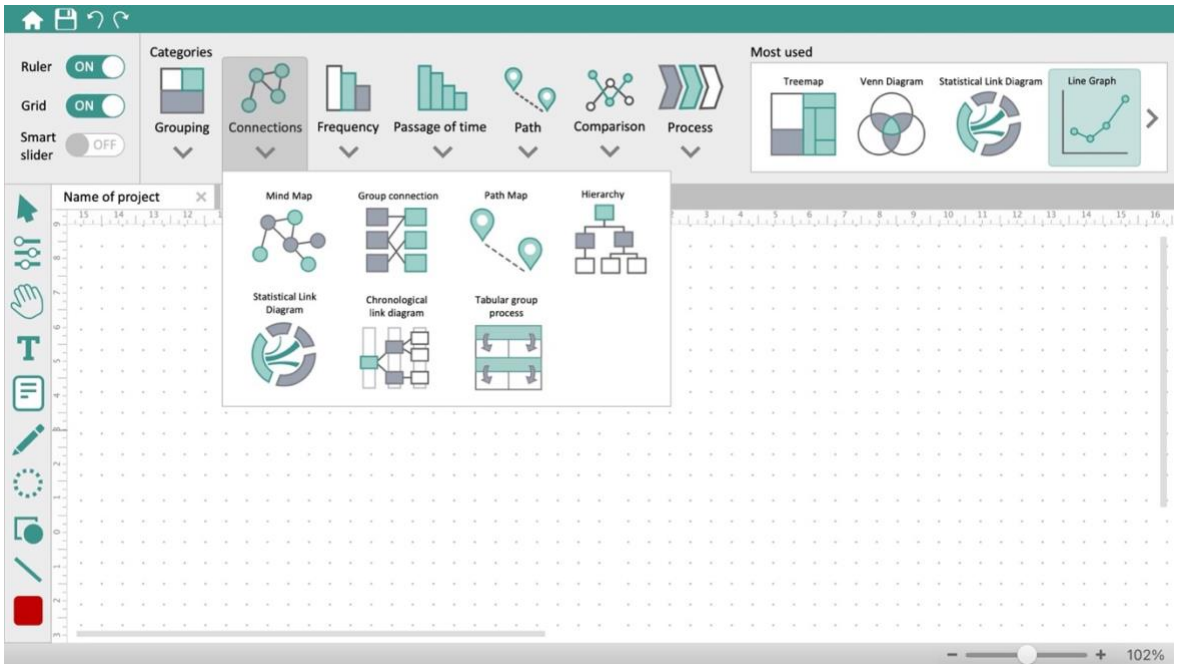

Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
4.3
4.4
2.5
3.5
3
5
4.5
2.4
2
2
1.8
2.8
January
February
March
April
102%

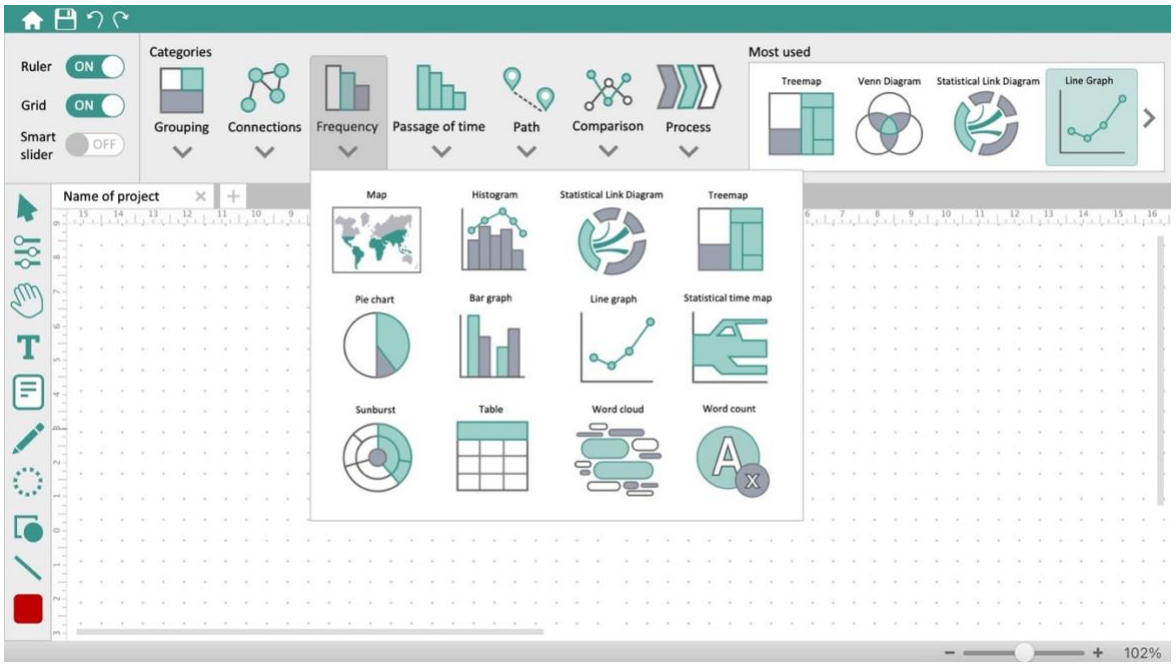

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
4.3
4.4
2.5
3.5
5
4.5
3
2.8
2.4
2
2
1.8
January
February
March
April
102%

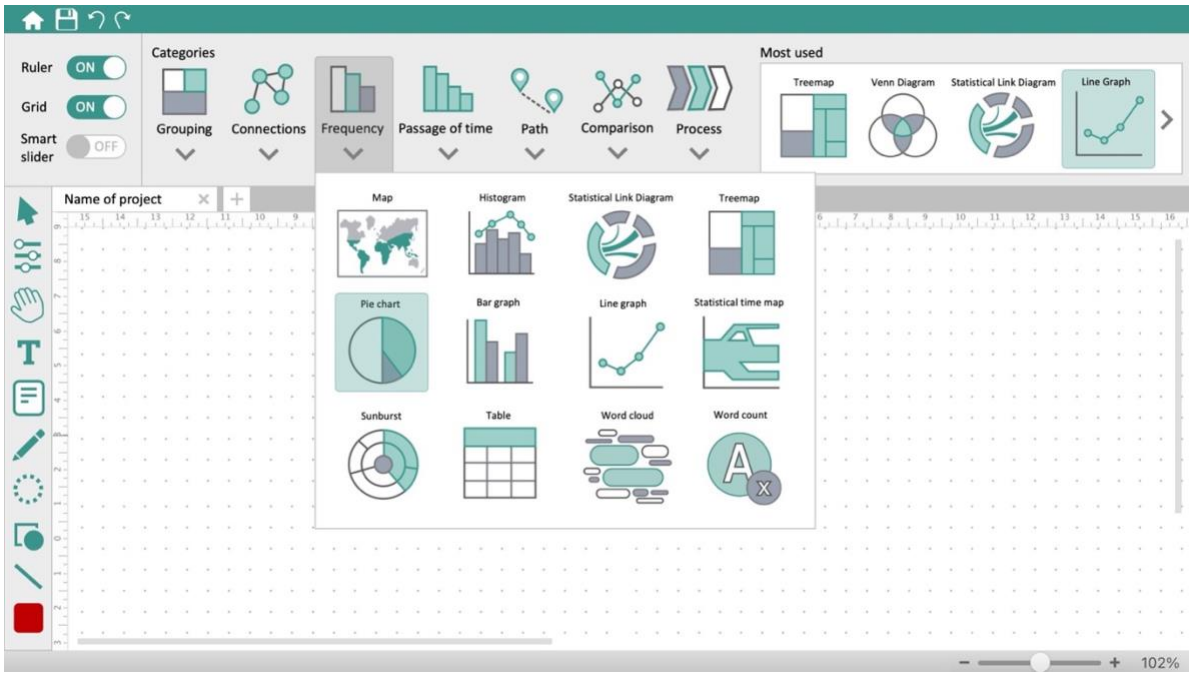

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
4.3
4.4
2.5
3.5
5
4.5
3
2.8
2
1.8
More colors
February
March
April
102%

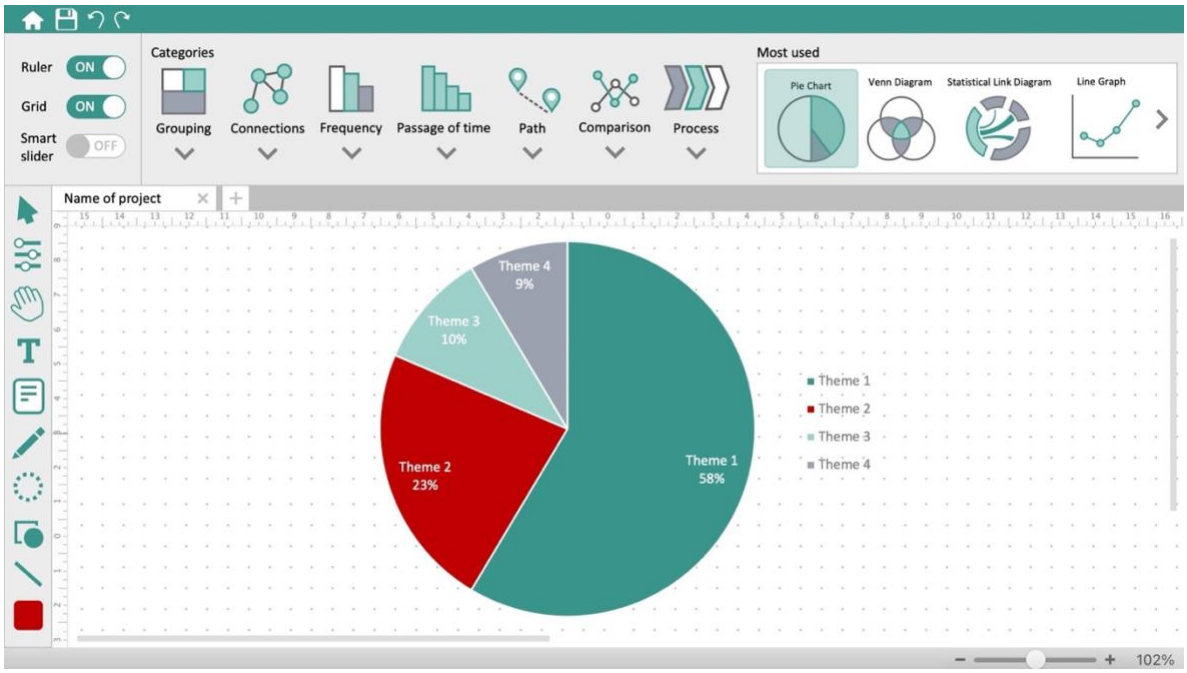

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Opacity
100%
HEX code
#000000
Cancel
Done
4.3
4.4
2.5
2
3.5
3
4.5
5
2.8
1.8
February
March
April
102%

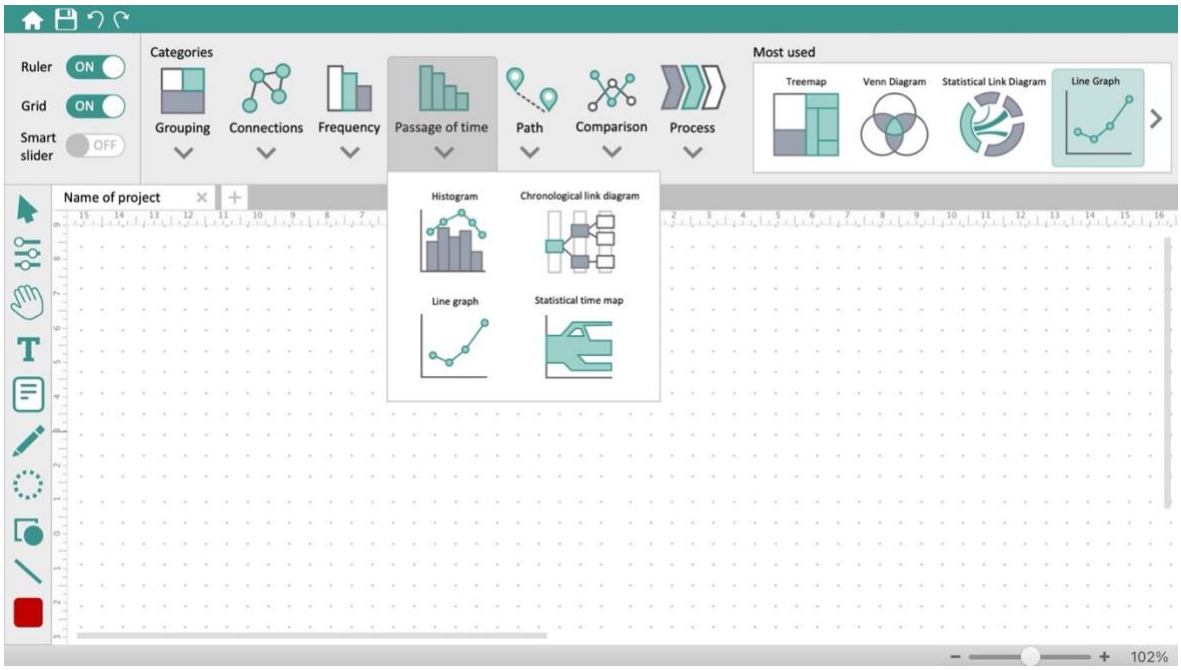

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Venn Diagram
Treemap
Sunburst
Tabular group process
Nesting group diagram
Statistical Link Diagram
Leaf 5
Leaf 8
Leaf 10
Stem 3
Leaf 9
Stem 2
Stem 4
Branch 2
Branch 1
Leaf 4
Leaf 16
Leaf 12
Stem 5
Leaf 13
Branch 3
Stem 1
Leaf 1
Leaf 6
Leaf 3
Leaf 2
Stem 6
102%

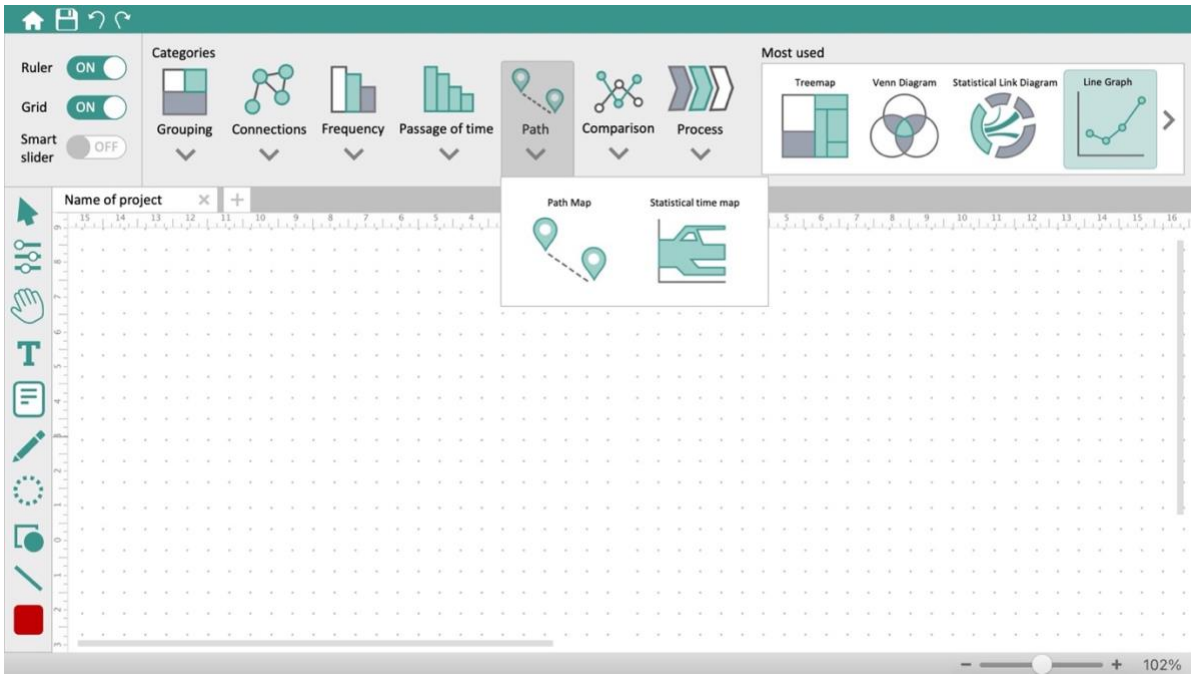

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Mind Map
Group connection
Path Map
Hierarchy
Statistical Link Diagram
Chronological link diagram
Tabular group process
102%

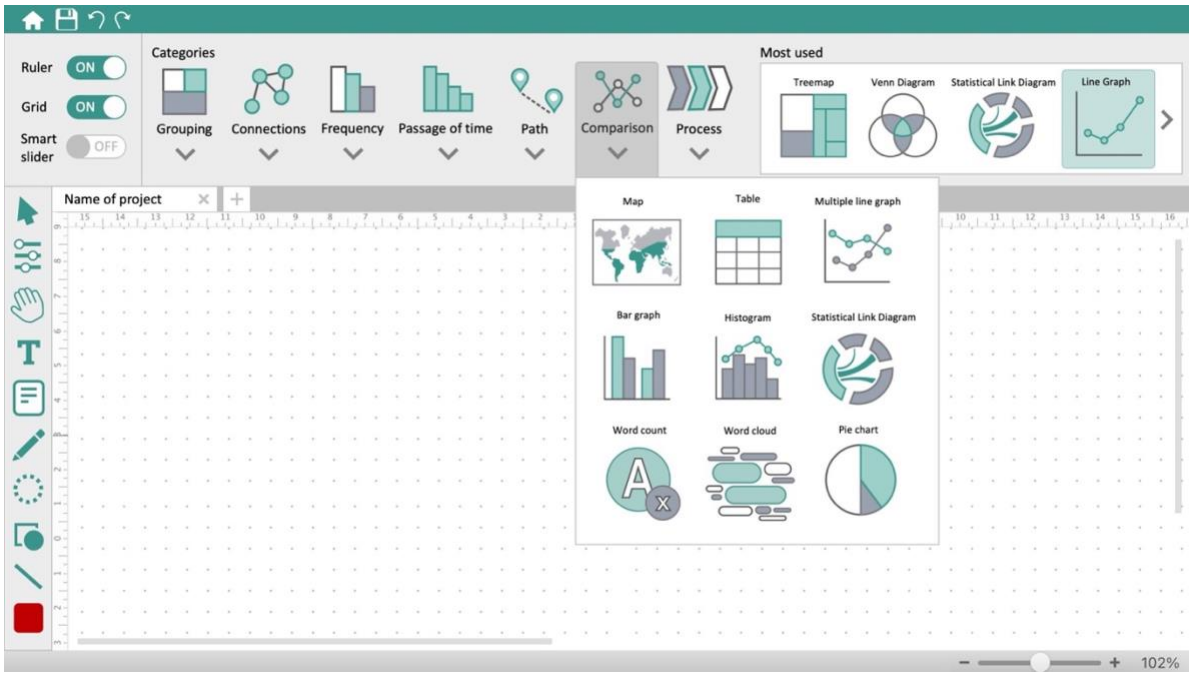

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Map
Histogram
Statistical Link Diagram
Treemap
Pie chart
Bar graph
Line graph
Statistical time map
Sunburst
Table
Word cloud
Word count
102%

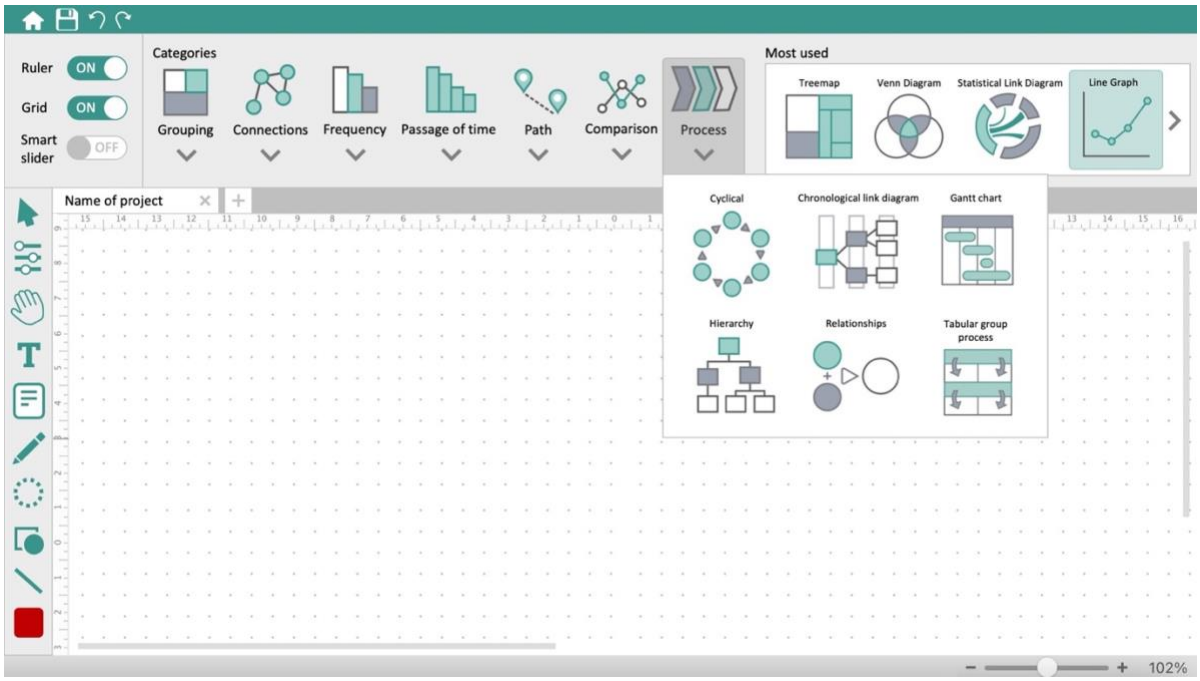

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Map
Histogram
Statistical Link Diagram
Treemap
Pie chart
Bar graph
Line graph
Statistical time map
Sunburst
Table
Word cloud
Word count
102%

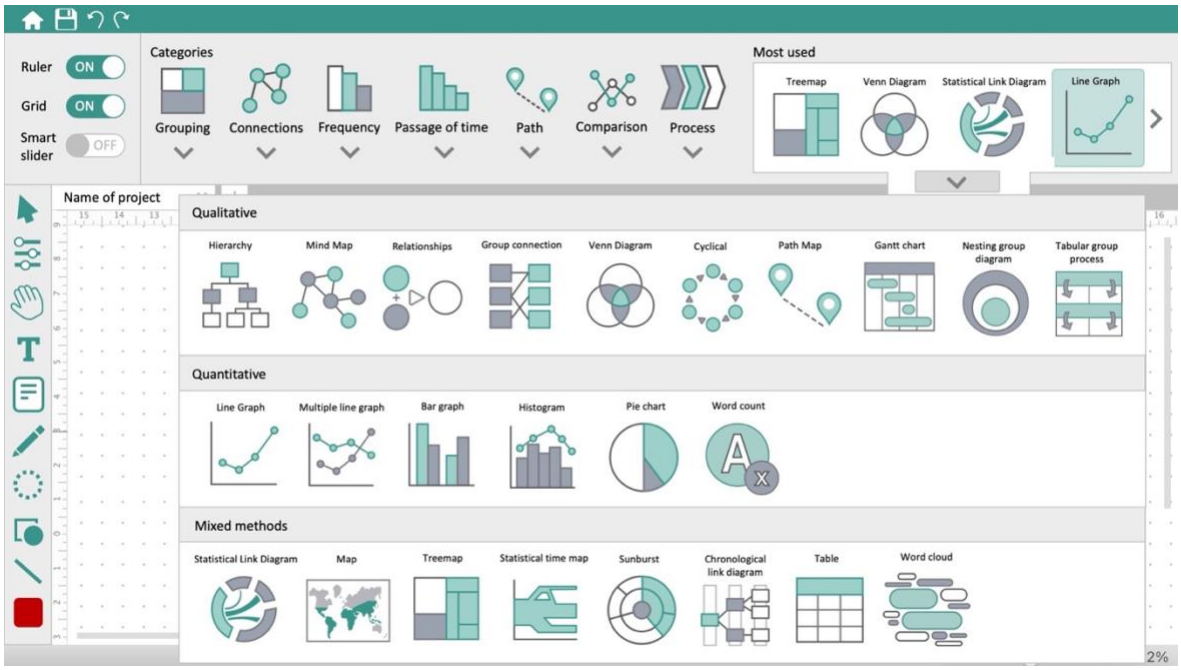

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Pie Chart
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Theme 4
9%
Theme 3
10%
Theme 2
23%
Theme 1
58%
Theme 1
Theme 2
Theme 3
Theme 4
102%

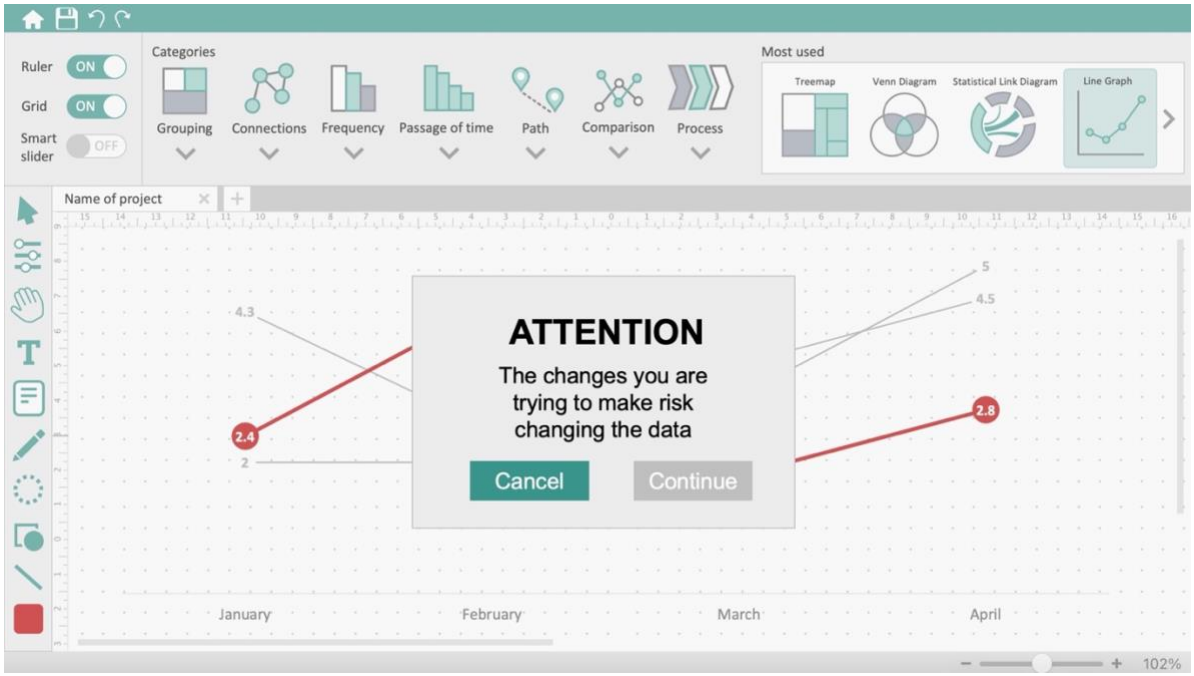

Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Histogram
Chronological link diagram
Line graph
Statistical time map
Name of project
102%


Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Path Map
Statistical time map
Name of project
102%


Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Map
Table
Multiple line graph
Bar graph
Histogram
Statistical Link Diagram
Word count
Word cloud
Pie chart
102%


Ruler
ON
Grid
ON
Smart slider
OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Cyclical
Chronological link diagram
Gantt chart
Hierarchy
Relationships
Tabular group process
102%


Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
Qualitative
Hierarchy
Mind Map
Relationships
Group connection
Venn Diagram
Cyclical
Path Map
Gantt chart
Nesting group diagram
Tabular group process
Quantitative
Line Graph
Multiple line graph
Bar graph
Histogram
Pie chart
Word count
Mixed methods
Statistical Link Diagram
Map
Treemap
Statistical time map
Sunburst
Chronological link diagram
Table
Word cloud
2%


Ruler ON
Grid ON
Smart slider OFF
Categories
Grouping
Connections
Frequency
Passage of time
Path
Comparison
Process
Most used
Treemap
Venn Diagram
Statistical Link Diagram
Line Graph
Name of project
4.3
2.4
2
5
4.5
2.8
ATTENTION
The changes you are trying to make risk changing the data
Cancel
Continue
January
February
March
April
102%

Appendix G

Ethics materials (consent letters)

Interview CONSENT FORM

By signing this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

Research Project Title: Exploring the use of data visualizations to make Artificial Intelligence more accessible for qualitative researchers conducting interviews in the field of healthcare.

Investigators: Luka Ugaya Mazza, University of Waterloo

James Wallace, University of Waterloo

Plinio Morita, University of Waterloo

I have read the information presented in the invitation letter about a study being conducted by Luka Ugaya Mazza of the School of Public Health Sciences, under the supervision of Professors James Wallace and Plinio Morita. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted.

I am aware that I may allow excerpts from the conversational data collected for this study to be included in teaching, scientific presentations and/or publications, with the understanding that any quotations will be anonymous.

I am aware that I do not need to share any images with the researcher but can do so if I desire to provide more visual context to the interview. I am aware that all images that might be shared with the research team must not have any confidential or identifiable data.

I am aware that any images I might share with the research team will not be published or made available for anyone outside of the research team.

I am aware that I may withdraw my consent for any of the above statements or withdraw my study participation without penalty by advising the researcher.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45268). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca.

Please Circle One

With full knowledge of all foregoing, I agree, of my own free will to participate in this study. YES NO

I agree to be audio recorded. YES NO

I agree to share any images that I think will be useful for the project with the researcher if I desire to.	YES	NO
I agree to **not** share any images with confidential or Identifiable data with the researcher.	YES	NO
I agree to let my conversation during the study be directly quoted, anonymously, in presentations of research results.	YES	NO
I agree to let the researchers use any sketches I might make and share on ZOOM Whiteboard.	YES	NO

Should you have any questions about the study, please contact us:

James Wallace	james.wallace@uwaterloo.ca
Plinio Morita	plinio.morita@uwaterloo.ca
Luka Ugaya Mazza	lugayama@uwaterloo.ca

_____

Participant Name (please print)

_____

Participant Signature

Date: _____

User testing CONSENT FORM (not used for thesis)

By signing this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

Research Project Title: Exploring the use of data visualizations to make Artificial Intelligence more accessible for qualitative researchers conducting interviews in the field of healthcare.

Investigators: Luka Ugaya Mazza, University of Waterloo

James Wallace, University of Waterloo

Plinio Morita, University of Waterloo

I have read the information presented in the invitation letter about a study being conducted by Luka Ugaya Mazza of the School of Public Health Sciences, under the supervision of Professors James Wallace and Plinio Morita. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted.

I am aware that I may allow excerpts from the conversational data collected for this study to be included in teaching, scientific presentations and/or publications, with the understanding that any quotations will be anonymous.

I am aware that I will need to share my screen during the interview and that my screen will be recorded, and that any images that come from it might be used in the future by the researchers. I am aware that those images and recordings will be completely anonymous and unidentifiable.

I am aware that I may withdraw my consent for any of the above statements or withdraw my study participation without penalty by advising the researcher.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45268). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca.

Please Circle One

With full knowledge of all foregoing, I agree, of my own free will to participate in this study. YES NO

I agree to be audio recorded. YES NO

I agree to let my conversation during the study be directly quoted, anonymously, in presentations of research results. YES NO

I agree to share my screen with the researcher and let my screen sharing be recorded. YES NO

I agree to let any images that come from screen sharing be used, anonymously, in presentations of research results. YES NO

Should you have any questions about the study, please contact us:

James Wallace james.wallace@uwaterloo.ca
Plinio Morita plinio.morita@uwaterloo.ca
Luka Ugaya Mazza lugayama@uwaterloo.ca

_____
Participant Name (please print)

_____
Participant Signature

Date: _____

Appendix H

Ethics materials (Information letters)

Interview

Information Letter – Invitation email

This email is being sent on behalf of the researchers:

Jim Wallace, University of Waterloo
Plinio Morita, University of Waterloo
Luka Ugaya Mazza, University of Waterloo

Subject: Exploring the use of data visualizations to make AI more accessible to qualitative researchers

Dear (potential participant),

This email is an invitation to participate in a research study we are conducting as part of my MSc degree at the School of Public Health Sciences (SPHS), University of Waterloo, under the supervision of Dr. Jim Wallace and Dr. Plinio Morita.

The goal of the study is to identify the needs and wants of qualitative researchers when it comes to using visual techniques to conduct data analysis. Our objective is to better understand how qualitative researchers might use visual language to make sense of their data in order to develop a data visualization tool that can in the future integrate AI with data analysis, without removing agency and control from the researcher.

We are seeking participants that have experience in conducting qualitative research and little to no experience with programming language.

The participants' role in this research study will be to provide insights on how they conduct their data analysis and participate in a co-design session with a designer in order to share what their vision for a tool that could assist in that process could look like. Participation will be voluntary and involve 2 (two) one-on-one 30 mins interviews conducted over the online platform, Zoom. Zoom has implemented technical, administrative, and physical safeguards to protect the information provided via the Services from loss, misuse, and unauthorized access, disclosure, alteration, or destruction. However, no Internet transmission is ever fully secure or error free.

You may decline to answer any of the interview questions if you wish. Further, you may decide to withdraw from this study at any time without any negative consequences by advising us, the researchers. Declining to participate in the study will not affect your relationship with the investigators or with the University of Waterloo. Only researchers associated with this study will have access to the records with your information which will be securely stored at the School of Public Health Sciences. We will keep our study records securely for a minimum of 7 years. We will remove all information that could identify you from the data within a couple hours of the interview and delete it permanently.

You can withdraw your consent to participate and have your data destroyed by contacting us. All records will then be destroyed according to University of Waterloo policies. Your data cannot be

withdrawn after it has been submitted for publication, but it could be potentially removed from any further use in other publications.

With your permission, the interview will be audio-recorded as a way to collect the information, and later transcribed for analysis. Your identity will be kept confidential. Your name will not appear in any thesis or report resulting from this study, however, with your permission, anonymous quotations from you may be used. You may share images to help illustrate how your data analysis process happens, however this is not required. The images shared must NOT contain any confidential or identifiable information. Additionally, a confidential ZOOM Whiteboard will be shared with you so you can sketch any ideas for the prototype you should desire. These sketches will only be saved as a screenshot and have no connection to your identity. However, with your permission, the anonymous sketches collected may be used for future publications.

There are no direct benefits to you the participants and no anticipated risks from participating in the study. Now more than ever healthcare needs to hear the experiences and voices of the people that are most affected by its need. Qualitative research is very important for understanding these voices and through our research we hope to develop a tool that will make conducting it easier on the researchers. Out hope is to provide a tool that will assist with qualitative research without making the researcher feel like they are being replaced by the AI, but instead work towards building a partnership between researcher and tool.

If you are willing to participate, please suggest a day and time that suits you and we will do our best to be available. Should you have any questions about the study, please email Luka Ugaya Mazza at lugayama@uwaterloo.ca.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45268). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca.

We look forward to speaking with you very much and thank you in advance for your interest in this project.

Thank you,

Jim Wallace
james.wallace@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Plinio Morita
plinio.morita@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Luka Ugaya Mazza
lugayama@uwaterloo.ca
Student Investigator
School of Public Health Sciences
University of Waterloo, Canada

User testing (not used for the thesis)

Information Letter – Invitation email

This email is being sent on behalf of the researchers:

Jim Wallace, University of Waterloo
Plinio Morita, University of Waterloo
Luka Ugaya Mazza, University of Waterloo

Subject: Exploring the use of data visualizations to make AI more accessible to qualitative researchers

Dear (potential participant),

This email is an invitation to participate in a research study we are conducting as part of my MSc degree at the School of Public Health Sciences (SPHS), University of Waterloo, under the supervision of Dr. Jim Wallace and Dr. Plinio Morita.

The goal of the study is to identify the needs and wants of qualitative researchers when it comes to using visual techniques to conduct data analysis. Our objective is to better understand how qualitative researchers might use visual language to make sense of their data in order to develop a data visualization tool that can in the future integrate AI with data analysis, without removing agency and control from the researcher.

We are seeking participants that have experience in conducting qualitative research and little to no experience with programming language.

The participants' role in this research study will be to test a data visualization prototype tool that assists in knowledge translation for qualitative research assisted by AI.

Participation will be voluntary and involve a one-on-one 30 mins interview conducted over the online platform, Zoom. Zoom has implemented technical, administrative, and physical safeguards to protect the information provided via the Services from loss, misuse, and unauthorized access, disclosure, alteration, or destruction. However, no Internet transmission is ever fully secure or error free.

You may decline to answer any of the interview questions if you wish. Further, you may decide to withdraw from this study at any time without any negative consequences by advising us, the researchers. Declining to participate in the study will not affect your relationship with the investigators or with the University of Waterloo. Only researchers associated with this study will have access to the records with your information which will be securely stored at the School of Public Health Sciences. We will keep our study records securely for a minimum of 7 years. We will remove all information that could identify you from the data within a couple hours of the interview and delete it permanently.

You can withdraw your consent to participate and have your data destroyed by contacting us. All records will then be destroyed according to University of Waterloo policies. Your data cannot be withdrawn after it has been submitted for publication, but it could be potentially removed from any further use in other publications.

With your permission, the interview will be audio-recorded as a way to collect the information, and later transcribed for analysis. Your identity will be kept confidential. Your name will not appear in any thesis or report resulting from this study, however, with your permission, anonymous quotations from you may be used. Additionally, a confidential ZOOM Whiteboard will be shared with you so you can sketch any thoughts you have about the prototype. These sketches will only be saved as a

screenshot and have no connection to your identity. However, with your permission, the anonymous sketches collected may be used for future publications.

During the interview you will be given a prototype of a data visualization in order to test it. With your permission, you will share your screen with the researcher and allow us to record your process of moving through and interacting with the prototype. Your identity will be kept confidential. Your name, face, or any other identifiable elements will not appear in any thesis or report resulting from this study, however, with your permission, anonymized screenshots of you using the prototype will be used for future publications.

There are no direct benefits to you the participants and no anticipated risks from participating in the study. Now more than ever healthcare needs to hear the experiences and voices of the people that are most affected by its need. Qualitative research is very important for understanding these voices and through our research we hope to develop a tool that will make conducting it easier on the researchers. Out hope is to provide a tool that will assist with qualitative research without making the researcher feel like they are being replaced by the AI, but instead work towards building a partnership between researcher and tool.

If you are willing to participate, please suggest a day and time that suits you and we will do our best to be available. Should you have any questions about the study, please email Luka Ugaya Mazza at lugayama@uwaterloo.ca.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45268). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca.

We look forward to speaking with you very much and thank you in advance for your interest in this project.

Thank you,

Jim Wallace
james.wallace@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Plinio Morita
plinio.morita@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Luka Ugaya Mazza
lugayama@uwaterloo.ca
Student Investigator
School of Public Health Sciences
University of Waterloo, Canada

Appendix I

Ethics materials (feedback letters)

Co-design

Dear ***(Insert Name of Participant),***

We would like to thank you for your participation in this study. As a reminder, the purpose of this study is to identify the needs and wants of qualitative researchers when it comes to using visual techniques to conduct data analysis.

The results from the study will help better understand how qualitative researchers might use visual language to make sense of their data in order to develop a data visualization tool that can in the future integrate AI with data analysis, without removing agency and control from the researcher.

Please remember that your identity will be kept confidential. The data collected will not include any personal identifiers. If you are interested in receiving more information regarding the results of this study, or if you have any questions or concerns, please contact us at either the phone number or email address listed at the bottom of the page.

Please keep in mind that the research team will contact you shortly to schedule the second 30min interview. Should you choose to contact the researcher before to schedule it sooner, please email Luka Ugaya Mazza at lugayama@uwaterloo.ca.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45268). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca. If you have any questions pertaining to the study specifically, please contact the researchers.

Thank you,

Jim Wallace
james.wallace@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Plinio Morita
plinio.morita@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Luka Ugaya Mazza
lugayama@uwaterloo.ca
Student Investigator
School of Public Health Sciences
University of Waterloo, Canad

User testing (not used for the thesis)

Dear (Insert Name of Participant),

We would like to thank you for your participation in this study. As a reminder, the purpose of this study is to identify the needs and wants of qualitative researchers when it comes to using visual techniques to conduct data analysis.

The results from the study will help better understand how qualitative researchers might use visual language to make sense of their data in order to develop a data visualization tool that can in the future integrate AI with data analysis, without removing agency and control from the researcher.

Please remember that your identity will be kept confidential. The data collected will not include any personal identifiers. If you are interested in receiving more information regarding the results of this study, or if you have any questions or concerns, please contact us at either the phone number or email address listed at the bottom of the page.

This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45268). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca. If you have any questions pertaining to the study specifically, please contact the researchers.

Thank you,

Jim Wallace
james.wallace@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Plinio Morita
plinio.morita@uwaterloo.ca
Associate Professor
School of Public Health Sciences
University of Waterloo, Canada

Luka Ugaya Mazza
lugayama@uwaterloo.ca
Student Investigator
School of Public Health Sciences
University of Waterloo, Canada

Appendix J

Interview protocol for user testing (not used in the thesis)

Interview guide phase 3: Measuring the efficacy of using data visualizations for researcher agency.

Part 1: Think-A-Loud

Prompt the participant to explore the data visualization template as they narrate their process and feeling while doing it.

Part 2: survey

Thank you for your participations, I will now ask you a couple questions and if you could answer them on a scale of 1 to 7, **1 being that you strongly disagree** and **7 being strongly agree.**

	Agree						Disagree
	1	2	3	4	5	6	7
1. I was always in full control while doing the task							
2. I felt like I was just an instrument at the hands of the AI							
3. My movements were automatic, my body just made them, and I didn't have to think							
4. The outcome of my actions generally surprised me							
5. The decisions of whether or when to act was withing my hands							
6. While I was performing an action, I felt like the AI was controlling the research more than me							
7. I felt completely responsible for the results of my actions							
8. I felt free to move always							
9. I always understood how to use the tool							
10. Using the data visualization tool felt very intuitive							

Appendix K

Ethics Protocol

PROTOCOLS

UNIVERSITY OF WATERLOO

#45268 - Exploring the use of data visualizations to make Artificial Intelligence more accessible for qualitative researchers conducting interviews in the field of healthcare.

Protocol Information

Review Type	Status	Approval Date	Renewal Date
Expedited	**Approved**	**Apr 28, 2023**	**Apr 13, 2024**

Expiration Date	Initial Approval Date	Initial Review Type
Apr 29, 2024	**Apr 28, 2023**	**Expedited**

Feedback

Approval Comment

This study has ethics clearance. Please direct any questions or concerns to Heather Dekker (hdekker@uwaterloo.ca), ext. 41506

General Information

Only the Principal Investigator/Faculty Supervisor can submit the application. This acts as a signature indicating approval of the application.

Principal Investigator / Faculty Supervisor

James Wallace

Department

School of Public Health Sciences

Study title

Exploring the use of data visualizations to make Artificial Intelligence more accessible for qualitative researchers conducting interviews in the field of healthcare.

General Questionnaire

Indicate the type of application you would like to complete

Standard application *

*** The Standard application is for faculty level research and thesis level research.**

**** The course project application is for single-term (non-thesis) course based research and can be completed by students or the course instructor**

Please confirm:

I understand that the type of applications listed above determine the form I am about to complete. If I have chosen the incorrect form I acknowledge that I may need to complete a new application.

People

University of Waterloo research team

Ensure all information in this table is completed.

Person

James Wallace

Waterloo Department

School of Public Health Sciences

Email Address

jrwallac@uwaterloo.ca

Phone

Researcher Role

Principal Investigator

Permissions

Full Access

Mandatory Training

All University of Waterloo undergraduate and graduate students, faculty, and staff must complete the TCPS 2 CORE tutorial prior to submitting an application for review. See instructions or email researchethics@uwaterloo.ca if you have questions.

REQUIRED: Upload a copy of the TCPS 2 certificate or a screen shot showing module completion. Applications where the certificate/screen shot is not uploaded will be sent back to the researcher and not reviewed.

TCPS2_CORE_CERTIFICATE_JIM WALLACE.PDF

As per the Waterloo policy on mandatory research ethics training, if you completed the TCPS2 tutorial more than 5 years ago, you may be asked to update your training within the next 6 months. You will be notified by email if this is the case.

Ensure all information in this table is completed.

Person

Luka Ugaya Mazza

Waterloo Department

School of Public Health Sciences

Email Address

lugayama@uwaterloo.ca

Phone

Researcher Role

Student investigator

Permissions

Full Access

Mandatory Training

All University of Waterloo undergraduate and graduate students, faculty, and staff must complete the TCPS 2 CORE tutorial prior to submitting an application for review. See instructions or email researchethics@uwaterloo.ca if you have questions.

REQUIRED: Upload a copy of the TCPS 2 certificate or a screen shot showing module completion. Applications where the certificate/screen shot is not uploaded will be sent back to the researcher and not reviewed.

TCPS2_CORE_CERTIFICATE.PDF

As per the Waterloo policy on mandatory research ethics training, if you completed the TCPS2 tutorial more than 5 years ago, you may be asked to update your training within the next 6 months. You will be notified by email if this is the case.

Ensure all information in this table is completed.

Person

Plinio Morita

Waterloo Department

School of Public Health Sciences

Email Address

pmorita@uwaterloo.ca

Phone

Researcher Role

Co-Principal Investigator

Permissions

Full Access

Mandatory Training

All University of Waterloo undergraduate and graduate students, faculty, and staff must complete the TCPS 2 CORE tutorial prior to submitting an application for review. See instructions or email researchethics@uwaterloo.ca if you have questions.

REQUIRED: Upload a copy of the TCPS 2 certificate or a screen shot showing module completion. Applications where the certificate/screen shot is not uploaded will be sent back to the researcher and not reviewed.

PLINIO MORITA - TCPS2_CORE_CERTIFICATE.PDF

Do you have any investigators external to the University of Waterloo?

No

As per the Waterloo policy on mandatory research ethics training, if you completed the TCPS2 tutorial more than 5 years ago, you may be asked to update your training within the next 6 months. You will be notified by email if this is the case.

General details

Is this new study related to any previous application?

No

What is the estimated start and end date for the study?

Start Date (Date you anticipate beginning the recruitment of participants)

April 3, 2023

End Date (Date you expect the study will end, for example, the date when there will be no further contact with participants or when the data analysis will be completed for US funded research)

April 3, 2024

Does this research require approval from a UWaterloo departmental committee?

Not a department requirement

What is the level of the research to be conducted? Choose one.

Master's thesis

Will this study involve Wilfrid Laurier University, Western University, Conestoga College or Local hospitals covered by the Tri-Hospital Research Ethics Board (Cambridge Memorial Hospital, Grand River Hospital and St. Mary's General Hospital)?

No

Has a version of this study been disapproved or rejected by any Research Ethics Board/Committee?

No

Special Instructions RE: Research During COVID-19 Pandemic
Are you proposing in this application a study that involves in-person (face-to-face) research activities either on-campus or off-campus?

No

Study description

State your research question(s)

How do healthcare qualitative researchers use visual techniques to make sense of their data? What visual elements do they find important to have when conducting research? What kinds of data visualization satisfy which needs of the qualitative healthcare research community?

Provide a clear, detailed description of the purpose, hypothesis, aim, and objectives of this study

The purpose of this study is to identify the needs and wants of qualitative researchers in the field of healthcare when it comes to the analysis of data collected from interviews. This will be done in order to develop a data visualization tool to make using AI more accessible to this demographic. Crucial to this is speaking to qualitative researchers that have experience in the field of healthcare to learn of their experiences and requirements, as well as testing the future prototype with users to ensure its efficacy. Our hypothesis is that through the use of data visualizations we can make AI more accessible to qualitative researchers that have interest in using the tool but no familiarity with programming language. Our goal is to better understand the needs of qualitative researchers in the field of healthcare to create a tool that can make AI more accessible for them.

Provide background information, a rationale, and justification for conducting this study. Describe why the research is being done and what research has already been done in this area. Be sure to explain why this research is important.

Healthcare is a field that can benefit from the collection of data about lived-in experiences of patients and healthcare practitioners. One way that researchers can gleam that kind of insight is by using qualitative methods of data collection and analysis. This kind of method is better suited for the analysis of rich and interconnected data, especially when dealing with the lived in experiences of people. However, conducting qualitative research is a strenuous and exhausting process. To assist with this, tools like Artificial Intelligence are being created to lessen the load on researchers. However, AI often times is inaccessible for those that are not familiar with programming language. This disconnect is a significant hurdle in the incorporation of AI in qualitative data analysis. To assist with this, we will be exploring the use of data visualizations to help translate information from Machine Learning based AI to qualitative researchers with little to no experience with programming language. A lot of the research that focuses on producing data visualizations for AI tools fails to take into consideration the specific needs qualitative researchers have that differ from quantitative researcher, including the way they feel to be used. For qualitative researchers, the process is as important as the conclusions and whatever

tool is introduced needs to take into consideration that process and how researchers move through it. With this project we hope to compile those needs and translate them into a data visualization. This will be done to help develop a translation tool for qualitative researcher that combines AI with the unique needs of this type of research.

In a maximum of 250 words, provide a non-scientific lay language description that summarizes the project outlining the purpose, anticipated benefits, and basic procedures. Write this summary as if it would be read by members of the general public who are not familiar with academic terms or acronyms. Use language suitable for a media release.

When it comes to healthcare, it is important to listen to the individual experiences of people and practitioners when doing research. The best way to do this is to use qualitative methods, which means focusing on the quality of the information and not just the quantity. This kind of research takes time and effort, which is not always available. A possible solution for this is to use Artificial Intelligence as a tool to lessen the researcher's load. However, AI uses a specific language that people not used to it often find difficult to understand. To bridge that gap, we propose using data visualizations to translate that information and make AI more accessible. Most research on data visualization for research has focused on the efficacy of using them to accomplish goals. While this is a valid way to assess their performance, it doesn't consider how a visualization may feel to be used. For qualitative researchers as important as the conclusions is the path taken to reach them. This can conflict with tools like AI, where the path to the answers can be obscured if one does not know programming language. Our team will explore the use of data visualizations as translation tools to make AI more accessible for qualitative researchers, focusing on how the data visualization feels to be used. Our team will seek to answer how researchers want to see their data and how we can best design data visualizations to increase their sense of agency.

What is the study design?

2 semi-structured interviews (conducted online) and 1 unstructured interview with co-design (conducted online)

Is this a pilot study?

No

Sample Size

What is the expected sample size? Outline the number of participants anticipated to take part in the study.

10 people, divided into 2 groups, each with 5 participants (i.e., 5 participants per group).

Was a formal sample size calculation completed?

No

Provide a rationale for the number of participants specified

This will be a highly specific study with a specific demographic. The focus will be on qualitatively analyzing the needs of a small population (qualitative researcher in healthcare) while doing a specific task (data analysis for interviews using thematic analysis), therefore the sample can be smaller.

Study sites

Where is this study taking place?

Remote (online survey, virtual/telephone interview, etc.)

Please note that different guidelines/policies may apply when participants are recruited from certain locations.

Funding

Is the study funded/will it be funded?

No

Conflict of interest

Are there any potential, perceived, or actual financial or non-financial conflicts of interest of the research team in undertaking the proposed research?

No

Benefits

Are there direct benefits of the proposed research to the study participants?

No

What are the scientific and/or scholarly benefits of the proposed research?

The development of a visualization tool that can assist in the accessibility of AI to qualitative researchers in the healthcare field.

Participants

Participant general categories

University of Waterloo undergraduate and/or graduate students
University of Waterloo staff and/or faculty
Adults (age 18-64 years)

Are you conducting research in classes with students as your participants to evaluate a teaching method or object?

No

Describe the sample in detail and list any specific inclusion/exclusion criteria for the study

This project will have 3 phases of participant involvement. All of them will have the overall criteria of: Qualitative researchers in the field of healthcare that have no, or limited experience, with programming language, and have experience conducting interviews and analyzing qualitative data. Phases 1

and 2 will be with the same 5 participants. This phase will be when the co-design of the visualization tool happens. Phase 3 will be with a new set of 5 participants that cannot have participated previously. This is in order to test the prototype with participants that did not help design it, in order to lower the bias of their answers.

If you are excluding people on certain characteristics provide a justification for the exclusion.

Only participants that speak English, as a translator was not considered for this project. If the participant is well versed with using AI to conduct their data analysis they will also be excluded, as they do not fall under the criteria of having difficulties understanding its language and thus will benefit less from out proposed project.

Will a screening process be used to determine eligibility in the study based on the inclusion and/or exclusion criteria identified above?

No

Recruitment

Identify from where/what sources potential participants will be recruited.

Through email/internet (e.g., social media networks)

Indicate what email listing, internet site or network you intend to recruit from

University of Waterloo Public Health Graduate students mail list.

What recruitment materials will be used? See sample recruitment materials.

Email script

Upload your recruitment materials

Upload your recruitment materials

INFORMATION LETTER_PHASES 1 AND 2_V2_20230419.PDF

Study group

Group 1, which will participate in phases 1 and 2.

Upload your recruitment materials

INFORMATION LETTER_PHASE 3_V2_20230419.PDF

Study group

Group 2, which will participate in phase 3

Will potential participants be recruited through pre-existing relationships with members of the research team (e.g., employees, students, or patients of research team, acquaintances, own children or family members, colleagues, etc.)?

Yes

Outline the relationship between the researchers and potential participants (e.g., professor-student, colleagues)

Colleagues (some participants may be colleagues of the investigator since we will be conducting research inside of our own faculty) Professor-student (some participants may be professors on the investigators since we will be conducting research inside of our faculty and some of them might have taught courses that the investigator took in the past)

Could this relationship compromise the potential participant's freedom to decline participation?

No

Explain

The investigators will contact the potential participants from the University of Waterloo that fit the criteria as outlined in the previous sections. These colleagues and professors will be contacted due to their expertise in the field of qualitative research and interest in the use of AI for their data analysis and will have the freedom to decline with no personal or professional consequences. The participants will also have the choice to quit at any moment with no repercussions or judgment from the investigators.

Methods and procedures

Which of the following will be conducted for this study?

One-on-one interviews

How will the one-on-one interviews be conducted?

Online – Video Chat

Will quotations be used in the write-up of the study

Yes

What type of quotations will be used?

Anonymous

For each of the procedures indicated above, provided a detailed, sequential description of how they will be used in the study. Provide one or more paragraphs describing how you are recruiting participants, obtaining consent, what participants are asked to do, and how the research team will be using the collected data.

A list of potential partners will be made based on the qualitative healthcare researchers located in the University of Waterloo. The recruitment will happen through email lists and personal contact between the student investigator and the potential participants. Once those connections are formed the study will happen in three phases. Phase 1: semi-structured interview with qualitative healthcare researchers with little to no experience with programming language. 1. An email will be sent to potential participants inviting them to participate; 2. An interview will be scheduled through Zoom at a time that suits both participant and investigator; 3. The interview will follow a semi-structured script. In case the participants wants to share images to help describe how they want to visualize their data, they can do so. However, this is not a requirement; 4. An email will be sent to the

participants with a feedback letter thanking the participants for their time and insights, as well as an invitation to schedule the second interview. The results will be analyzed and used to develop a template of data visualizations to assist qualitative researchers. Phase 2: co-design with participants 1. The participants from the previous phase will be contacted to confirm the second interview; 2. An interview will be scheduled through Zoom at a time that suits both participant and investigator; 3. The interview will be unstructured and consist of a 30mins to an 1 hour session of co-designing a data visualization to fit their specific needs; 4. A feedback letter will be sent thanking the participants for their time and insights. The result will be a prototype that will be tested in phase 3. Phase 3: testing to prototype with qualitative healthcare researchers 1. An email will be sent to potential participants inviting them to participate; 2. An interview will be scheduled through Zoom at a time that suits both participant and investigator; 3. The interview will be semi-structured with Think-A-Loud 4. A feedback letter will be sent thanking the participants for their time and insights. The results will be analyzed in order to further perfect the data visualization tool.

Please upload any study materials related to the procedure(s)

Study material

INTERVIEW1QS_V2_20230419.PDF

Study material

INTERVIEW2QS_V1_20230419.PDF

Study material

INTERVIEW3QS_V1_20230316.PDF

Does the study involve the administration or use of an approved drug or natural health product?

No

Will you be collecting any biological specimens?

No

Will you be creating or contributing to a bio-bank, bio-repository, registry, as part of the study?

No

Will you be doing any genetic testing or analysis?

No

Incidental and secondary findings

See Guideline for reporting incidental and secondary findings to study participants

Are any of the methods or procedures used likely (i.e., a real possibility and probability) to reveal an incidental finding (i.e., discoveries made in the course of research but that are outside the scope of the research and/or results that are outside the original purpose for which a test or procedure was conducted)?

No

Are any of the methods or procedures used likely to reveal a secondary finding (i.e., findings that are not the primary target of the test or procedure; rather, it is an additional result that is actively sought)?

No

Equipment use

Will there be any equipment used as part of this study?

No

Deception

Does the study involve deception or partial disclosure?

No

Risks and safeguards

Considering each method or procedure to be used in this study, indicate if participants might experience any of the following risks or harms

No known or anticipated risks

Outline the criteria for stopping the study early due to safety concerns/other issues.

None.

Privacy

Will demographic and/or background information be asked of participants? If so, ensure that the demographic questions have been uploaded in the methods section.

No

Participant identification - If applicable, include how participants will be referenced in study results.

By a study number

If applicable how will the key/list that links participants' codes with their actual name and/or consent forms be stored and protected? Also, outline how long the key/list will be stored.

Data will be stored in a password protected computer and deleted after the project has been completed. Any identifiable information that might connect the participant to their answer (like their ID) will be encrypted and stored in a different folder than the data from the interviews.

Are there any limitations to the promise of confidentiality?

No

Will any study data be leaving the University of Waterloo, the province, or country (e.g., member of research team is located in another institution, province, or country, etc.)?

No

Will any collected data or information be entered into a database for future use?

No

Are there other members of the research team who are not named on this application (e.g., co-op students, research assistants, or other temporary personnel) who may carry out specific tasks involved in your study?

No

Will individual participant identities be confidential in the publication or release of the study findings?

Yes

Data storage

What type(s) of data will be collected for this study?

Audio-recordings
Video-recordings
Electronic files

If identifiable features are captured, ensure this is described as part of the procedures section

---

For each type of information collected, identify where the data will be stored

Audio recordings will be kept in a local server in a password protected computer until they are transcribed. After the transcription the audio files will be deleted. The electronic files will be kept in a password protected local computer for up to 7 years after the conclusion of the project. After that date the files will be permanently deleted. The screenshots of the sketches made using ZOOM Whiteboard and any images the participants choose to share with the researchers will be kept in a local server in a password protected computer. The electronic files will be kept in a password protected local computer for up to 7 years after the conclusion

of the project. After that date the files will be permanently deleted. The screen recordings of the participants using the prototype will be stripped of any identifiable features as soon as the interview is done. After, the recordings will be kept in a password protected local computer for up to 7 years after the conclusion of the project. After that date the files will be permanently deleted. The written consent documents will be similarly kept in a password protected local computer for up to 7 years after the conclusion of the project. After that date the files will be permanently deleted.

For each type of data collected, identify the minimum retention period

Audio recordings: deleted after transcription Electronic files (including the screenshots, screen-sharing recordings, transcripts, and images shared): Erasing of electronic files after 7 years.

Data Management

Are there plans to link the data collected with other data sets, databases, or registries?

No

The Tri-Agency Open Access Policy on Publications and some journals are requesting that research data be provided to an open access repository to promote the availability of findings, to enhance transparency and share with the widest possible audience.

Do researchers plan to make the data-set available in an online repository/archive?

No

Do you have a data management plan?

Yes

Consent and Withdrawal

What member(s) of the research team will be responsible for obtaining informed consent?

Student investigator Luka Ugaya Mazza

Is there a relationship between the potential participant(s) and the person obtaining consent?

Yes

Explain the nature of the relationship (e.g., professor/course instructor, teacher, supervisor, employer, health care professional/physician, etc.)

Colleagues

Describe how undue influence and/or authority to participate will be mitigated.

Colleagues will be recruited based on their expertise and, as professor and investigators themselves, will have the option to decline participation without any pressure or commitment from the study researchers. It will be explained to them that they can decline participation and quit partway through with no consequences.

How will consent be obtained

Written signature on a consent form

Upload Information and Consent Materials - See resources and samples for creating information consent letters. Refer to the guide for creating an information consent letter on this webpage.

Upload Information and Consent Materials

CONSENT LETTER_PHASE 1 AND 2_V2_20230419.PDF

Study group

Group 1, which wil participate in phases 1 and 2.

Upload Information and Consent Materials

CONSENT LETTER_PHASE 3_V2_20230419.PDF

Study group

Group 2, which will participate in phase 3.

Do you anticipate that you will need to make special accommodations for your participant group?

No

Do you anticipate needing to put in place any special procedures when obtaining informed consent?

No

Will consent need to be re-documented throughout the life of this study?

No

Describe the participants rights and processes for withdrawing consent.

The participants will be informed by e-mail of their right to withdraw, in addition to the inclusion of this right in the consent forms. During the interviews, they will also be informed of their right to withdraw. In case they wish to withdraw their consent, they will contact the researchers through e-mail stating their wish. A follow-up e-mail will be sent to the participants in order to confirm their withdrawal. This e-mail will confirm that the participant contacted the researchers about a withdrawal and that the e-mail is being sent to confirm that the participant is no longer a part of the study.

Outline what will be done with the participant's information (data, samples, etc.) if they withdraw from the study.

The participant's data will be deleted.

Will any individuals taking part in this study be unable to provide their own informed consent?

No

Remuneration

Will there be remuneration provided to show appreciation for a participant's time, effort, skills, etc. to take part in the study?

No

Will participants incur any expenses by participating in the study?

No

Feedback and Appreciation

How will you show appreciation to participants for taking part in the study?

The participants will receive a follow-up email with a Feedback Letter, thanking them for participating in the study and restating information regarding the study, anonymity and contact from the researchers. They will also be sent a copy of the results when it is available.

When will feedback/appreciation be provided to participants (e.g., immediately after the session, at the end of a survey, mail results at time X.)?

Immediately after the session.

Upload Feedback/Appreciation materials

Upload Feedback/Appreciation materials

FEEDBACK LETTER_PHASE 1_V2_20230419.PDF

Study group

Group 1, but it will only be sent after phase 1

Upload Feedback/Appreciation materials

FEEDBACK LETTER_PHASES 2 AND 3_V2_20230419.PDF

Study group

Group 1 and 2, but it will only be sent after phases 2 and 3.

How can participants learn about the study results/obtain a summary of the findings if interested?

Results of the study will be published on a MSc thesis as well as at least one online publication.

Other Details

Provide any other information relevant to this study you wish to explain to the Research Ethics Board reviewers or to the staff in the Office of Research Ethics.

N/A

Other Attachments

Upload any additional study documents

Attachments

Attestation

As the Principal Investigator/Faculty Supervisor/Local Investigator, I attest to the following:

- I will ensure all co-investigators, collaborators, and student investigators listed on this application have reviewed the application contents and will conduct the study according to the application/protocol.
- I am aware that any changes made to the research must be reviewed and provided clearance before the changes are implemented. Change requests (i.e., an amendment) are to be submitted through the system. I am also aware ethics clearance for this study is valid for only 12 months unless I renew the study prior to the ethics clearance expiry date. If an annual renewal report is NOT submitted through the system prior to the expiry date, the study will be suspended, all work on the study must stop, and Research Finance will be notified which will result in a hold being put on the funds associated with this study.
- I agree to comply with the Tri-Council Policy Statement (TCPS2) for conducting research with human participants and with University of Waterloo policies and

guidelines when conducting this study (e.g., statement on human participant research, IST policies, etc.).

- I confirm I have read the University of Waterloo Research Integrity guidelines and I agree to comply with the policies and guidelines of my profession or discipline regarding the ethical conduct of research involving humans.

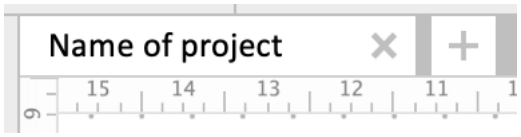
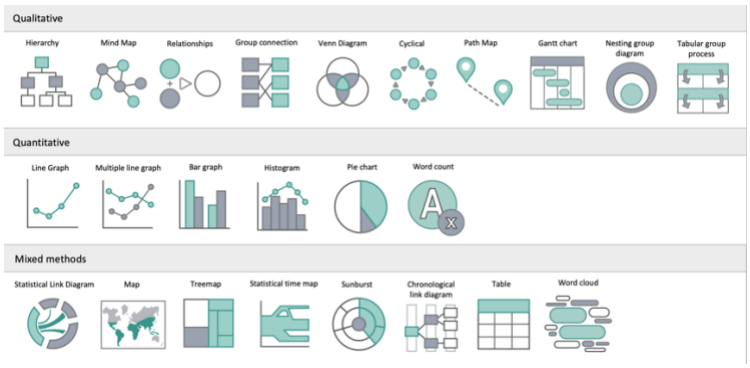
By submitting this application I agree to the above attestations and will ensure the research is conducted accordingly

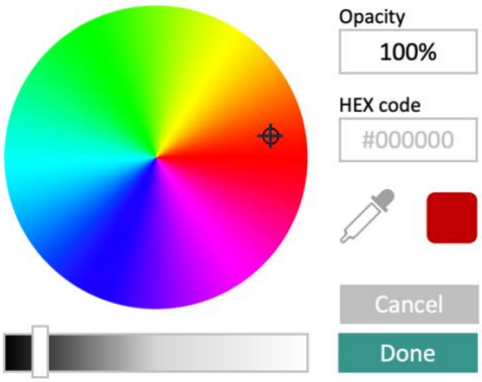
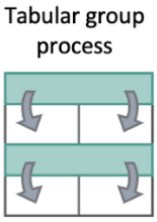
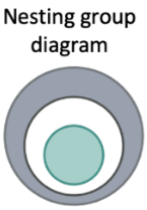
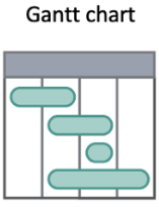
Only the Principal Investigator/Faculty Supervisor can submit the application. This acts as a signature indicating approval of the application.

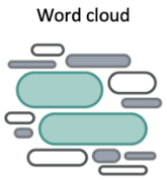
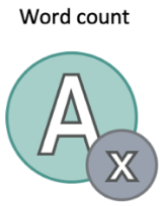
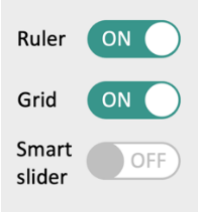
This is the end of the application form. Click submit in the right menu if you are ready to send it to the Research Ethics Office.

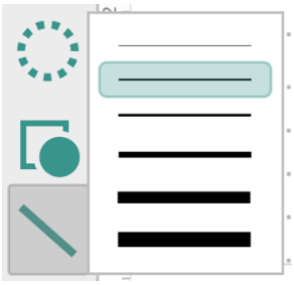
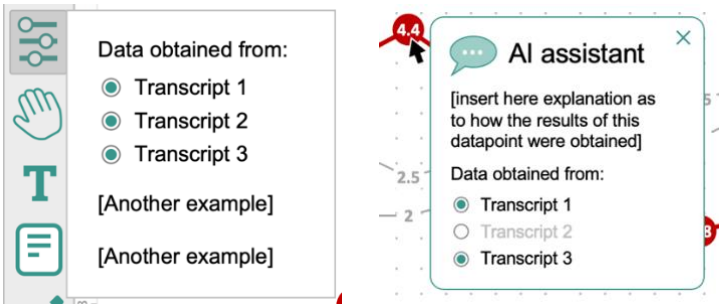
Appendix L

Changes made to prototype

Changes suggested by participants	Changes made to the prototype
Addition of tabs.	Added below the top bar. 
Addition of a text tool.	Option added to the editing side bar.
Adding a group for what datavis is good for qualitative, quantitative, and mixed methods.	Options added to the pop-up that shows all options for data visualizations. 
Adding a heading for the side of the top menu that is separated into categories.	Done. A heading of ‘categories’ was added.
Adding undo arrows.	Added to the top bar, alongside a menu and save button. 
Eye-drop, HEX code, opacity, and color wheel for colors.	Added as more options when the user selects colors.

	
Tool that allows for checking the design for color-blindness.	Not possible to do for the prototype.
Tabular grouping process.	
Nesting group diagram.	
Adding hierarchy into the possible datavis for processes.	Added.
Gantt chart.	

Word cloud.	
Precise word count datavis.	
Save visualizations both as project and as image files.	Not possible to do for the prototype.
Add a tutorial.	Not possible to do for the prototype.
Warn the user for any changes that might alter the visualization to the point of misinterpreting the results.	A warning was added in case the users try to do any change that will corrupt the data behind the data visualization. 
Employ smart editing that will change the visualization when the user tries to move elements that are connected (e.g., change the entire graph if one of the data points is moved).	Not possible to do for the prototype.
Add a smart slider to control the density of information presented, a ruler, and a grid option.	Added as toggle options to the top bar. 

Ability to change the thickness of lines.	Added as an editing feature with the lines. 
Responsive window for ‘all options’ that shows the newest and most used options by the user.	‘All options’ window was now changed to ‘most used’.
Make the window for ‘all options’ a pop-up.	It was made a pop-up with separations between qualitative, quantitative, and mixed methods appropriate data visualizations.
Double finger trackpad scroll functionality in the window for ‘all options.’	Not possible to do for the prototype.
Ability to edit the data that was input into the visualization (e.g., excluding specific data points, adding data that was not originally there, etc.).	Two options were added for editing data. The firsts one is on the side bar, and It allows for the user to edit all of the data that went into the visualization. The second is with the ‘AI assistant’ tool, in which the user right clicks over an element and a pop-up appears explaining how those results were reached and offering some editing options for the user. 
Fast changes to the visualization when something is altered.	Not possible to do for the prototype.