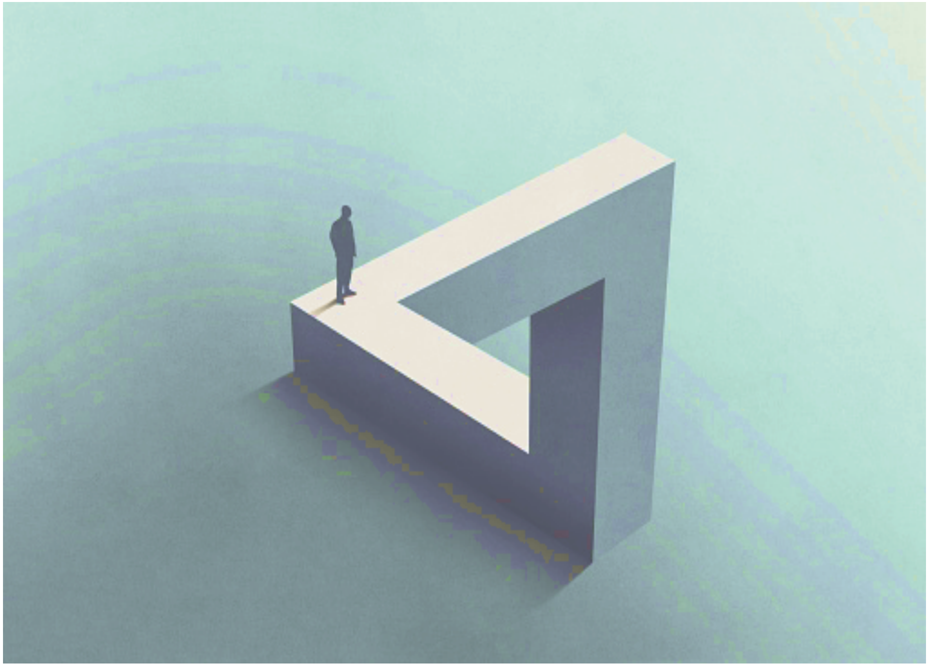
DEPARTMENT OF
APPLIED IT

A META-SYNTHESIS STUDY OF DATA VISUALIZATION IN CITIZEN SCIENCE



Xing Liu

Thesis:	15 hp
Program:	Learning, Communication and IT
Level:	Second Cycle
Year:	2023
Supervisor:	Marisa Ponti
Examiner:	Karin Ekman
Report nr:	2022:070

Abstract

How can citizen science use data visualization technology? What are the current status and the boundaries? As part of the systematic meta-synthesis study, a feedback loop has been developed to show how current statutes on citizen science data visualization can be interpreted in a comprehensive manner. In a Citizen Science context, data visualization goes beyond technical efficiency, abstract data, and symbolic meaning, incorporating material traces, and social domains such as viewer perceptions, individual experience, and pre-knowledge into meaning-making interpretations of visualization. Visual practices are available throughout all the phases of Citizen Science projects. Furthermore, it has been shown that the process of visualizing data in Citizen Science is also a learning experience. Lessons learned from best practices in the literature lead to recommendations that can assist educators in using data visualization technology in Higher Education to support students' learning.

Keywords

Data Visualization, Visualization, Citizen Science, Learning, Higher Education

Titel

EN META-SYNTESSTUDIE AV DATAVISUALISERING I MEDBORGAR-FORSKING

Sammanfattning

Hur kan medborgarforskning använda teknik för datavisualisering? Vad är nuvarande status och vilka är gränserna? Den här systematiska meta-syntesstudien ger en omfattande översikt över datavisualiseringens feedbackloop i medborgarforskning. I ett medborgarforsknings-sammanhang går datavisualisering bortom teknisk effektivitet, abstrakta data och symbolisk mening genom att inbegripa materiella spår och sociala domäner som åskådaruppfattning, individuell erfarenhet och förkunskaper i meningsskapande tolkningar av visualisering. Visuella metoder är tillgängliga genom alla faser hos medborgarforskningsprojekt. Det har även påvisats att processen att visualisera data i medborgarforskning också är ett lärtillfälle. Lärdomar från bästa praxis i litteraturen leder till rekommendationer som kan hjälpa lärare att använda datavisualiseringsteknik för att stödja studenters lärande i högre utbildning.

Nyckelord

Datavisualisering, Visualisering, Medborgarforskning, Lärande, Högre utbildning

Foreword

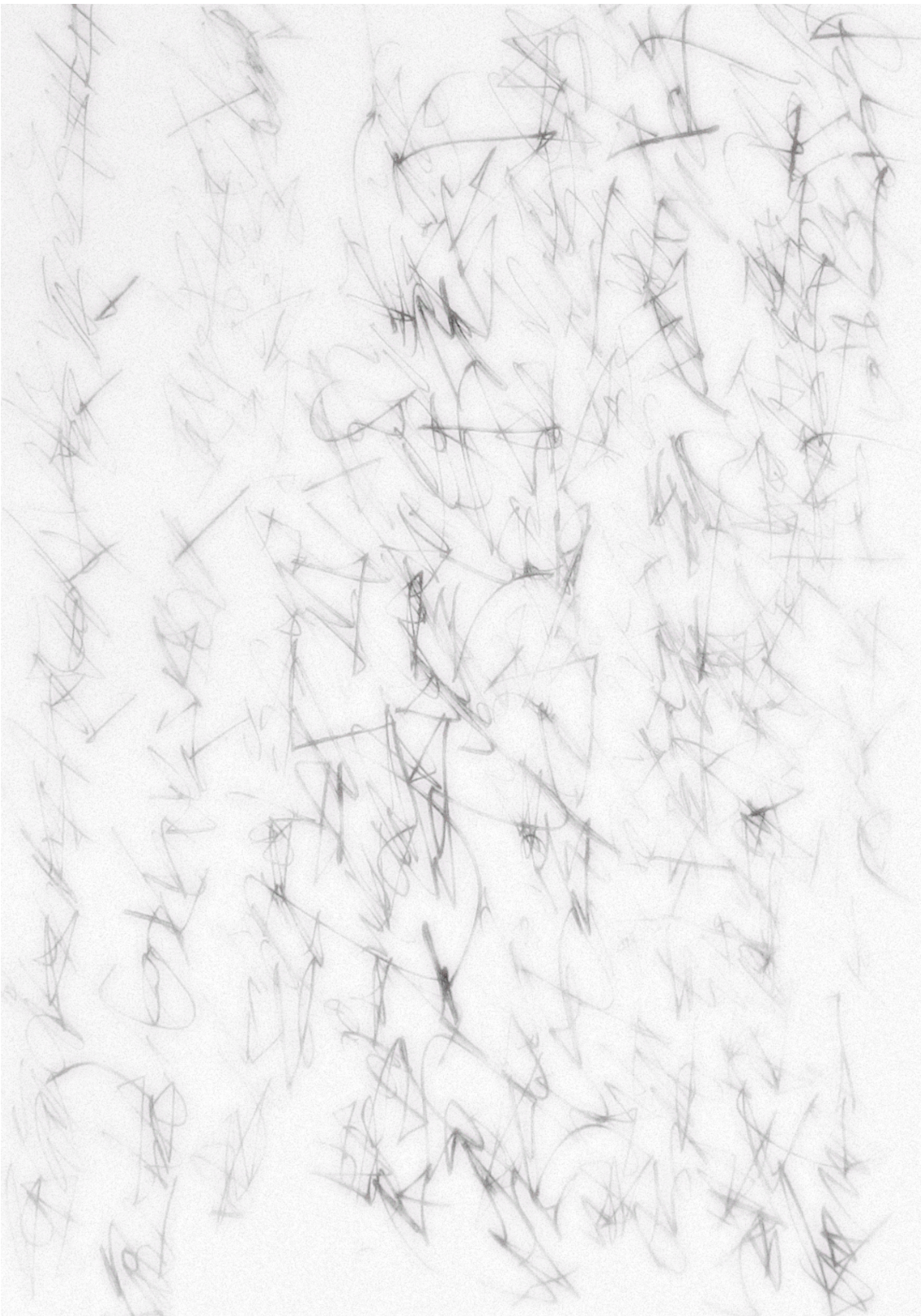


Table of contents

Figures

Tables

Acronyms and Abbreviations

CS	Citizen Science	
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	
Wos	Web of Science	
GPS	Global Positioning System	
IoT	Internet-of-Things	(Claes et al., 2018)
InfoVis	Information visualization	
DV	Data Visualization	
VGI	Volunteered Geographic Information	(Navarra et al., 2020)
PPSR	Public participation in scientific research	(Sandhaus et al., 2019)
NEUVis	Non-expert's user visualization design	(Snyder, 2017)
UEMs	Usability Evaluation Methods	(Biraghi et al., 2020)
EDSS	Environmental Decision Support Systems	(Gray et al., 2002)
IDE	Interactive Development Environment	(Navarra et al., 2020)
S2S	Source-to-Sea	(Clark et al., 2021)
AMDI	Australian Marine Debris Initiative	(Clark et al., 2021)
eDNA	Environmental DNA	(Kandlikar et al., 2018)
WFD	Water Framework Directive	(Biraghi et al., 2020)
NASA-TLX	NASA-Task Load Index	(Hoyer et al., 2021)

1 Introduction

In science, data visualization techniques are key to making sense of data and communicating it effectively. Citizen Science (CS) is a collaborative practice of engaging citizens in scientific research projects (Kullenberg & Kasperowski, 2016a). While Citizen Science data can reveal fascinating discoveries, communicating them efficiently can be challenging. According to Bonney et al. (2009a), Citizen science involves engaging citizens in various phases of authentic scientific research, especially data collection and data analysis (Bonney et al. 2009a). An example is the eBird project (https://ebird.org/). As one of the largest citizen science projects, eBird alone contributes to over 100 million bird sightings worldwide (Sullivan et al., 2016). In another example, over two million volunteers have worked on Zooniverse to analyze and classify data from galaxies, fossils, and much more. Involvement in citizen science activities fosters knowledge generation, sharing of ideas, and evidence-based decision-making, which has great potential to enhance scientific literacy among the public (Roche et al., 2020). Additionally, the possibility of analyzing the participant's own data will provide motivation and support for ensuring real engagement as well as educational impact (Vries et al., 2019). As Sherbinin et al. (2021) emphasize that the contribution of non-professional participants to scientific research has become increasingly important. However, to be understood and interpreted clearly and accurately, data must be transformed into meaningful information. Taking Citizen Science data and transforming it into meaningful information is challenging.

Data heterogeneity and quality are key barriers to transforming, manipulating, and visualizing citizen science data because citizen science initiatives cover a wide range of topics, designs, and research needs. As Pocock et al. (2014) point out, analyzing citizen science data can be challenging because of its complexities. Another barrier is typically caused by the lack of knowledge and skills of non-expert scientists when interpreting and communicating data. Consequently, Heer et al. (2010) emphasize the importance of understanding data, relating it, and communicating it in a meaningful way (p.59). It has been proven that well-designed graphics are much more effective than other forms of data interpretation and communication, such as text, numbers, and spreadsheets, by Tufte (1983). Visualization enhanced by advanced digital technology has already been widely developed and applied to support the data analysis of expert scientists (Wong, 2012). Well-designed data visualization is even more crucial when trying to simplify complex scientific data into an easy-to-understand format for citizen scientists, who may lack technical or subject-domain knowledge. Apart from this, the research community has already proved that data visualization has the potential to appeal to a broad audience, motivate participation, improve data collection, control data quality, and promote decision-making (Gray et al., 2021). It is no doubt that there is a growing interest in and emphasis on best practices in data visualization in general. However, there is a lack of research to provide an overview of the current state of this topic, and best practices are hidden, which makes it difficult to guide researchers seeking to achieve the same objectives.

In this thesis, a meta-synthesis of the literature is employed in a systemic manner as a means of better understanding the current research status on the application of data visualization technology in citizen science. The theory, technical features, opportunities, and challenges of a wide range of research disciplines are examined in a systematic manner. Each research question is outlined with its key findings and followed by a list of best practices for data visualization in citizen science. Based on the lessons learned from these best practices, recommendations are made for educators in Higher Education to use data visualization to improve student learning. Ultimately, this thesis is evaluated for its limitations and its contribution to the field of Learning, Communication, and Information Technology, as well as suggestions for future research.

Research Questions:

RQ1: What research has been carried out on the application of data visualization in citizen science?

RQ2: What are the technical aspects of the applications used to visualize data?

RQ3: What are the challenges and opportunities of applying data visualization to citizen science?

2 Background

This chapter firstly presents an overview of data visualization in section 2.1. By introducing concepts such as data, information, knowledge, and the flow of data into knowledge, the question of why visualize data is reasoned and argued. In section 2.2, the discussion of data visualization focuses on the citizen science context. After conceptualizing citizen science, why visualizing data is essential to citizen science is grounded.

2.1 Data Visualization

2.1.1 Data and how we process it

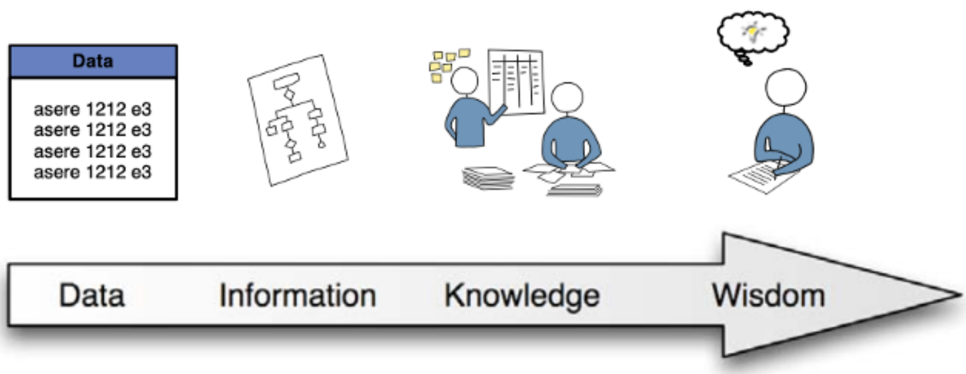
What will the weather be like later today? Do I need to bring rain clothes in case it rains when I get home later in the evening? It is highly unlikely that anyone is going to be able to determine what the weather would look like with just satellite image data or other sources of observed data alone. What we need instead is a form of information that is understandable to help us with our individual decisions, such as a weather forecast. As Few (2009) points out, information is generated by items, entities, and things that do not directly correspond to physical structures or objects. In this example, data is satellite images and other kinds of raw data. Information is the weather forecast, which is created by processing raw data to provide context and meaning. Furthermore, knowledge is created when information is integrated with human experience (Mazza, 2009). Every individual has different experiences and insights, making the interpretation unique. Mazza (2009) argues “the development of knowledge should be the principal aim of any communication process.” (p. 9). Knowledge is what we know, and knowledge may result in action. According to Vries (2018), Table 1 shows a brief overview of the main differences between the three basic concepts of “data” “information” and “knowledge” (p. 4).

Table 1
Main differences between data, information, and knowledge. Adapted from Vries (2018, p. 4).

	DATA	INFORMATION	KNOWLEDGE
	Objective	Objective	Subjective
MEANING	No meaning	Has a meaning	Has a meaning for a specific audience
PROCESS	Unprocessed	Processed Organized and presented in a suitable format	Processed & understood
EXAMPLES	Satellite image data	Weather forecast	Disaster warning/ Climate change

Mazza (2009) illustrates the flow of "From Data to Wisdom" (see Figure 1) to visualize how we process data to knowledge and wisdom. A piece of data is meaningless on its own. Mazza (2009) argues that "to give meaning to this data, data must first be processed, organized, and presented in a suitable format. This transformation and manipulation of data produce information." (p. 9). When information is integrated with a specific audience's experience, it creates knowledge. Wisdom is the highest level of comprehension. It can be defined as the stage in which a person has acquired such an advanced level of knowledge of processes and relationships that it is then possible to express a qualified judgment on data.

Figure 1
The continuum of understanding.



Note: The continuum of understanding [Illustration], according to Nathan Shedroff (as cited in Mazza, 2009, p. 9)

The first step of the "From Data to Wisdom" flow is to make the transition from data to information in the continuum of understanding. According to the book Information Design, in order to create meaningful information, it must be organized and presented in a way that is accessible to the target audience and communicates context around it (Jacobson & NetLibrary, 1999). A central question is then, how to transform and manipulate data into information? In the next section, data visualization is introduced as a potential solution.

2.1.2 Why visualize data?

The way how humans see plays a crucial role in the way process information. Vision is known as a dominant sense compared with smell, taste, and hearing. That is because, on one hand, human vision is pre-attentive, which means we see things before we pay attention to them (Appelbaum & Norcia, 2009). Human brains process information from some basic visual features, such as shapes, sizes, colors, positions, etc., in an extremely selective manner during the pre-attentive process. On the other hand, Koch and Tsuchiya (2007) report that human vision is capable of consuming a wide variety of information for sensing. Brains are capable of processing more information via human vision in both a cognitive and memory-enhancing way. Humans tend to prefer visualizing complex information over poring over numbers, texts, spreadsheets, or other forms of information, especially when dealing with abstract data that does not correspond to physical space (Mazza 2009, p. 11). In short, Pocock et al. (2016a) argue that the human brain is extremely good at interpreting visual information, detecting patterns, and detecting outliers. They also point out that data visualization is a way for us to access information through our eyes and through manipulation with interactive systems (Pocock et al., 2016a). As Tufte (1983) claims, well-designed graphics communicate messages from data effectively.

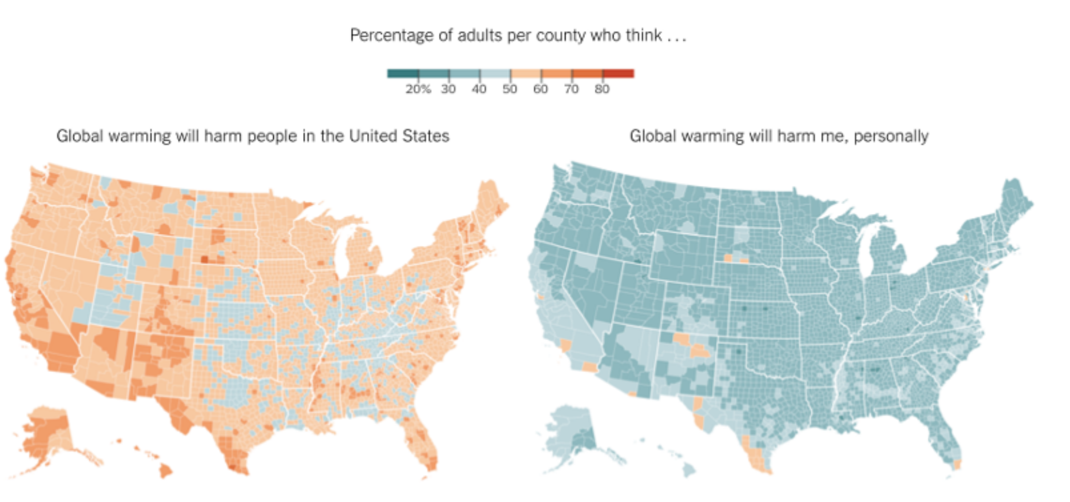
2.1.3 Conceptualizing Data Visualization

There are numerous definitions of data visualization, which means the concept cannot be formally defined (Chen, 2017). In line with the research questions, this study is well served by the explanation of Few (2009) argreed by Chen in 2017: "the use of images to represent information . . . it provides a powerful means both to make sense of data and to then communicate what we've discovered to others" (p. 2). Regarding the definition of data visualization, analysis, and communication are the two primary goals. The visualization aiming for analysis is also called "visual analytics", as it helps make sense of the data visually. The analytical data visualizaiton divides into two types, *exploratory visualization,* and *confirmatory visualization*. Exploratory visualization helps extract objectives from data and generate new hypotheses based on the data. If there are already hypotheses and it needs to be confirmed or checked whether the hypothesis holds the data or not, then is called *confirmatory visualization*. In contrast, *explanatory* visualization is used for communication purposes. An explanatory visualization supports the generation of a message from the meaning of the data to reach the target audience. When using data visualization technology for communication, Nicholson-Cole (2005) suggests there are some important aspects need to be taken into account. Such as who is the target audience, and what view of the data they want or need to know, as well as based on previous experience and knowledge, can the target audience understand the visualization effectively and accurately? In conjunction with visualization technology, what and how information is communicated has a significant impact on individual interpretation, as well as the choices they make about future issues. Data visualization approaches are designed for certain goals. Therefore, it is important to analyze the goals of the data visualization approaches to identify which approaches and tools will be most appropriate for a given task.

A Data Visualization Example

The Yale Program on Climate Communication conducted the "Climate Change in the American Mind" project. Based on an analysis and visual representation of survey data, the pattern was found by Popovich et al. (2017) that most Americans were concerned about climate change, but they didn't expect to be impacted personally. As shown in the left visualization in Figure 2, most Americans know global warming is happening, and most agree it harms people in the United States. However, the data visualization on the right, reveals most Americans believe global warming won't harm them personally. In this couple of visualizations, the pattern of the problem is clearly visible, namely the problem of people's risk perception of global warming. The visualizations indicate the message that the global warming is known for its enormous long-term consequences but has a little short-term impact on individuals.

Figure 2
How Americans Think About Climate Change.



Note: From *How Americans Think About Climate Change*, [Illustration], by Popovich et al., 2017, Source (https://www.nytimes.com/interactive/2017/03/21/climate/how-americans-think-about-climate-change-in-six-maps.html)

2.2 Data Visualization in Citizen Science

This section will discuss data visualization in the context of citizen science. The first step will be to define citizen science and then clarify the purpose of using data visualization technology in citizen science. Additionally, the opportunities and challenges of incorporating such technology into citizen science projects are discussed.

2.2.1 Citizen Science

Science, which has traditionally been practiced by professionals, has been significantly impacted by the rapid development of information technology. By utilizing emerging technologies, public members who are not experts can participate in authentic scientific processes, including digital platforms, mobile applications (apps), geographic information systems (GIS), and graphical user interfaces (Pocock et al., 2015). In scientific research, citizen scientists have become increasingly important sources of data collection (Sherbinin et al., 2021). In addition to collecting data, non-expert citizen scientists address scientific questions (Bonn et al., 2016) communicate with other interested parties, share ideas, and contribute to decision-making (Hoyer et al., 2021). A citizen science approach has ample potential for transdisciplinarity and for integrating natural, physical, and health sciences with the humanities (Roche et al., 2020), and leads to an overall advancement in scientific understanding (Bonney et al., 2009b). The evolution of citizen science can be seen in the bird observation research example.

An example of Citizen Science

Back in the year 1900, citizens started to participate to observe, classify, and collect information about the Christmas Bird Count (CBC) for the National Audubon Society annually. Citizen science began in this way, but it wasn't until the 1990s that the concept of citizen science was introduced. Over the past three decades, Cornell Laboratory of Ornithology has engaged the public from around the world in observing birds and collecting data. In 2002, they launched eBird (https://ebird.org/), one of the largest citizen science projects in the world, which contributes more than 100 million bird observation data each year. With the help of citizen scientists involved in eBird, scientists can identify how pollution, disease, climate change, and other factors affect birds.

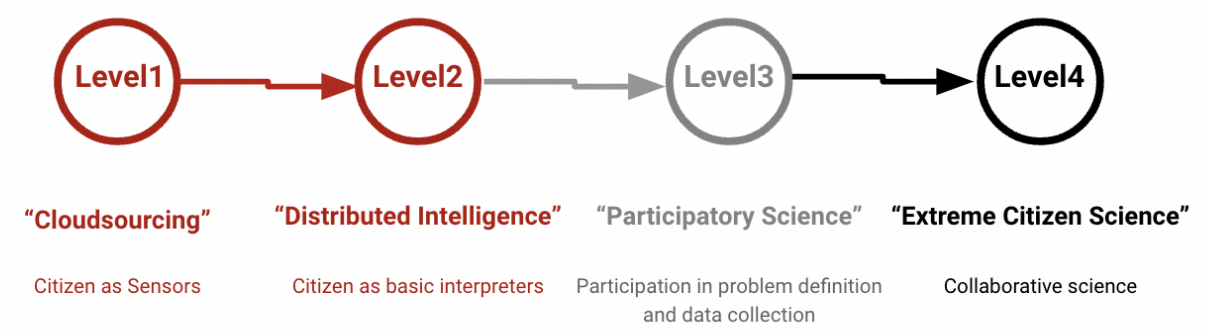
Non-expert Citizen Scientists

In the field of Citizen Science, there are various terms for citizens who are engaged, such as participant, citizen scientist, volunteer, and so on and so forth. It is worth mentioning that in this study, the term non-expert citizen scientist, short for citizen scientist, is used. Non-experts stand in contrast to experts, who are typically professional scientists in the respective scientific field, or a professional with expertise in organizing Citizen Science projects, as well as expertise in other professional donations such as data scientists (Kasperowski & Hillman, 2018). Even though citizen scientists may from many perspectives not necessarily be experts, they are increasingly involved as more access to data and digital tools are made available (Sherbinin et al., 2021). As Newman et al. (2012) cite several studies on citizen science, such as Danielsen et al. (2009), Dickinson et al. (2012), and Miller-Rushing et al. (2012), they highlight that the support of accessing data and applying digital tools, advanced non-expert citizen scientists in analyzing their own data, interpreting, and communicating them with their own concerns and interests.

In line with Arnstein (1969), Roche et al. (2020) argue that the role citizen science may play in activism and advocacy is crucial when it comes to education and learning. Citizen science practice could be exercised as one means of educating active citizens by empowering communities to advocate for their local environment through research, or by enabling citizens to gather evidence on, and articulate, pressing issues (Roche et al., 2020, p. 4). Building on Arnstein's concept, Haklay (2013) designs a framework for classifying the level of participation and engagement of citizen scientists is proposed against the technical, social, and cultural background of citizen science. He suggests learning will be enhanced as the public gets more involved and engaged in citizen science.

At the most basic level, Haklay (2013) explains that it is *crowdsourcing* since volunteer computing relies on a large number of individuals providing resources, with very little cognitive engagement. A second level consists of *distributed intelligence*, which uses the cognitive abilities of participants as a resource. As part of this type of engagement, volunteers may ask questions during their time working on the project, and it is important to be aware of how to support their learning beyond the initial training. Next, *community science* involves participants defining the research questions and devising data collection methods in consultation with scientists and experts. Participants collect data, but they are not involved in analyzing their own collected data, Haklay (2013) claimed it is "perhaps because of the level of knowledge that is required to infer scientific conclusions from the data." (p. 117). In the end, collaborative science is integrated into one piece at the fourth level, which is referred to as *extreme citizen science*. At this level, it is possible for participants to be involved in the analysis, publication, or use of the results, depending on the level of engagement they choose. As Roche et al. (2020) argue that knowing and acquiring knowledge are integrated as active and collaborative processes in this level of engagement.

Figure 3
Levels of participation and engagement in citizen science projects.

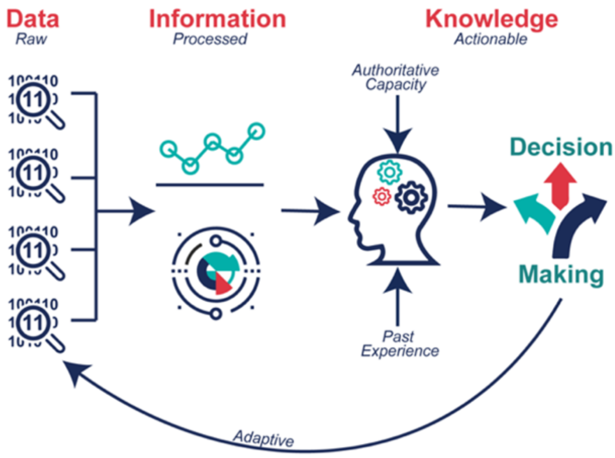


Note: *Levels of participation and engagement in citizen science projects* (adapted from Haklay, 2013, p. 116)

2.2.2 "From Data to Wisdom" flow in Citizen Science

The "From Data to Wisdom" flow discussed previously emphasizes the importance of organizing, processing, and presenting data in order to understand it, which also holds true in a citizen science context. In Figure 4, the illustration by Patterson (2018) shows an example of the flow from data to information to knowledge in water data in citizen science. The raw collected data is processed into understandable information by data visualization. When information is integrated with the background or experience of an individual, it produces knowledge that has the potential to influence decisions. The flow back to data can also be continued as new questions can be directed, such as what data is missing, or how the data collection method could be improved. In citizen science, the iterative process acts as a loop to dynamically improve the "From Data to Wisdom" flow.
Figure 4

The flow from data to information to knowledge.



Note: *The flow from data to information to knowledge*, [Illustration], Source (https://internetof-water.org/valuing-data/what-are-data-information-and-knowledge/)

2.2.3 Applying Data Visualization in Citizen Science

While the use of data visualization is a powerful tool for dealing with data in general (Pocock et al., 2016a), there are certain context-specific considerations that need to be addressed when applying it to citizen science. For scientists, according to Pocock et al. (2016b), there is growing evidence showing enhancing analyzing and communication data by applying data visualization technology. However, non-expert citizen scientists, who often have diverse backgrounds or skills, may not be capable of processing data in a meaningful way (Clark et al., 2021), since citizen science involves large and complicated scientific data. Nevertheless, access to scientific data with digital tools such as data visualization has opened opportunities for non-experts to participate in scientific research processes in depth. Newman et al. (2012) claim that for non-expert citizen scientists, visualization of data can efficiently support analyzing, interpreting, and communicating their own contributed data relating to their own concerns and interests. For example, through the investigation of the eBird CS project, the research by Sullivan et al. (2016) has found preliminary evidence showing the process of citizen scientists processing the data, generating information by data visualization, further taking evidence-based actions, and achieving conservation goals. The researchers demonstrate that in eBird, a web-based visualization application is provided to enable simple exploration, interpretation, planning, as well as decision-making functions for non-expert citizen scientists. It is reported that when new features are launched in data visualization to allow citizen scientists to track their own observation data and compare their own data to others, engagement increased nearly threefold (Sullivan et al. 2016).

Advancing scientific knowledge and communicating science are part of the primary goals of citizen science projects. In citizen science, participants want to communicate the scientific findings to know what their own data was telling them (Vries et al., 2019). Their study also highlights the findings of Krebs (2010) and Rotman et al. (2012), which argue that a lack of clear communication of scientific findings could cause dissatisfaction and demotivation among participants. For citizen science projects to be successful, engagement and interest are enhanced by good communication. Consequently, communication and dissemination are crucial, according to Rufenacht et al. (2021). Data visualization is a powerful means of communicating extensively, enhancing public engagement (Nicolson-Cole, 2005). By providing data, tools, and instructions for data analysis (Bonney et al., 2009a), non-expert citizen scientists had the ability to interpret their own collected data (Bonney & Dhondt, 1997). Citizen scientists could formulate local, evidence-based solutions by independently analyzing, visualizing, and verifying data, and can communicate these solutions to a broader scientific audience or decision-makers as needed (Gray et al., 2021, p. 2). According to Paleco et al. (2021), this communication process contributed to the democratization of science (p. 262). Furthermore, McCallie et al., (2009), and Haywood and Besley (2013) have identified two communication approaches adopted in CS. A one-way communication strategy is used to boost outreach and dissemination, including disseminating results and recruiting participants. Two-way communication is often adopted to emphasize interaction and dialogue between participants, and foster collaborative work, relationship building, and learning (Roche et al., 2020. p 5). As Roche et al., (2020) argued a two-way participatory approach bridges the gap between science education and science communication and poses science as one of many types of knowledge (p. 6). Examples of recommendations were proposed by Hecker et al. (2018) as below, regarding communication in the CS field.

Hecker et al. (2018) propose recommendations for communication in citizen science, such as:

- *Websites and mobile apps* that are easy to use and built with different user groups in mind will attract wider participation.
- New technologies such as *social media networks and platforms* can help project managers to reach more potential participants, support a sense of camaraderie and community amongst participants, provoke more discussion amongst participating volunteers and scientists about the research question, and improve the flow of data outcomes to the participants and the flow of feedback to the organizers (Newman et al. 2010).
- Providing for and encouraging *participant feedback* throughout the project can reveal new opportunities to share informative materials, improve the research and data quality, and increase the educational potential of the project (Druschke and Seltzer 2012).
- *A good communication plan* is crucial for the success of a citizen science project and needs to be developed at the outset. Communicating with participants throughout the research process and sharing progress and interim outcomes can increase the engagement of participants and the learning of all involved significantly. (p. 9)

Every participant in citizen science is also involved in a learning process and effective communication enhances learning in citizen science (Veeckman et al., 2021). Regarding what learning means in online citizen science, Aristeidou and Herodotou (2020) in their systematic review identify four main learning categories such as attitudes toward science, understanding of the nature of science, topic-specific knowledge, and science knowledge. Besides this in-depth understanding of science and the scientific research process, learning the skills necessary to participate in a CS project is also considered a learning process (Miller, 1983), for example, the necessity of training to support data collection and analysis. Meanwhile, the training requirements were identified as a challenge in Citizen Science (Kloetzer et al., 2021, p. 301). Applying data visualization in CS brought opportunities for enhancing the learning of citizen scientists. By providing well-designed visualization applications (Bonney et al., 2009a), citizen scientists were able to integrate and explore their own data to encourage real engagement (Bonney et al. 2009b), and further strengthen learning (Dickinson et al., 2012).

3 Method

The study aimed to investigate current trends in applying data visualization technology in the Citizen Science context. The primary aim of the study was to map, interpret, and characterize current theoretical understandings, technical features and best practices, as well as identify opportunities and challenges in this field. Lessons learned were also used to develop recommendations based on evidence-based findings. Considering the lack of familiarity with this topic, performing a literature review would be an appropriate research strategy. Various types of literature reviews were compared, resulting in two alternatives for quantitative analysis and a meta-synthesis for qualitative research. Due to the collected literature sample being qualitative, a quantitative meta-analysis that interprets a particular phenomenon was likely to be the best choice for the research questions (Grant et al, 2009).

A Meta-Synthesis literature review was conducted to examine how data visualization technology was used in citizen science (RQ1, RQ2, RQ3). Later, based on the insights learned in the literature, recommendations for applying data visualization techniques in higher education were constructed. The Meta-Synthesis literature review was based on the Grounded Theory Literature Review Method presented by Wolfswinkel et al. (2013) and combined a guideline for Structured Literature Reviews from Karolinska Institutet University Library. A grounded theory analysis was an exploratory method that provided a means of investigating a phenomenon that has little prior research attraction (Milliken, P. 2010), which fits into the nature of the topic. The research design is illustrated in Table 2. Using the PRISMA flow, the three steps of defining, searching, and selecting were performed systematically. In addition to open coding based on Wolfswinkel et al. (2013), content analysis was also used to analyze the data.

Table 2
Research Design of Meta-Synthesis.

Structures	Stages	Task
Data Collection	1. DEFINE	Identify the fields of research Formulate and delimit research questions Define the criteria for inclusion/exclusion Find search terms and create search blocks
	2. SEARCH	Apply PRISMA flow to search in a structured way Select Database, Search structures
	3. SELECT	Refine the sample iteratively Two catalogs, different aspects Validity of the selected literature
Data Analysis	4. ANALYZE	Grounded Theory

3.1 Data Collection

3.1.1 Define

The first stage of data collection was to identify and refine the search terms in line with the research questions. Since Citizen Science and data visualization span various disciplines, a thesaurus might be helpful to identify search terms and synonyms. The ERIC database's terminology function was used to search for prior studies on this topic. In the early days of Citizen Science and Data Visualization, neither thesaurus existed, which may indicate a different approach was needed. The prior literature review studies were searched for using keywords such as "technology", "communication", and "learning" in "Citizen Science" within a broad range. A total of three literature reviews were analyzed, and key search terms related to citizen science were identified based on those reviews. In addition, several key articles and books were used to locate keywords related to data visualization. The definition was iterative and used techniques such as snowballing, as suggested by Waddington et al. (2012). In Table 3, search terms, words, or phrases are divided into two groups. Citizen science-focused terms are in group 1 (#1) and data visualization-focused terms are in group 2 (#2).

Table 3
Search Terms

Group 1/ #1 Citizen science focus	Group 2/ #2 Data visualization focus
"Citizen scien*" OR "participatory science" OR "crowdsourced science" OR "community science" OR "community-based monitoring" OR "public participation in scientific research" OR "Community science" OR "Amateur science" OR "volunteer monitoring"	"Data visualization*" OR "information visualization*" OR "visualization*" OR "Visual Analytics"

#1 represents group 1, #2 represents group 2.

3.1.2 Search

In the beginning, an experiment was carried out at the University Library in Gothenburg to select databases relevant to the research field. Scopus and Web of Science (Wos) were included as a result of two factors. Both databases were multidisciplinary databases having quality sources with peer-reviewed literature. In addition, both databases were recommended by the University of Gothenburg library. Regarding search strategies in this study, boolean operators "AND" and "OR" were used to link search terms to database logic when a topic contained multiple search terms. Quotation marks were also used to search for phrases.

To optimize the search structure based on the databases used, a test search was conducted in Web of Science and Scopus. Searching in all searchable fields can be done using one query in Web of Science by using the "All Fields" option. The "Topic" option only searches for titles, abstracts, author keywords, and Keywords Plus. As compared to a search for "All Fields"in the Web of Science database, a search for "Topic" was used as it provided more relevant hints. To make it equal in both databases, a "TITLE-ABS-KEY" search in Scopus was selected. Additionally, as neither Scopus nor Web of Science databases contained a thesaurus function, a free word search was used. All databases conducted the same search query (see Table 3) on the same date 2022-06-20. Languages other than English were excluded. Other document types like "data papers" and "early access" were also excluded. A total of 170 document results were found in Scopus and 76 were found in Web of Science (Table 4).

Table 4
Search Strategy.

Database	Search query		Included/ Excluded	Results
SCOPUS	#1 TITLE-ABS-KEY	8,628	Excluded n=6 • Data Paper (1) • Editorial (1) • Erratum (1) • Other languages than English (3)	170
	#2 TITLE-ABS-KEY	364,819		
	#1 AND #2	176		
Web of Science	#1 (Topic)	7001	Exclude n=4 • Data Paper (2) • Early Access (1) • Proceeding Paper (1) • Other languages than English (0)	76
	#2 (Topic)	182,891		
	#1 AND #2	80		

Only free text search in Scopus, and Web of Science, 2022-06-20.
#1 represents group 1, and #2 represents group 2 as presented in Table 4.

3.1.3 Select

PRISMA flow

In performing the literature review, the PRISMA diagram was adopted in Figure 6 to report the results. PRISMA stands for preferred reporting items and is largely used for Systematic Reviews and Meta-synthesis (Page et al., 2021).

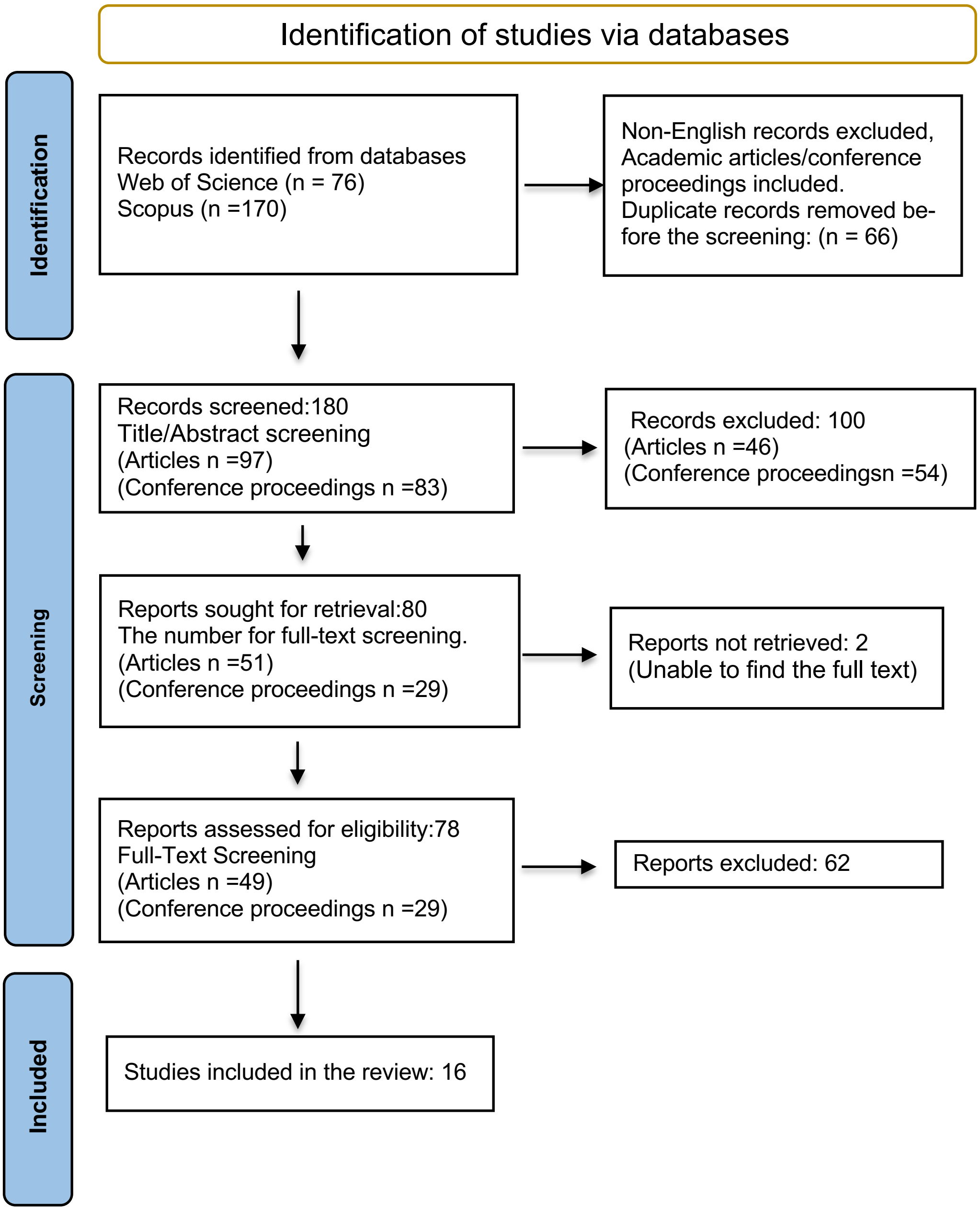


Figure 6. PRISMA flow, adapted from Page et al., (2021).

Iterative selection process

After removing duplicate articles, the final search was conducted on 2020-07-03 resulting in 187 hits in two selected databases. The 187 hits were grouped into two categories regarding the document type. Group One was peer-reviewed journal articles (n=97) and Group Two included conference proceedings (n= 83). Firstly, group one skimmed through the title, abstract, and keywords which resulted in 51 articles. The purpose of skimming group one first was to test how many samples could get from peer-reviewed articles. After a full-text scanning, 7 articles were included. The scope of the peer-reviewed academic articles had not fulfilled the requirement for the validity of a literature review study. Therefore, the same procure applied to group two conference proceedings. Group two contributed 9 hints. A total of 16 articles strictly conformed with the search criteria forward and backward in an iterative manner which illustrates in Figure 7, regarding the Grounded Theory by Waddington, et al. (2012, p. 49).

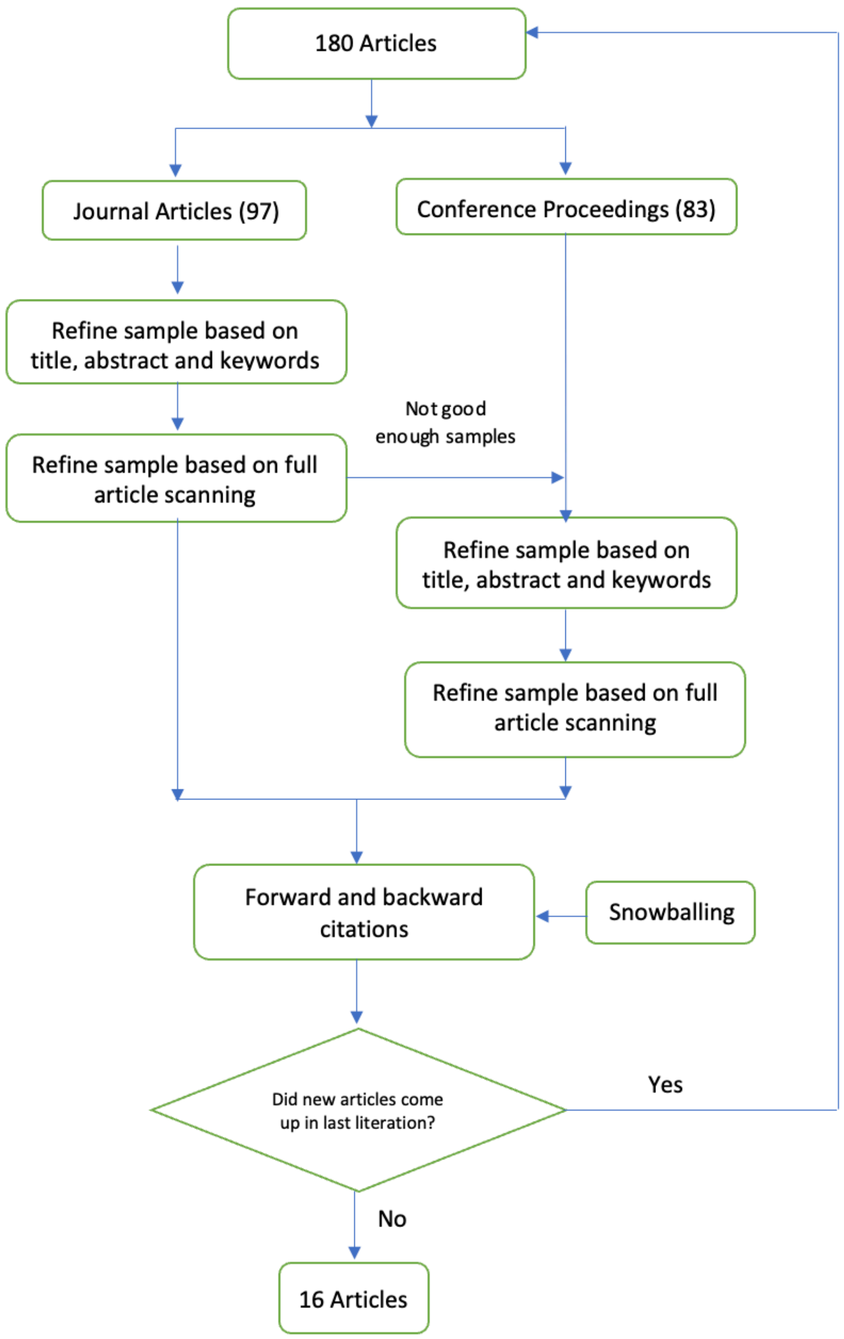


Figure 7. Iterative selection process, adopted from Waddington, et al. (2012, p. 49).

Inclusion criteria:

- Report on visualization technologies applied in the Citizen Science context.
- Examine or evaluate visualization practices.
- Written in English.

Exclusion criteria:

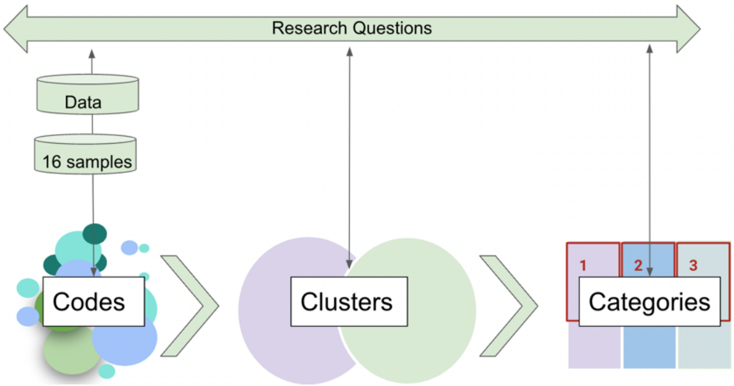
- Mainly introduce a citizen science project or an application associated with citizen science. Visualization was briefly mentioned as one of the functions. For example, the article "Reputation-aware filtering services for citizen science data' by Brooking and Hunter, 2011".
- Conceptual studies or models.
- No source found/available, i.e. "The Globe Zika education and prevention project".

3.2 Data Analysis

The overall structure of the data analysis of the literature review is organized on the research questions. There is an emerging study aimed at investigating the current trends in citizen science when using data visualization technology. In order to investigate phenomena with little prior research attraction, according to Milliken, P. (2010), a grounded theory approach is recommended as appropriate regarding the topic's nature.

In response to RQ2, nine (9/16) articles on technical aspects of data visualization were grouped together. In these nine articles, the data visualization application discussed was often not only developed for a particular CS project but could also be used for citizen science projects in general. In some cases, the application was part of a CS project, but its technical aspects were mostly examined. There were seven (7/16) studies in which the data visualization solution was part of a larger CS project. Therefore, the data visualization solution was discussed and examined within the Citizen Science project. In response to RQ1 and RQ3, all 16 articles were examed. The data analysis processes are illustrated in Figure 7. It should be noted that, due to the lack of an existing analysis framework in this field and the inherent nature of the grounded theory analysis method, another researcher may reach a different conclusion and choose a different grouping.

Figure 7.
Data Analysis Process.



Note: *Data Analysis Process* [illustration], adapted from Social Research (2021, March 1). Qualitative Content Analysis [Video]. YouTube. https://www.youtube.com/watch?v=arXQ91uh9rY

3.2.1 Process

Based on Wolfswinkel et al. (2013), a three-step analytical coding process (open, axial, and selective coding) was conducted in this study. They stated that the three-step analytical coding process was "performed in an intertwined fashion, going back and forth between papers, excerpts, concepts, categories, and sub-categories" (p. 51). Coding was done manually on paper since the data set was relatively small. Additionally, since the coder was new to the grounded theory approach, memoing was also used as a tool of grounded theory suggested by Cohen et al. (2018). A memo is a note written to oneself by the coder, as Strauss and Corbin (1990) argued, memoing contributes to the formulation of the emerging theory. In accordance with the recommendation by the researchers, the coder carefully manually read, continuously compared, and linked identified categorizations from 16 articles and memoing in Zotero software, until there were no new concepts, or properties according to the timeframe and other resources of the thesis work. An operative description for each coding process is given in Table 5 according to Wolfswinkel et al. (2013).

Table 5

A matrix of each coding process. Adjusted by (Wolfswinkel et al., 2013, p. 51).

Analytical coding steps		
1 open coding	inductive	Generating higher-abstraction level type categories from sets of concepts.
2 axial coding	deductive	Further developing categories and relating them to their possible sub-categories.
3 selective coding		Integrating and refining categories.

The first step was to select key excerpts from the articles. There are questions to guide the open coding such as, what is this data a study of? What is happening in the data? As a deductive process, axial coding establishes relationships between categories and subcategories. Tucker (2016) argues that this phase seeks to distinguish concepts based on whether they refer to causes, consequences, or intervening or contextual circumstances. In this study, *causes* and *consequences* seemed more natural. The third step elective coding identifies and develops links between the main categories to answer the research questions (Wolfswinkel et al., 2013, p. 51).

Following the advice of Hsieh and Shannon (2005), additional codes were developed as the analysis progressed, and the initial coding scheme was revised and refined (p. 1286). The initial coding scheme of open coding for this study was approached as follows:

- What it is?
- How does it test? What does it show?
- What are the opportunities and challenges according to their findings?

This stage resulted in a concept matrix using Microsoft Excel and memos in Zotero Note. As a result of size considerations, this thesis did not include the resulting matrix. An example of coding, categorizing, and memoing is presented in the box below:

An example of coding, categorizing, and memoing:

Text: The paper focuses on three challenging aspects for academics who would want to build on citizen-birthed information sets in order to better document and analyse minor heritage items:

- The heterogeneity of the Information (in terms of scope, of editorial choices, of quality, etc.), and behind it the heterogeneity of the Information Providers themselves. We present and discuss the strategy adopted in order to demonstrate potential added values of such a research on both sides (academics as well as information providers), and in order to pinpoint profiles of information providers.
- The factors of imperfection that are likely to be met when handling such information sets. We first give a global view of the data and information harvested from citizen-birthed e-sources, and then propose an exemplified list of the key factors of imperfection we came across during the research up to now.
- The design and implementation of visual solutions supporting analytical tasks in the specific context of imperfect information sets. We present and discuss the learnings of some of the solutions we have developed, solutions that are today available online (territographie.map.cnrs.fr). (Blaise et al., 2018, p. 19)

Coding, categorizing, and Memoing: Blaise et al. (2018) drew the conclusion of challenges in this text, which means the three challenges were not theoretical, instead they were proven challenges (context of "challenges"). The terms "heterogeneity" and "imperfection" appeared in other articles, as part of the issue of "data quality" in the context of "proven challenges". Since "the design and implementation" challenge had wide meaning, it remained open here to see if more precise categories could form from other articles to better describe the challenge of "the design and implementation". This process led to the formation of three levels of categories, such as "Challenges"-- "Data quality"--"Heterogeneity"/ "Imperfection". Additionally, "visual solutions supporting analytical tasks" were interpreted as "to analysis data was the data visualization goal" and grouped into categories "visualization goals".

As a result of the large number of column categories developed in this thesis, it was difficult to include these categories in a table. The abridged coding and categories resulted in three matrices (Table 6,7,8) using Microsoft Excel. More detailed information was listed as follows based on each research question:

- The categories such as Publication Outlines and Demographics, Subject Areas, Wordcloud, Networks Analysis, and Concept mapping were investigated to answer RQ1: *What research has been carried out on the application of data visualization in citizen science?*

The first research question in this study is to determine what research has been carried out on the application of data visualization in citizen science. Based on RQ1, the overview of CS projects that have been facilitated by data visualization technology in existing studies was analyzed. An analysis from a statistical as well as a theoretical perspective was beneficial since it was an emerging topic. Publication Outlines and Demographics, Subject Areas, Word Could, and Network Analysis have been examined for a statistical overview of this field from multi angels. The goal of Concept Mapping was to provide a better understanding of different emerging data visualization-related theories in the Citizen Science context.

In Table 6, A description of the data was provided, including Article name/Type, Publication, Place, Data type/Domain, and CS Project/ Application Name. Using the All Science Journal Classification (ASJC) function in Scopus was used to classify the journals and conference proceedings, the 16 articles were saved in the Scopus database and were classified, and analyzed based on ASJC standards. In Table 7, the subject area of each article is presented. Based on Tables 6 and 7, Publication Outlines and Demographics, as well as Subject Areas, were analyzed in Scopus and Microsoft Excel. This step helped to understand the scope of the datasets (16 articles) by time and country, to answer the underlying question of RQ1such as whether this topic was studied globally or only in certain geographic areas, or certain subject areas. As well, in order to get the perspective from different angles to better answer RQ1, Word Cloud and Networks Analysis were also applied. Using a web-based application (https://www.wordclouds.com/), the Word Cloud analysis conducted a visual representation of the most prominent terms from the abstract and conclusion parts of 16 articles, capturing important and frequently occurring terms. Networks Analysis (using https://www.researchrabbit.ai/) could be helpful to analyze citation connections between 16 articles, such as whether the research topic was dispersed or focused, and to what extent. In addition, the citation result from the Networks Analysis could be used to gain a better understanding of the reliability of the datasets by showing how often the 16 articles were cited. The results were discussed in section 4.1 and represented in Figures 9,10,11,12,13.

Furthermore, the investigation of Concept Mapping of visualizations in Citizen Science resulted from the direct observation of the datasets. There were two motivations for employing Concept Mapping. Firstly, based on the discussion of conceptualizing data visualization in the Background section, there were new concepts were introduced in the datasets. It was necessary to clearly describe the definitions of different data visualization-related concepts that appeared in the context of Citizen Science. Secondly, it might have the potential to contribute a more comprehensive conceptual understanding and therefore form a new landscape for applying data visualization in Citizen Science. Based on the relevance of the concepts, three sub-categories were grouped and formed: 1) concepts associated with data type and data quality; 2) concepts associated with a design perspective; and 3) concepts associated with a critical perspective. The results were discussed in section 4.1 and the results contributed "data visualization feedback loop in Citizen Science" (Figure 14) which was presented in the Discussion section.

- Three categories Primary Technical Features, Application Goals, and Usability Evaluation were grouped regarding answering RQ2: *What are the technical aspects of the applications used to visualize data?*

Frequently appeared terms in relation to answers' initial coding scheme "what it is?" helped to identify additional codes which present in Table 8. It included Data Type, Application type/name, Open Source, Visualization Goal, and Target Audience. These additional codes in Table 8 were divided into two categories, Primary Technical Features, and Application Goals. To determine which data visualization approaches and tools would work best for a particular task, it was important to analyze the goals of the data visualization approaches. Besides Visualization Goals, Target audiences were also important technical aspects that were frequently mentioned in the datasets to answer RQ2. Furthermore, since different types of data require different approaches to visualization (Hsu et al., 2018), it was necessary to determine which Data Type was visualized. Additionally, there was a high frequency of *open source* in the datasets. According to Gray et al., (2021), an open-source application is one that does not require licensing and is free to use. By providing access to data operations and code modifications, Gray et al., (2021) argue that the open-source data visualization solution increases accessibility, transparency, participant engagement, and trust among participants in Citizen Science. Thus, the open-source feature was chosen for analysis. However, there needed to be a clear distinction here to help clarify the definition of *open source* and to distinguish it from another term *open data* in order to ensure correct coding. *Open data* is also often mentioned in Citizen Science. *Open data* refers to data or datasets that are open to the public, open in format, and openly licensed (definition drawn from the Open Knowledge Foundation).

When it came to the initial coding scheme "How does it test? What does it show?", referred to the evaluation. Mazza (2009) claimed that a data visualization solution should be evaluated based on its usability, functionality, effectiveness, efficiency, and usefulness. In data visualization, usability and functionality were most frequently mentioned in the selected literature. Typically, usability is determined by asking questions from a user experience perspective, such as, "is the graphical interface easy to use and intuitive?" (Mazza, 2009, p. 127) and whether the visualizations do enable the exploration of data (Blaise et al., 2018, p. 18). Functionality refers to the set of operations of an application. It can negatively affect usability if there are too many or too complicated functions in an application since it can be difficult to learn and use. Consequently, this study focuses on Usability Evaluation in an attempt to find evidence-based best practices. In section 4.2, the results of the three categories and their subcategories were discussed.

- In line with RQ3, the constant comparison of the literature identified the importance of discussing opportunities and challenges within the technical and social domains. In addition to open coding, keyword searching in Zotero software was also used to locate the context of opportunities and challenges. Since a variety of writing styles, focus, depth, and quality were observed in the articles, the synonym of challenge/challenges such as question/questioning, obstacle, hinder, drawback, risk, and the synonym of opportunity/opportunities (chance, possibility, benefit) were used in the search.

The defined additional codes developed by constant comparison included: Data Quality (heterogeneity, incomparability, uncertainty), Data Privacy (when it occurs and how to deal with it), Applicability (applicable data visualization frameworks or applicable design strategies), Knowing the target audience, Data visualization as an aid for communication and learning (using data visualization in formal and informal education, training and learning materials for data visualization such as explanatory text, video tutorials, support for understanding and analyzing/visualizing data, support for understanding the analyses/visualization), The impact of collaborative design on the social domain (open-ended design in developing open source applications, co-authoring CS narrative in public visualization with polling interaction).

The limitation of resources (such as time, financial and human resources) was one of the most discussed challenges in literature, motivated the investigation of the concern of applicability in the technical domain. In addition, Gray et al. (2021) highlight that the technical domain may not have the same long-term value as the social dimensions expressed by integrating technology as a bridge in Citizen Science (p. 13). How the data are visualized was crucial i.e., for effectively communicating science, however, so is how the target audience perceives the data visualization (Nicholson-Cole, 2005). Therefore, the categories of social dimensions in this study were divided into three focuses: 1) knowing the target audience; 2) data visualization as an aid for communication, learning, and education; and 3) the impact of collaborative design on the social domain. The results were coded on papers and memos stored in Zotero Note and organized by themes in text in Section 4.3.

3.2.2 Reflection

Despite being particularly useful for investigating what the data demonstrated, grounded theory analysis also had some drawbacks. Only one coder in this study made the interpretation and construction of the data subjective. In addition, a variety of factors were noted in relation to the articles, such as their quality, depth, and style. It was observed that the coding effort was a highly subjective endeavor since the author acknowledges that factors such as background, education, and understanding may have affected the coding effort. It is likely that other researchers would code the data set differently. Besides producing large amounts of data, Bryant and Charmaz (2007) criticize that standard rules for identifying and categorizing are lacking, which makes it difficult for researchers without skills to identify and categorize. I experienced that when analyzing data with grounded theory, it was easy to assume the most frequently occurring categories and concepts were the most relevant. However, I tried to keep in mind that concepts that appear just once in the data set may be just as relevant as those appearing more than once. More reflection on the analysis method was also discussed in section 3.4.

3.3 Limitations of Method

The collected dataset was small, only containing 16 peer-reviewed journey articles and conference proceedings. Since the results were drawn from materials that were published, it was difficult to claim that they are representative of the entire field. Time and resources were constrained, as were language, the background, and the judgment of the researcher. This research consisted of the part-time work of one individual over a period of four months. There could also be limitations associated with the method chosen for collecting and analyzing data. However, the study matures with time and peer reviews.

3.4 Ethics, Validity, and Reliability

In this study applying grounded theory, researcher bias was considered an important ethical consideration. Research biases can be deliberate (intentional) or unconscious (unintentional), according to Byrne, D. (2017). In line with unconscious (unintentional) bias, there was a good thing about the researcher's inexperience in citizen science and data visualization. The researcher was motivated to do this study mainly out of interest. For example, as a result of the lack of pre-knowledge, it's easier to keep open and immerse in the data so that the data spoke for itself. Additionally, the researcher was mindful that maintaining objectivity throughout the entire research process was essential to avoid deliberate (intentional) bias.

Further, the research method was carefully selected according to the research questions, the knowledge requirements, and the appropriateness of the study. A systematic PRISMA data collection flow was applied to improve the validity and reliability of the study. Manually analyzing 16 articles was used to gain a thorough understanding of the literature. This might be a limitation, but it was necessary for the research at this stage. Since this study had only one coder, establishing peer debriefing from the supervisor enhanced its reliability. As Lincoln and Guba (1985), and Manning (1997) suggested, negative case analyses were used in this study to achieve neutral or unbiased results (Hsieh & Shannon, 2005). However, there might occur limitations during the data analysis process, such as coding bias discussed in section 3.2.2. The detailed limitation caused by the research method was reflected after discussing the results in section 5.5.

4 Results

As stated in the previous section, the literature search inquiry yielded 246 documents, which were reduced to 180 after removing duplicates. Seventy-eight full texts were selected after screening titles and abstracts according to the selection criteria presented in the method chapter. Ultimately, 16 peer-reviewed journal articles and conference proceedings were retained in the review corpus after applying the selection criteria.

The 16 samples were divided into two clusters, where related codes were divided into categories based on the research questions. Nigh (9/16) articles addressed the technical aspects of data visualization. In response to RQ2, they identified primary features, application goals, and usability evaluation of the applications. As part of the CS Projects, seven (7/16) studies examined the design and impact of visualization solutions. In these studies, the visualization solution was often viewed in context. All 16 articles addressed RQ1 by providing profiles of the CS projects and mapping the concepts to identify emerging trends and critical aspects as well as RQ3 to identify opportunities and challenges in this field. Based on the detailed explanations for the formulation of categories and sub-categories have been discussed in the Method section. Each research question was answered thematically. Figure 8 illustrates the core structure of the results.

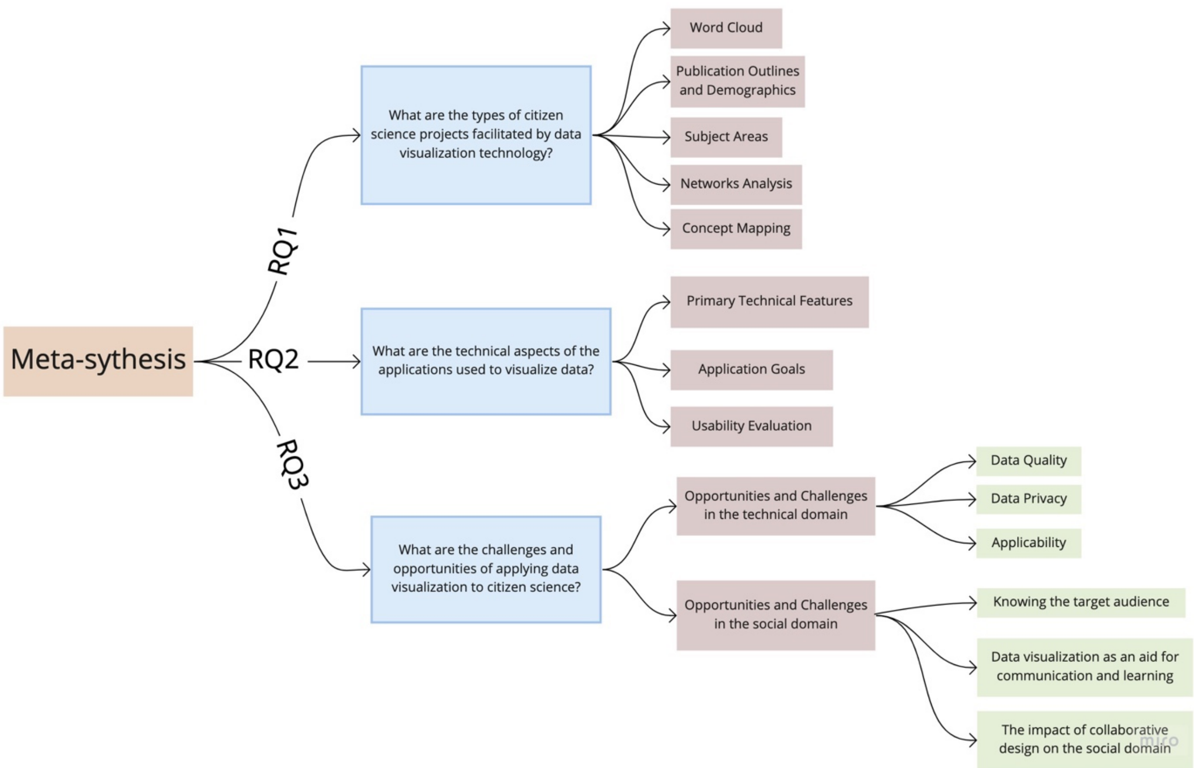


Figure 8. Research question thematic categories.

4.1 Citizen Science facilitated by data visualization (RQ1)

RQ1: What research has been carried out on the application of data visualization in citizen science?

The first research question in this study seeks to determine an overview of existing studies of what studies were conducted for CS projects have been facilitated by data visualization technology. An analysis of the current state of citizen science visualization studies was conducted from a statistical and a theoretical perspective. Statistical analysis examed Publication Outlines and Demographics, Subject Areas, Word Could, and Networks Analysis. The investigation of existing data visualization-related theory in CS resulted in Concept Mapping. Responses to RQ1,16 articles are summarized in a matrix in Table 6, and the detailed subject area matrix is presented in Table 7.

Table 6
Studies of data visualization in Citizen Science.

Article	*Type*	*Publication*	*Place*	*Data type/ Domain*	*CS Project/ Application Name*
(Snyder, 2017)	Conference Paper	ACM Conference on Computer Supported Co-operative Work, CSCW	U.S.A, Canada	the health of marine ecosystems	Coastal Observation and Sea-bird Survey Team (COASST)
(Offenhuber, 2020)	Journal Article	IEEE Transactions on Visualization and Computer Graphics	Germany	Environmental pollution	Staubmarke/ Public Lab
(Nicholson-Cole, 2005)	Journal Article	Computers, Environment and Urban Systems	UK	Climate Change	--
(Sandhaus et al., 2019)	Journal Article	International journal of environmental and science education	U.S. A	Environmental health	Gardenroots
(Blaise et al., 2019)	Conference Paper	2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)	EU, France	Minor heritage	--
(Hoyer et al., 2021)	Journal Article	KN - Journal of Cartography and Geographic Information	Eu, German	Biodiversity	Experiencing Biodiversity
(Claes et al., 2018)	Conference Paper	ACM International Conference Proceeding Series	EU	Air pollution	Part of OrganiCity, European H2020
(Navarra et al., 2020)	Journal Article	Workshop on Visualisation in Environmental Sciences (EnvirVis)	EU, Norway, Sweden,	spatiotemporal user-generated climate-related data	CitizenSensing
(Oliver et al., 2020)	Conference Paper	ACM Designing Interactive Systems Conference	Australia	acoustic sensing, bird, wildlife	--
(Tuppen et al., 2016)	Journal Article	Fontes Artis Musicae	UK	Music- bibliographic data	A Big Data History of Music
(Biraghi et al., 2020)	Conference Paper	ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences	Eu, Italy-Switzerland	lake monitoring	SIMILE (Informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems)
(Hsu et al., 2018)	Conference Paper	DIS 2018 - Companion Publication of the 2018 Designing Interactive Systems Conference	U.S.A,	Environmental sensing and public health data	EHC (Environmental Health Channel)
(Clark et al., 2021)	Journal Article	Water International	Australia	Environmental, marine debris	CesiumJS
(Abramov et al., 2022)	Journal Article	IEEE Transactions on Visualization and Computer Graphics	U.S.A,	Cosmological simulation	CosmoVis
(Gray et al., 2021)	Journal Article	ISPRS international journal of geo-information	Canada	Environmental, water quality monitoring (CBWQM)	CWDAT (Community Water Data Analysis Tool)
(Kandlikar et al., 2018)	Journal Article	F1000Research	U.S.A,	Environmental, eDNA	Ranacapa

Publication Outlines and Demographics

Despite one article's publication in 2005, the other 15 articles were published between 2015 and 2022. The publication outlets included nine journal articles (56.3%) and seven conference proceedings (43.8%), see Figure 9. Research is conducted in several countries (see Figure 10): the U.S.A. (6 articles), Australia and the UK (two papers each), Belgium, France, Germany, Italy, Netherlands, Norway, and Canada (one paper each).

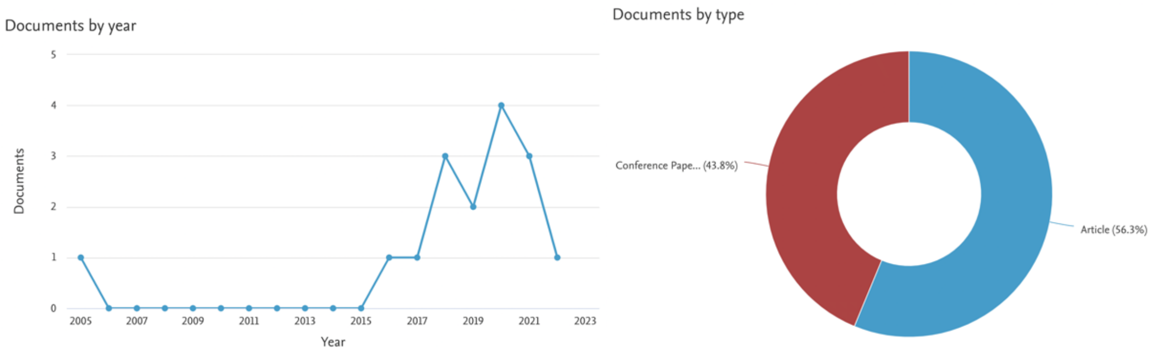


Figure 9. Documents by year and documents by type of publication.

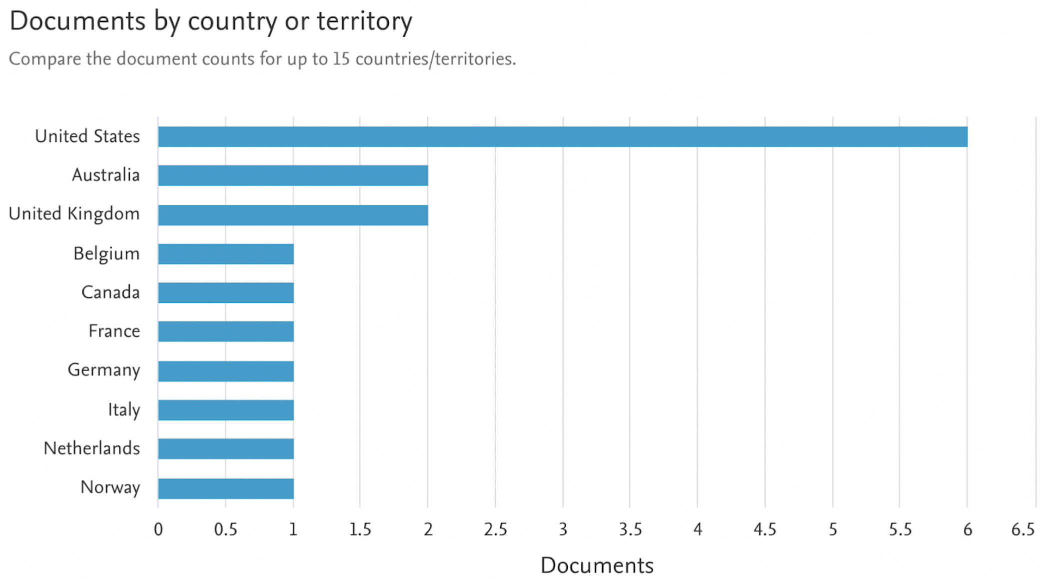


Figure 10. Documents by country.

Subject Areas

There were nine domains (see Figure 11) in the data set regarding the citation analysis method All Science Journal Classification (ASJC), which was used in Scopus to classify journals and conference proceedings. Computer Science accounts for 37.5% of all research and Social Sciences for 20.8%. Environmental Science accounted for 12.5% and Earth and Planetary Sciences accounted for 8.5%. Arts and Humanities and Decision Sciences contributed one study each (4.2%). Kandlikar et al., (2018)'s article alone covered three domains: Biochemistry, Genetics, and Molecular Biology, as well as Immunology and Microbiology.

Table 7
Subject Areas.

	computer science	*social science*	*environmental science*	*Earth and Planetary Sciences*	*Pharmacology, Toxicology and Pharmaceutics*	*Biochemistry, Genetics and Molecular Biology*	*Arts and Humanities*	*Immunology and Microbiology*	*Decision Sciences*
Number	9	5	3	2	1	1	1	1	1
(Snyder, 2017)	computer science								
(Offenhuber, 2020)	computer science								
(Nicholson-Cole, 2005)		social science	environmental science						
(Sandhaus et al., 2019)		social science							
(Blaise et al., 2019)	computer science								Decision Sciences
(Hoyer et al., 2021)				Earth and Planetary Sciences					
(Claes et al., 2018)	computer science								
(Navarra et al., 2020)	computer science		environmental science						
(Oliver et al., 2020)	computer science								
(Tuppen et al., 2016)		social science					Arts and Humanities		
(Biraghi et al., 2020)	computer science	social science							
(Hsu et al., 2018)	computer science								
(Clark et al., 2021)			environmental science						
(Abramov et al., 2022)	computer science								
(Gray et al., 2021)		social science		Earth and Planetary Sciences					
(Kandlikar et al., 2018)					Pharmacology, Toxicology and Pharmaceutics	Biochemistry, Genetics and Molecular Biology		Immunology and Microbiology	

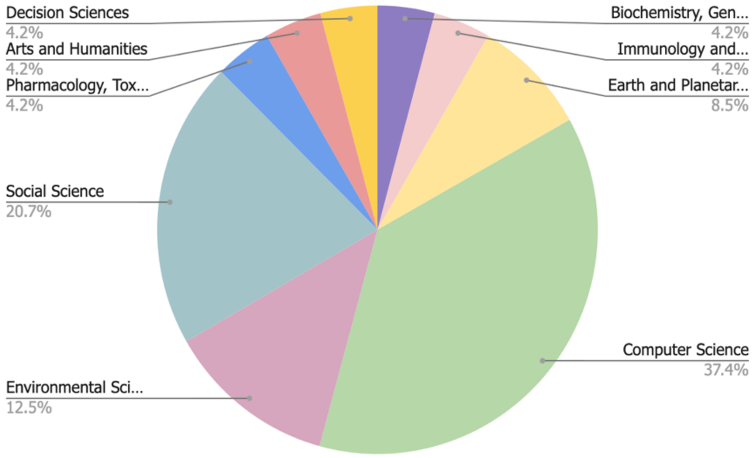


Figure 11. Documents by subject area.

Word Cloud

Visual word clouds in Figure 12 provide a quick view of the most prominent terms from the "Abstract" and "Finding" sections of the 16 articles. As seen in the image generated, the larger the word appears, the more frequently it appears in the text data. The data sample included "Abstract" and "Finding" text data generated by Scopus.

Figure 12. Word Cloud.

Networks Analysis

A visual analysis of the citation connections between 16 articles (see Figure 13) revealed that this research topic of data visualization in Citizen Science was dispersed. In this network analysis visualization, the green dots indicated 16 articles, and the blue dots represented papers that connected 16 articles, not 16 articles themselves. The larger the dots, the more citations. The only direct connections were between Snyder (2017) with Offenhuber (2020), and Sandhaus (2019) with Nicholson-Cole (2005). There were six articles with no connection at all (37.5%), and there were six with indirect connections (37.5%).

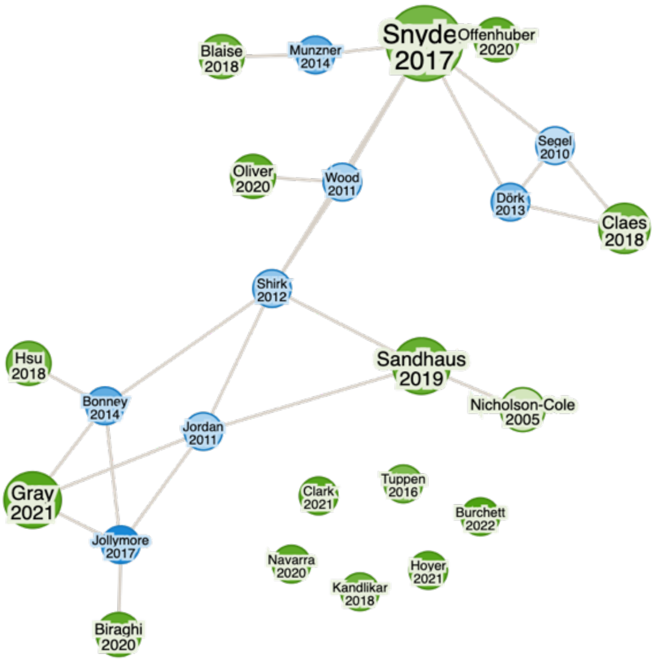


Figure 13. Networks Analysis, *ResearchRabbit, Version 2.0 Human Intelligence Technologies, Incorporated.* (n.d.). https://www.researchrabbitapp.com

Concept Mapping

Although dominant concepts such as "data visualization" or "information visualization (InfoVis)" were identified in the Background chapter, other data visualization-related concepts emerged in the literature review. Concept mapping can be especially useful in identifying blank spots and blind spots in a cross-discipline study. New approaches and emerging trends were identified, as well as critical aspects.

Concepts associated with data type and data quality

The concepts of Volunteered Geographic Information (VGI) visualization and Cartographic visualization, refer to visualizing geospatial data. Geospatial data contains spatial components and georeferenced information. According to Hoyer et al. (2021), many Citizen Science topics have geospatial data which can be locations of animal and plant species, data of the water quality in lakes and oceans, or places with environmental pollution. The study of Biraghi et al. (2020) focuses on the role of VGI visualization in encouraging participation in environmental Citizen Science. An empirical study of Cartographic Visualization focused on the uncertainty issue associated with unsystematic collected CS data (Hoyer et al., 2021). In a case study, Blaise et al. (2018) explored challenges encountered in data modeling and visualization challenges in the collection of geo-related spatio-historical data in the field of minor heritage. Researchers Navarra et al. (2020) developed and examined two different approaches to visualizing VGI.

In contrast to geospatial data, other forms of data, such as audio, were less plentiful but quite complex. Oliver et al. (2020) revealed a number of interaction challenges in connection with audio data in the CS projects. Their study examined how visual transformation could aid the education and engagement of citizen scientists to collect better data. Despite the study's small sample size (12 participants), the results suggested a new approach to analyzing audio data through visualization. Another niche area focuses on analyzing music-bibliographic data regarding public involvement. Tuppen et al. (2016) provided data visualization examples to encourage open access to music-bibliographical datasets to reach out to more scholars and citizen scientists.

When it came to the data quality issue across the literature, the complexity caused by heterogeneity arose regularly as a concern which is discussed in section 4.3.1.1. Furthermore, a discussion about imperfection in minor heritage datasets was raised by Blaise et al. (2018). Imperfection, which was caused by citizens' self-knowledge, played a crucial role in preserving and sharing minor heritage. Contrasting to collecting or decrypting massive datasets, the study reasoned why imperfections should not be reduced in Historical Sciences, rather it was necessary to design visualization solutions that led to better assessments of imperfections.

Concepts associated with a design perspective

Along with the fact that visual practices are deeply entwined with the type of citizen science data, Snyder (2017) provides a theoretical understanding of Vernacular Visualization in relation to non-expert user visualization design (NEUVis). This concept proposes a broader view of visualization in the citizen science context. As part of this concept, vernacular visualizations are created by and for non-experts in order to support data-rich collaborative and coordinated CS projects, distinct from visualizations created by experts. NEUVis practices can positively impact the technology development of data visualization itself through a co-design process. Furthermore, different approaches to involving non-experts directly in visualization design have been discussed in other practices. Gray et al. (2021) emphasize the benefits to engage end-users directly in the design of the data visualization application to enable them to visualize data independently to address personal interests.

Other aspects of visualization solutions in connection to CS, i.e., where visualizations are displayed, a recent paper by Clase et al. (2018) discusses the concept of public visualization, which refers to data visualization exhibited in public spaces with the goal of facilitating social discourse and promoting reflection. Advanced by Internet-of-Things (IoT), data visualization, and interactive polling system enables participants to contribute to a civically motivated co-authoring data-driven narrative.

A number of other theories from different disciplinary sectors have been applied to aid in designing data visualization in CS. For example, Biraghi et al. (2020) are interested in exploring the difference between the meaning of color in absolute terms and it's meaning when applied to a specific context such as water quality. The concern about color has received some attention but has only been briefly addressed in other studies, as well as a focus group study by Hsu et al. (2018). In spite of this fact, in the environmental field, Biraghi et al. (2020) have made the only empirical attempt to use the color meaning test method based on Won and Westland (2017) among the 16 articles. To the best of my knowledge, the color meaning study in data visualization remains underexplored or underresearched in other scientific fields.

Concepts associated with a critical perspective

Offenhuber (2020) explained "Autographic Visualization" as a speculative counter-model to address some limitations of data visualization. Data visualization offered many advantages, including generalizable methods of pattern discovery and computation. However, data visualization was only possible when data was already present. In this way, the material processes and circumstances of data collection remained largely hidden (Offenhuber, 2020). In addition, through an exploratory qualitative study, Nicholson-Cole (2005) criticized the role of visualization in communicating climate change in terms of validity, subjectivity, and interpretation. Even though Trumbo (1999) argues "contemporary science communication relies on visual representation to clarify data, illustrate concepts, and engage a public informed through an ever-increasing arsenal of computer graphics and new media tools" (p. 421). However, in line with Stamm et al. (2000), Nicholson-Cole (2005) declares that people's perceptions and the nature of climate change make it challenging to communicate a motivating message visually. In the article, Nicholson-Cole (2005) analyzes the factors that impact data visualization in climate change and provides practical recommendations for its implementation.

4.2 From a Technical Perspective (RQ2)

RQ2: What are the technical aspects of the applications used to visualize data?

In line with RQ2, three aspects will be examed in this section: 1) the application's primary technical features, including data type, application type (web-based or/and mobile application), and open-source; 2) application goals; and 3) usability evaluation, including what, and on whom was the evaluation conducted.

4.2.1 Primary Technical Features

Data type

There were seven studies (7/9) related to the environmental datasets, one study focused on cosmological simulation datasets, and one study on Minor Heritage. Environmental datasets were the focus, two investigated water monitoring data, and the other five contributed one study each in four different domains: climate-related, marine debris, environmental eDNA, biodiversity, and environmental sensing and public health. As compared with the findings of concept mapping to RQ1, four articles reported VGI data access through different disciplines.

Web-based or/and Mobile Application

In all nine studies, interactive web-based applications were reported, two of which also mentioned mobile versions (see Table 8). The study by Abramov et al. (2022) was the only one that discussed why they considered web-based to be the best solution. It was time-consuming and complex to visualize simulation datasets, according to Abramov et al. (2022). As a result, CosmoVis was developed as an interactive web application due to 1) minimize disk space and memory requirements, 2) maximize accessibility across operating systems, and 3) eliminate the need to compile or install software locally because of its massive disk space and memory requirements (Abramov et al., 2022).

Table 8
Primary Technical Features of Data Visualization in Citizen Science.

Article	Data Type	Application Type/Name	Open Source	Visualization Goal	Target Audience
(Blaise et al., 2019)	Social Historical Minor Heritage	Web-based	Open source	Analysis	Non-expert and Experts
(Hoyer et al., 2021)	Environmental Biodiversity	Web-based, Biodiversity project	Partly	Analysis	Non-experts
(Navarra et al., 2020)	Environmental Climate	Web-based, CitizenSensing Project	Open source	Analysis and Communication	Non-expert and Experts
(Biraghi et al., 2020)	Environmental lake monitoring	Web-based, mobile application SIMILE Project	Open source	Analysis and Communication	Non-expert and Experts
(Hsu et al., 2018)	Environmental sensing and public health	Web-based, EHC	Partly	Analysis and Communication	Non-expert
(Clark et al., 2021)	Environmental marine debris	Web-based, Mobile application	Partly Will be	Analysis	Non-experts
(Abramov et al., 2022)	Cosmological simulation	Web-based CosmoVis	Open source	Analysis	Non-expert and Experts
(Gray et al., 2021)	Environmental water monitoring	Web-based CWDAT	Open source	Analysis and Communication	Non-expert and Experts
(Kandlikar et al., 2018)	Environmental eDNA	Web-based Ranacapa	will be	Analysis	Non-expert and Experts

Open source

There were five (5/9) studies reporting open-source data visualization applications (Navarra et al., 2021; Biraghi et al., 2020; Blaise et al., 2019; Abramov et al., 2022; Gray et al., 2021). Three (3/9) reported it as restricted to a limited audience due to i.e., data use agreements (Clark et al., 2021). It has been stated by two articles (Kandlikar et al.,2018; Clark et al., 2021) that the application may become open source in the future. Kandlikar et al. (2018) stated that it would become an open source for undergraduate education in the future.

4.2.2 Application Goals

Data visualization applications can be classified into categories for analysis and communication based on the application goals discussed in the Background section. All studies (9/9) reported visualizing data for analysis, while three of them (3/9) addressed both analysis and communication as purpose. Examining the application's goals in detail, many of the application goals were research-oriented, including making citizen-contributed data explorable (Hsu et al., 2018;

Navarra et al., 2020) such as patterns (Navarra et al., 2020), and investigating a range of scientific questions (Abramov et al., 2022; Kandlikar et al., 2018). Several of the goals were integrated into Citizen Science, including encouraging Citizen Science participation (Biraghi et al., 2020); keeping the public informed and updated, and communicating and disseminating evidence-based results (Hsu et al., 2018; Blaise et al., 2018); learning and promoting citizen scientists' sustainable behaviors (Biraghi et al., 2020); empowering scientific evidence-based action and decision-making (Hsu et al., 2018).

4.2.3 Usability Evaluation

4.2.3.1 What was evaluated?

Only six articles reported the evaluation process (Blaise et al., 2019; Hoyer et al., 2021; Biraghi et al., 2020; Abramov et al., 2022; Gray et al., 2021; Kandlikar et al., 2018). Among the six, one study (Kandlikar et al., 2018) discussed the results of the assessment alone without going into sufficient detail. Taking a closer look at the literature, even though the evaluation was discussed in light of the six studies, the reason for a certain assessment method was often unfounded. Furthermore, all six studies used different criteria. Moreover, the other three articles do not discuss evaluation (Navarra et al., 2020; Hsu et al., 2018; Clark et al., 2021), rather suggesting evaluation as a further study.

A key finding of Hoyer et al. (2021)'s empirical studies on the gradation of reporting activity have discovered that citizen scientists who report well do not automatically have better experiences using ornithological data and visualizing them. Further, NASA Task Load Index, a subjective, multidimensional assessment instrument, was used to assess perceptions of task workload (p. 157). Biraghi et al. (2020) presented and validated a mixture of Usability Evaluation Methods (UEMs) such as heuristics and field tests to test issues such as recognizability, learnability and memorability, efficiency, and others (p. 243). In another water monitoring data visualization study, Gray et al. (2021) evaluated the intuitiveness of the interface, its relevance to the users' questions, and the generation of actionable information (p. 10). Abramov et al. (2022) provided evaluation results from an open-ended study with detailed qualitative feedback from eight researchers recruited from the workshops (p. 2922). In a study by Blaise et al. (2018), a set of eight criteria were tested: 1) readability assessment, 2) knowledge communication; 3) reasoning and hypothesis generation; 4) adoption and reuse; 5) user guidance; 6) legibility; 7) adaptability; 8) consistency.

4.2.3.2 Who was evaluated?

Another important question about the evaluation of usability is to evaluate whom. In studies of Hsu et al. (2018) and Abramova et al. (2022), the evaluation population was limited to only experts, although the application was also intended for both scientists and citizen scientists. In the other studies, the way to define the application tested on whom remains unclear. The end-users term was used in the study by Gray et al. (2021), which referred to anyone who could use the application, regardless of whether they were involved in the CS project. End-users could be interpreted as non-experts and/or experts. In another study by Biraghi et al. (2020), they tested the application with "people at the very first experience with the app and no preliminary introduction" (p. 243), which could be interpreted as they may not be the end-user. Moreover, according to Blaise et al. (2019), their empirical study evaluated the visualization by the participants who had no background knowledge of heritage sciences, which caused inconclusive results. Instead, they evaluated direct feedback from information providers, which led to success in detecting misinterpretation (p. 19). Similarly, the definition of no background knowledge in heritage sciences in the study of Blaise et al. (2019), was also not necessarily the end-users mentioned in Gray et al. (2021).

4.3 Opportunities and Challenges (RQ3)

RQ3: What are the challenges and opportunities of applying data visualization to citizen science?

In this section, opportunities and challenges associated with applying data visualization in a Citizen Science context have been discussed in both the technical and the social domains among 16 articles (see Table 6).

4.3.1 Opportunities and Challenges in the technical domain

4.3.1.1 Data Quality

Heterogeneity

Data quality has been identified as a key barrier in the literature (e.g., by Hoyer et al., 2021; Blaise et al., 2018; Gray et al., 2021). In Social Science, data heterogeneity (high variety of data) as well as the integration of heterogeneous data, was identified as the main challenge (Tuppen et al., 2016; Blaise et al., 2018). Tuppen et al. (2016) demonstrated how music-bibliographical data could be analyzed and visualized, and how scholars and citizen scientists could engage with it through hackathons, large-scale data analyses, and database construction (p. 87). In order to engage in the analysis and visualization of music-bibliographical data, researchers urged making datasets more accessible. Another article discussed another challenge caused by the data heterogeneity was the imperfection in minor heritage data. Blaise et al. (2018) described that minor heritage data was "personal memories, records of individual experiences, self-knowledge about a time gone by" and such data were imperfect, they were "detailed yet less precise, often subjective, and unverifiable" (p. 13). Blaise et al. (2018) argued that the imperfection of "uncovering personal, individual self-knowledge" was crucial to the understanding of minor heritage and it was not a good idea to reduce these imperfections in Historical Sciences. According to the researchers, designing visualizations for imperfect minor heritage datasets was particularly challenging. They claimed that although there were efforts made in investigating visualizing such data, there was still significant work to be done in adapting methods of collecting heterogeneous data to provide “spatial, temporal, and thematic knowledge.” (Blaise et al., 2018, p. 11)

Incomparability

In Environmental Science, the incomparable characteristics of citizen scientists' data present another challenge (e.g., Hsu et al., 2018; Clark et al., 2021; Gray et al., 2021). State and non-state professionals were unable to retrieve information, visualize trends, and share findings due to the incompatibility of data stored in multiple incompatible systems (Hsu et al., 2018). It was noted by Clark et al. (2021) that a lack of coordination and harmonization in data collection limited comparability between groups, which hindered state and non-state actors from developing solutions at scale. In Gray et al. (2021), the possibility of using an open-source application approach was explored to allow users to check their data against an accepted reference to ensure quality and reliability before submitting it. For established community-based users, this capability was important for reviewing their data before it is submitted to a larger project database, as well as for training purposes where new participants can compare their observations to historical or regional norms.” (p. 8). However, it has been questioned whether open source could successfully address this challenge in other data types. By using open-source tools, Blaise et al. (2018) sought to answer whether an open-source solution could be reliant on, and if so, how to visualize the imperfect data efficiently (p. 12). Despite concerns, there was not enough evidence or discussion in the literature to assess the risks and benefits of using open-source applications.

Uncertainty

Spatial features gained concern in CS data (e.g., Blaise et al., 2018; Hoyer et al., 2021), and uncertainties could appear in any spatial dataset regardless of the CS subject. According to Hoyer et al. (2021), both Zhang and Goodchild (2002), and the study of Kinkeldey et al. (2017) have alarmed that the uncertainty could result in a negative effect on decision-making processes and trust in results. Uncertainty challenges made it difficult to recognize the limits of the validity of presented data and reduced data transparency. They stated that it was possibly due to a lack of communication with visualization experts, programming skills, or resources (p.159).

4.3.1.2 Data Privacy

The literature identified a privacy challenge, and some solutions were discussed. According to Gray et al. (2021), end users had concerns about privacy when communicating/sharing data via visualization. Hsu et al. (2018) raised data privacy issues when using data visualization to share results and communicate with the public. Personal stories and images were displayed in a heatmap visualization to give insight into how people experience oil and gas exposure. Participants' privacy was protected by de-identifying data and preventing re-identification by Hsu et al. (2018, p. 3). Another study by Navarra et al. (2020) restricted the submission of collected data to geographical locations only, assuming that the user enabled retrieval of this information. Additionally, Navarra et al. (2020) addressed the issue of data privacy in an iterative design process. In the first version of the web portal, personal information was not displayed due to data privacy concerns. In turn, the second version was password-protected and displayed personal data only for analysis and evaluation (p. 3).

4.3.1.3 Applicability

Several researchers have noted that developing well-designed data visualization tools requires a significant amount of time, human and financial resources. As a result of resource constraints, Hsu et al., (2018) reported that CS projects or organizations typically could not develop visualization tools independently to support project goals like decision-making (p. 2). It is possible, therefore, that well-designed data visualization applications should be developed efficiently. The articles mention that in the design process, they studied the relevant research first to see if there were useful related studies or even existing applications that aligned with their CS project goals (Abramov et al., 2022; Hoyer et al., 2021; Clases et al., 2018; Navarra et al., 2020; Biraghi et al., 2020; Gray et al., 2021; Kandlikar et al.,2018). In some cases, it may not be necessary to develop from scratch. In analyzing and visualizing heterogeneous music-bibliographical datasets, for example, Tuppen et al. (2016) found both free and commercial tools could be used. They introduced three free tools, Google Fusion Tables, OpenHeatMap, and Palladio, that were well suited to their requirements and easy to use, no programming knowledge was needed. CS projects may also require the development of their own applications based on their specific goals and limitations. According to Gray et al. (2021), a number of tools have been developed to support the analysis and visualization of water monitoring data. Most tools, however, were not accessible due to the cost, system requirements, program requirements, and/ or typically designed for specific protocols, and/or formatted input data that was not applicable to citizen scientist collected data (Gray et al., 2021). Abramov et al. (2022) decided to develop a new application since certain objectives could not be met with the existing applications. In addition, they noted that the high requirement for prerequisite skills to use an existing application was also considered a barrier.

Furthermore, Citizen Science data visualization presents additional challenges due to its subject- and application-specific nature. However, some studies have considered and examined the possibility of applying data visualization practices and the conditions under which they can be applied. In CS, four models or frameworks for data visualization practices have been identified in the literature: 1) visualizing audio; 2) Source-to-Sea (S2S) management for political decision-making; 3) a *Gardenroots* model for grassroots communities; and 4) the COASST model for vernacular visualization practices. The following section examines the applicability of these four data visualization practices.

Visualizing audio

Visualizing bird audio data, Oliver et al. (2020) explored citizen science interactions with it. Interacting with a variety of media visually and acoustically had several benefits (p. 1687). Design lessons were offered to bridge the knowledge gap and promote long-term engagement in audio-focused CS projects. As part of the design process for bridging the knowledge and content when interacting and reviewing the audio, the following elements were necessary according to Oliver et al. (2020): "1) media salience; 2) accessible associations; and 3) species information & perceptions. To facilitate long-term engagement, they included 1) Growth of knowledge with purpose, 2) Diversity in Task Difficulty, 3) Customizable Complexity, and 4) Collaborative Puzzle Solving" (p. 1694f).

Source-to-Sea (S2S) management for political decision-making

Using the Source-to-Sea (S2S) framework, Clark et al. (2021) provided a design structure for the Australian Marine Debris Initiative (AMDI) visualization tool. By incorporating these five "source-to-sea" steps, users could characterize, engage, diagnose, design, act, and adapt when addressing leading environmental concerns (p. 211). Despite the fact that the tool has not been empirically evaluated in this study, Clark et al. (2021) provided examples of its application to state policy implementation, monitoring, and reporting, as well as to local non-governmental organizations monitoring and reporting. Although there was a lack of evidence on the feasibility of the applicability of the design structure, CS projects for political decision-making may be able to benefit from considering the applicability of this design structure in their design thinking.

Gardenroots model for grassroots community

Gardenroots was a collaborative PPSR (public participation in scientific research) in environmental health and informal science education. The Gardenroots model demonstrated that community-first reporting and effective data visualizations combined in a collaborative citizen science project enable participants to make personalized decisions about their risk and develop individual prevention and intervention strategies based on their own self-efficacy and capacity (Sandhaus et al., 2019, p. 56). These design structures were applicable to all populations, particularly underserved ones.

COASST model for vernacular visualization practices

A vernacular design aspect expanded the boundaries of data visualization. According to Snyder (2017), the vernacular version of visualization developed by and for non-expert users was more useful than a version created by an expert for a specific audience. An example of vernacular visualization practices in data-rich collaborative and coordinated CS contexts is the University of Washington's Coastal Observation and Seabird Survey Team (COASST). The COASS (see Table 6) Field Guide ensured high data quality by providing hands-on visual support for non-experts. Researchers in the project usually created computer-based data visualizations in-house using R, Inkscape, and PowerPoint after the collected data converting to abstract data (Snyder, 2017, p. 2102). As part of the analysis process, data visualization was used to inform and update the scientific community. The collaborative vernacular design approach discovered that different data visualization forms offered different opportunities to present information to the viewer. The collaboration between COASST staff and designers collaborated to ensure that the visual design and content were correct and consistent (Snyder, 2017, p. 2105). Coordination, communication, and data infrastructure were supported by vernacular visualization practices in the COASST project.

Applicable design strategies

Adopting the entire existing design solution or framework may not be possible, but parts of it may work within the special requirements of the application goals. The results of an empirical study adopted color theory techniques to investigate color meaning in the environmental citizen science domains (Biraghi et al., 2020). They examined the difference between preserved and absolute meaning by using the color theory methods of Won and Westland (2017). Red was found to have a negative acceptance, which aligned with the WFD (the international Water Framework Directive) but was not evident from the investigation of absolute color meaning alone (Biraghi et al., 2020, p. 239).

There were other design strategies that were suggested to be adopted for a wide range of use. Besides the implementation of color theory, Biraghi et al. (2020) offered the visuospatial designer a design strategy of a map view visualization which was proved by positive results through various UEMs. The researcher identified the content overload might be interpreted by participants as that data collection was unnecessary, thereby discouraging contributions. Thus, “expiration time and map limit density” were proposed as solutions for controlling data quantity and encouraging participation (Biraghi et al., 2020, p. 243). Also worth noting, the two design strategies were documented and evaluated in detail and could therefore be easily adopted by other researchers with similar needs.

4.3.2 Opportunities and Challenges in the social domain

4.3.2.1 Knowing the target audience

It has been found that identifying the target audience is particularly crucial, but it remains challenging due to limited resources. As presented in the Background section, visualizations can provoke different reactions in accordance with “background, previous knowledge, motivations, and personal beliefs”, so it is imperative to understand the target audience to ensure that visualizations were accurate, reliable, and understandable (Sandhaus et al., 2019, p. 56). Sandhaus et al. (2019) conluded frequently asked questions based on Nicholson-Cole (2005), which included “*learn who the intended audience is, what their needs are, and how they will perceive the data*” (p. 56).

Scientists and non-expert citizen scientists were often identified as the target audience. In the humanities, Tuppen et al. (2016) investigated how music-bibliographical data could be analyzed and visualized with the aid of technology, especially for scientists. Their study showed that data visualization allowed scientists to identify potential points of interest at a macro level, then examined the data more closely at a micro level. Furthermore, in another study in the humanities, Blasise et al. (2018) tested the visualizations by the wrong audience who "had no background knowledge or involvement in heritage sciences at large" and it led to unreliable results in evaluation. So, they decided to test the "information providers" instead, to prevent misinterpretations when visualizing minor heritage data, rather than scientists (p. 19).

Even so, the target audience differed not just between scientists and citizen scientists, but also between citizens from different backgrounds, which presented a challenge. The COASST project (see Table 6) made assumptions about the different needs of potential participants based on their 17-year experience (Snyder, 2017, p. 2102). Their method might not be so helpful for a newly established CS project. As a result, the researcher asked, "how can vernacular designers advance their practices without assuming they need professional skills or training" (Snyder, 2017, p. 2107). To address this issue, the researcher urged expanding the evaluation technology in vernacular visualization so that the intended audience could be learned.

The reviewed literature, however, showed that very few studies have attempted to identify a target audience, although some insights have been provided. Using participants' roles and motivations as relevant criteria, Gray et al. (2021) developed their data visualization. Thanks to the identification of different roles and motivations, they detected the contract feedback for default templates and settings in the interactive data visualization interface. The researchers concluded that it might feel restrictive to advanced users, but overwhelming to those with less knowledge of technology. Based on their findings, an independent visualization approach has been proposed as a possible solution to aid efficient communication. However, there was another problem here, since it was not always clear which criteria were relevant. The experiments conducted by Hoyer et al. (2021) provided some insight into this issue. Their study examined whether a low or high level of reporting activity helped the relevant target group interpret dissemination and visualization results. This study found that a more active reporting behavior may not automatically result in better interpretations of the data visualization and the reason for this remained unclear (Hoyer et al., 2021, p. 167). In other words, it may not be necessary to consider reporting behavior of the participants when evaluating the visualization.

4.3.2.2 Data visualization as an aid for communication and learning

Using data visualization in formal and informal education

It has been shown that data visualization applications can be used as a teaching tool in Citizen Science in a variety of studies. In informal education, visualizations were used by experienced team members to teach students about "the mechanics of data representation, pointing out specific visual features in order to flag potential errors or limitations in statistical models." (Snyder, 2017, p. 2107). Empirically, Kankliar et al. (2018) examined the use of data visualization tools in formal education. The results revealed positive evidence for using such a tool to reduce the time and difficulty associated with visualizing biodiversity patterns and help students better understand data. Professional scientists believed that CosVis could be an effective tool for disseminating scientific research findings and they also expressed interest in using the application as a teaching tool in various (both formal and informal) educational contexts or as an outreach tool to the public (Abramov et al., 2022).

Training and learning materials for data visualization

Synder (2017)'s concept of a wider view of visualization was applied to examine the training and learning materials for supporting data visualization in a citizen science setting. In line with Keim et al. (2008), two aspects were discussed: for understanding and analyzing/visualizing data, and for understanding the analyses/visualization (as cited in Blaise et al., 2018, p. 19).

- Support for understanding and analyzing/visualizing data

Educating the non-technical audience to visualize and analyze results independently is very challenging, not only for professional scientists but also for non-expert scientists (Kandlikar et al., 2018, p. 3). Based on what has been discussed before, by Kloetzer et al. (2021), diverse audiences should be considered when designing and developing the training to increase participants' scientific competence, support, and ensure the intended outcomes of Citizen Science projects (p. 302). Scientists and citizens face different challenges during training. According to a study by Abramove et al. (2022), detailed documentation was required to assist scientists in learning the data visualization tool. However, when facing the non-expert scientist, it was particularly crucial to change the specific science language to plain language to be understandable to improve communicative clarity (Snyder, 2017; Gray et al., 2021; Sandhause et al., 2019). To grassroots community members, it was suggested to provide informal explanations which differed from the wider scientific community (Gray et al., 2021).

Aside from text support, various forms of visual support were also discussed in the literature. Gray et al. (2021) noted that there was a preference for "graphs, and maps using color to spatially display water quality parameters, their values, and associated criteria." (p. 11). According to Snyder (2017), vernacular designers combined different types of visual representation supported by plain explanatory text**,** such as icons, photographs, and verbal descriptions of visual features (p. 208) to guide the collection of high-quality data and ensure data validity. Photographs helped to stabilize a transition from physical observation to abstracted data, as well as enable experts to provide feedback to citizen scientists regarding the quality and accuracy of their observations. The use of multimedia such as blogs, E-newsletters with photographs was an effective way to reach out to a broader audience.

In addition, Gray et al. (2021) suggested using the application itself to train for assessing data validity and reliability. The function of allowing less experienced citizen scientists to compare their collected data to an accepted reference could be seen as an important issue for training purposes (Gray et al., 2021). Kandlikar et al. (2018) suggest providing brief explanations and links to additional educational resources. Yet, it should be noted that extra support material should be subject-specific. As an example, the solutions suggested in observation CS projects don't fit simultaneous data. Astrophysicists suggested video tutorials in the study of Abramove et al. (2022), but there was no explanation for their recommendation. However, Hoyer et al. (2020) showed empirical evidence that a video tutorial could enhance understanding of data visualization. Challenges were also reported such as being less mentally demanding and exerting less effort by providing video tutorials. According to Hoyer et al. (2020), external supporting material made maps more understandable when multiple layers of information are combined. Their results highlighted the challenge to understand the "missing presence data" caused by the incompleteness of CS data (p. 155). An example of "the missing presence data" is that: if a bird is not reported in an area, either it can be interpreted as the bird does not exist in that area, or it can be understood as there are no observations available. If missing presence data (either the bird doesn't exist, or no bird observation is available) isn't explained clearly, it may lead to misunderstandings owing to the lack of transparency in the way the data visualization is presented. Based on the NASA-TLX (Task Load Index) method, Hoyer et al. (2021) found that the additional video tutorial could not resolve the transparency issue caused by the missing presence data. Their result suggested that regarding tutorials and supporting materials, it was crucial to explain not only the visualization and interpretation of the data but also "where the limits of validity of the data depicted lie" in the citizen science context (p. 169).

- Support for understanding the analyses/visualization

As well as training and supporting materials discussed previously, Kloetzer et al. (2021), Tuppen et al. (2016) and Sandhaus et al. (2019) suggested providing individualized solutions for a diverse audience. It was worth noticing that individualized data visualizations referred to as visualization were generated by experts for non-experts, which differed from what Gray et al., (2021) have discussed that users analyzed data independently for their own. An individualized visualization that was well-designed and appropriate for a variety of audiences required careful consideration and design. How best to communicate the scope of complex information to fit the needs of a diverse audience and maintain data validity was seen as one of the biggest challenges (eg., Sandhaus et al., 2019; Snyder, 2017). Sandhaus et al. (2019) called for sharing project results in a personalized visualization and suggested simplifying the figures and only showing pertinent information within participants' expectations and relevant to the participants. Furthermore, even though they emphasized the time-consuming nature of a personalized approach, they also highlighted how crucial and worthwhile it is.

4.3.2.3 The impact of collaborative design on the social domain

The literature identified two collaborative approaches focused on design for the social impact of Citizen Science data visualization practices, open-ended design, and co-authoring. Besides, there is another a co-design approach (*COASST model by* Snyder in 2017) that has been described as a framework in section 4.3.1.3.

Open-ended design in developing open-source applications

An open-end design approach has been investigated and evaluated by Gray et al. (2021) in aiding independent data analysis and visualization of water quality data. By using an open-source language (R) through Github, the visualization tool CWDAT can be modified and customized as needed by end-users Gray et al. (2021). The ability to independently analyze and visualize the end users' own data and compare it to reference datasets (as discussed in RQ2), could enable the end users to better understand the quality of their own data and therefore improve it. Being involved in design directly, citizen scientists had the opportunity to ask questions on their own, as well as learn about potential questions they were unaware of. They would also be able to better understand and utilize their own data independently or as part of collaborative efforts. These social dimension concerns could promote independent development for individuals and two-way communication between diverse audiences. According to Gray et al. (2021), the customized and independent features could address data quality limitations and bridge the gap between non-experts and scientists. They also pointed out that these design features were critical in data analysis and visualization tools that support decision-making. As a result, the application could be considered as a solution that would expand the range of capabilities and applicability of Environmental Decision Support Systems (EDSS).

Co-authoring CS narrative in public visualization with polling interaction

Through a public visualization and polling systems approach, Claes et al. (2018) discovered the tactic of social relationships between participants in a co-design approach promoting both opportunities and challenges in participatory engagement, actionable behavior, and distribution potential. Based on their study, social relationships encourage reflection more than the data trends and patterns in the visualizations themselves (p. 8). Unlike the study of Gray et al. (2021), which focused on co-designing the application itself, this co-authoring approach focused on encouraging participants involved in the CS project by reflecting and acting on the collected data. Participants in a physical public environment could be engaged by displaying visualization with polling interaction. Therefore, they advocated taking social dimensions into account in distributing a public visualization. Detailed design considerations were provided that could be used for "crowdsourced, bottom-up public visualization of civic concerns evidenced in data on distributed displays." (Claes et al., 2018, p. 7). It should be noted that this study has been conducted over a limited length of time, so it does not provide long-term effects.

5 Discussion

5.1 Discussion of RQ1

CS projects facilitated by data visualization technology

The first research question in this study sought to determine an overview of existing studies of what CS projects have been facilitated by data visualization technology. It was evident from the subject area and network analysis results that studies covered a wide range of domains and were cross-disciplinary in nature. There is a possibility that citizen science alone is increasingly being used in different fields of science, as mentioned in the Background section. In addition, data visualization technology applied in CS has been studied across the boundaries of technology and citizen science, which may contribute to understanding its cross-disciplinary nature. What is surprising is that the result of the network analysis has shown very few network connections among 16 samples. This result could be explained by several factors. It may be due to the fact that the research in this emerging field is still in its early stage. It may also be a lack of collaboration in this specific field, or it is determined by the subject-specific and application-specific nature of the research. As discussed in the Background section, these results confirm the need for an investigation into data visualization technology in citizen science.

Concept mapping

A concept mapping process resulted in various concepts covering CS-related visualization issues that were not discussed in the Background section. The implication is that, beyond a technical perspective, a citizen science context can add special meaning to technology. The most frequently studied in the literature is Volunteered Geographic Information (VGI) visualization which referred to visualizing geospatial data. This result may be explained by the fact that the vast majority of Citizen Science observation data are geospatial in nature. In addition to geospatial data, other types of data, such as audio, and music-bibliographic data, were also discussed.

The field has also practiced other emerging trends, such as collaborative design strategies. An interesting finding was that non-expert user visualization design (NEUVis) thinking as a co-design approach was applied to design visualization for and by the non-expert (Snyder, 2017). It is possible that the visualization concept in CS may extend beyond the visualization of abstract data to provide visual support throughout the entire project lifecycle. Another interesting finding about collaborative design was that end-users were directly involved in the application design through open-source solutions, visualizing data independently based on their preferences.

As well, the concept of "Public Visualization" suggested a civically motivated approach to co-authoring data-driven narratives (Claes et al., 2018). Across the literature, heterogeneity, uncertainty, unsystematic, and incomparable were frequently cited as concerns about data quality.

Among the most important findings was that the literature featured critical voices. By looking at research from a critical perspective, it may be possible to gain a deeper understanding. Offenhuber (2020) extended the scope of visualization beyond symbolic data to understand how material traces were perceived, and how these perceptions related to individual knowledge and skills. By considering the structures in the physical world as a form of data, autographic visualization was speculatively opposed to data visualization, which was confined to symbolic representations (Offenhuber, 2020, p. 106). The researcher suggested a combined model could combine both InfoVis and autographic visualization to track data as well as to display it graphically. Using this model requires the presence of visible material traces which can explain the phenomenon. As a result, it would not work in the context of audio visualization in Oliver et al. (2020) since there are no visible material traces. Although autographic visualization lacks agility, versatility, and scalability compared to computational analysis (Offenhuber, 2020), in the presence of the material traces available, a raw trace may elicit greater attention and trust in citizen scientists' data than a well-designed chart, and thereby more likely to evoke behavior change and decision-making.

The validity, subjectivity, and interpretation of visualization have been criticized in an exploratory qualitative study by Nicholson-Cole (2005). The findings asserted that merely designing an effective visualization was insufficient, suggesting considering how viewers receive visualization was essential. The following key questions should be raised when utilizing visualizations:

> *What kind of images portrays the appropriate level of scientific information? What choices are made by the scientist to develop or select an appropriate visual representation? How is the audience influenced? Will these images change behavior?* (Nicholson-Cole, 2005, p. 269)

The finding has important implications for how people perceive climate change when communicating it to them. This perspective and these questions are likely to be particularly relevant when a CS project focuses on public engagement. The insights gained from this work can be applied to further research in the field of climate change visualization and similar contexts focused on public engagement and action.

As a result, importantly, merged with the summative findings of the concept mapping, the illustration of the "data visualization feedback loop in Citizen Science" (Figure 14) is formed based on the "from data to wisdom" flow presented by Mazza (2009). "Data visualization feedback loop in Citizen Science" illustrates how data to knowledge flows in loops, supporting a more comprehensive understanding of what data visualization means in the context of Citizen Science. Through feedback loops, data visualization technology assists data to knowledge flows in citizen science projects. The arrows in Figure 14 present where the data visualization takes place throughout the citizen science project. According to the literature, data visualization can take a variety of forms in the feedback loop. As an example, the autographic visualization based on the material traces is performed from phenomenon to information loop. Besides, the wisdom feedback to data process using a public data visualization approach presents the wisdom obtained and can also assist in improving data collection by identifying data gaps. This improvement in data collection methods, therefore, advances the investigation a step forward. A variety of visual forms are available in the feedback loop throughout Citizen Science projects. Data visualization can go beyond technical efficiency, beyond visualizing abstract data, beyond symbolic data, and instead combine the material traces, to make sense of processed abstract data more easily to be understood. Additionally, how viewers perceive, and how their individual experience and pre-knowledge relate to the interpretation of visualization are also vital factors to concern when considering what scope of data needs to be visualized, and in what way.

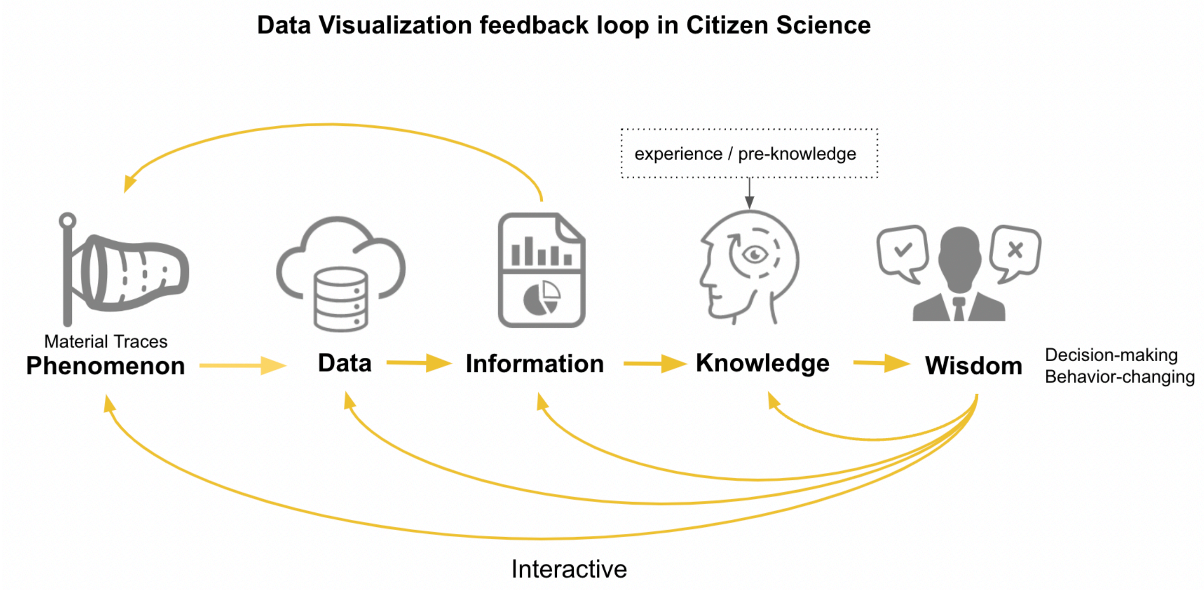


Figure 14. Data visualization feedback loop in Citizen Science.

5.2 Discussion of RQ2

The second question in this research was to identify data visualization applications in CS from a technical aspect. Nine (9/16) articles with diverse subject domains were grouped together due to their focus on technical features in applications. However, other articles also address their technical concerns, albeit in a less focused way. The other seven articles were not included in the meta-analysis to answer RQ2 based on the consideration of efficiency and reliability. There is a possibility that another researcher would make a different decision and produce a different result.

Primary technical features

Environmental datasets were significantly emphasized when it come to data type features. Among the nine studies (see Table 8), seven (7/9) dealt with environmental datasets, one with cosmological simulation data, and one with Minor Heritage datasets. Additionally, four articles reported on studies on VGI visualization (discussed in RQ1) across a variety of scientific disciplines. Furthermore, all nine data visualization applications in nine studies (100%) were interactive web-based, and two of them also mentioned mobile applications. This result further supports the idea of Heeker et al. (2018), who suggested that easy-to-use websites, mobile apps, and those designed with different user groups in mind will be more likely to attract wider participation in communication in CS (p. 6). However, only Abramov et al. (2022) provided insights into why a web-based solution was considered to be the best in the literature. there was no discussion of potential risks or challenges associated with using web-based and/ or mobile solutions.

Additionally, another important finding was obvious to see a trend emerging in the development of data visualizations in CS using open-source solutions. The present study raises the possibility that using open-source solutions can promote collaborative design, which is crucial for citizen science. On the other hand, the collaborative potential in developing applications might require extra resources, such as high technical requirements for human resources. However, it is worth noticing that several articles failed to differentiate between *open-source* applications and *open data* access, making it difficult to determine whether an application is open-source. The author has made a judgment call.

Application goals

All studies (9/9) reported visualizing data for analysis purposes, while four of them (4/9) addressed both analysis and communication as purpose (see Table 8). Examining the application's goals in detail, it was interesting to note the visualization goals were integrated in the line with the goals of the Citizen Science project. Many of the application goals were research-oriented, including making citizen-contributed data explorable such as patterns and investigating a range of scientific questions. Meanwhile, several of the goals were incorporated from a social perspective, including encouraging Citizen Science participation, keeping the public informed and updated, communicating, and disseminating evidence-based results; learning and promoting citizen scientists' sustainable behaviors; empowering scientific evidence-based action and decision-making.

As shown in Figure 15, a prior study by Kloetzer et al. (2021) has illustrated the scientists' and citizen scientists' learning in citizen science projects, how they learn, and what they learn (p. 299). It is interesting to see that the application goals identified in the literature are closely connected to these phenomena describing scientists' and citizen scientists' learning. That is to say, the goals of visualizing data for analysis and communication in CS are closely related to what the scientists and citizen scientists learn as well as how they learn. Perhaps this is because visualizing data for analysis and communication is a learning-oriented activity in the citizen science context. It can therefore be assumed that employing data visualization is part of the learning process in CS.

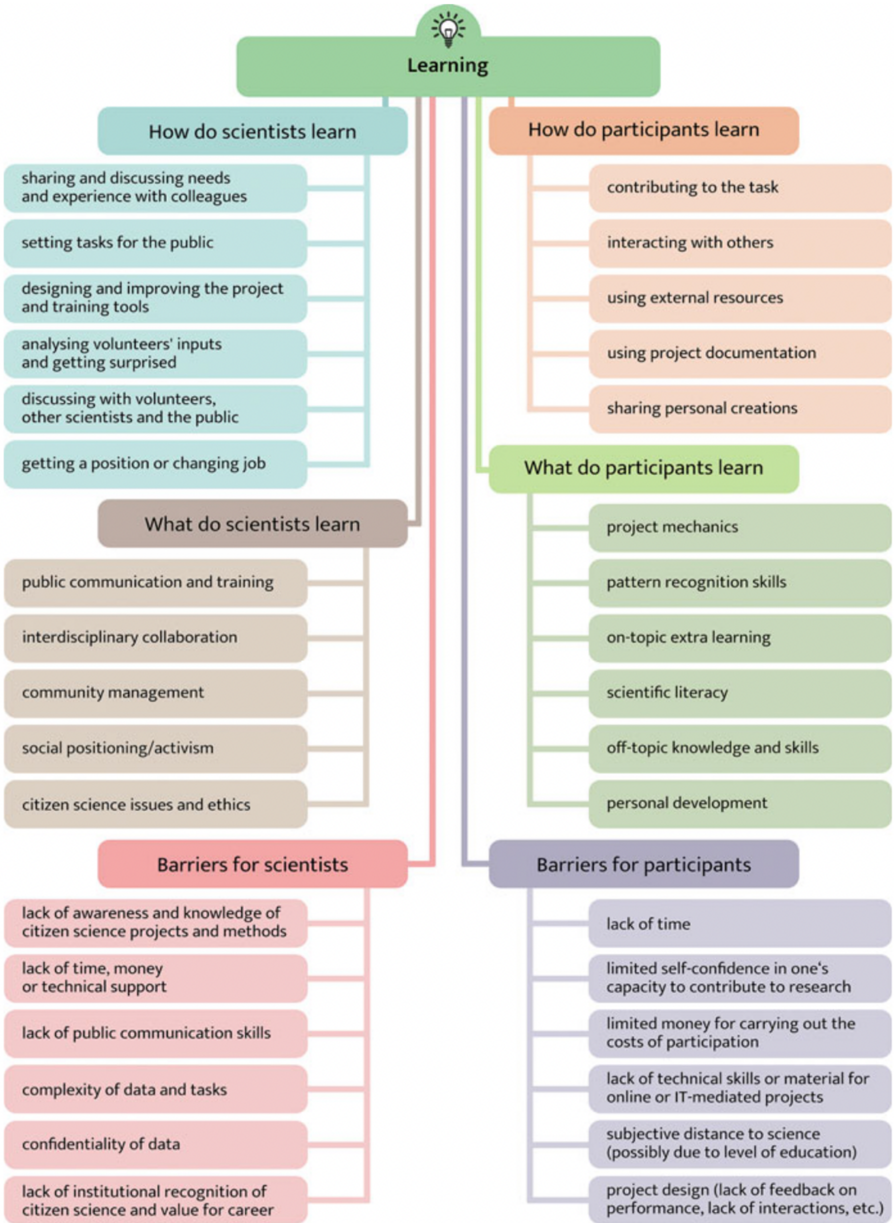


Figure 15. Extended thematic map of volunteers' learning, Source (Kloetzer et al., 2021, p. 299)

Usability evaluation

In the visualization literature, evaluating visualizations is becoming increasingly important, according to Tory and Moller (2005). It is crucial to conduct user studies of perceptual phenomena and compare tools to determine whether visualizations are useful for people working in real life

(p. 8). Six (6/9) articles evaluated usability in the literature. Taking a closer look at the literature, the reason for a certain assessment method was often not clearly stated among the six articles. According to the findings, there was a lack of proper evaluation of usability when visualizing data in CS. Furthermore, the most obvious finding to emerge from the analysis was the diversity in both what criteria and techniques to choose and which target audience should be evaluated. This result indicates there might be a lack of standardization in usability evaluation. There might be several reasons for this. A possible explanation was that Mazza (2009) argued that evaluation could be limited to only a subset of criteria depending on the type of application. Besides, in line with Mazza (2009), Blaise et al. (2018) pointed out that resources (such as time and cost) might be another reason. They also mentioned that implementing a grounded assessment of usability was difficult since its impact depended on the time taking the effort to "call contributors in and to analyze feedback." (p. 20). However, evaluation affects applicability. In the discussion of RQ3, it will be discussed further why the lack of proper studies in usability evaluation hinders the applicability of data visualization applications.

A small sample of available studies, as well as diversity in evaluation methods and design, caused a lack of transparency in analysis. Consequently, doing a meta-synthesis analysis based on such a diverse sample was not easy. However, these limitations did not prevent valuable insights from being uncovered. The practices of Different Usability Evaluation Methods (UEMs) used by different research in the literature might be a good place to start for considering choosing appropriate criteria. In one sense, the academic community should pay more attention to the gaps; on the other hand, different practices might provide inspiration for further investigation. Further steps should be taken such as how to identify the target audience for a reliable evaluation, which facts need to be taken into account so on and so forth. For example, some existing practices could be used as a reference when designing evaluations for data visualization. In a study by Hoyer et al. (2021) on the Gradation of Reporting Activity, one of the main findings was that good reporting by citizen scientists does not automatically indicate a better experience using ornithological data and understanding the visualizations (p. 157). As a result, in similar settings, the reporting activity may not be a relevant factor when selecting who will be tested in terms of the validity of the evaluation.

A variety of evaluation techniques and methods have been successfully applied to human-computer interaction (HCI), including focus groups, field studies, and expert reviews (Tory & Moller 2005, p. 8). Data visualization is part of human-computer interaction. The results showed that all these methods and techniques have been studied in the literature. In relation to the question of who should be tested, Tory and Moller (2005) suggest that expert evaluation complements formal user studies. For example, experts can evaluate early prototypes (formative evaluations), then end users can evaluate refined versions (summative evaluations) (p. 11). The literature has reported practices of evaluation conducted on experts, and/or non-experts. However, due to a lack of time and resources, as well as limited knowledge of application evaluation, the methods of evaluation have not been analyzed. The methodological evaluation remains relevant but is beyond the scope of this study. Future research should focus on analyzing the strengths and weaknesses of evaluation techniques in the field of visualization, as suggested by Tory and Moller (2005).

5.3 Discussion of RQ3

The third question in this research was to identify opportunities and challenges associated with applying data visualization in a Citizen Science context in both the technical and the social domains.

Opportunities and Challenges in the Technical Domain

Data quality

Data quality issues have implications for data visualization design strategies and validity. The finding of data quality-related challenges has been a focus on data heterogeneity (high variety), data incomparability, and uncertainty in the spatial dataset. Based on the findings from the uncertainty challenge, it seems that explanations of dataset scope are just as important as explanations of visualization features when interpreting data visualizations. In an interesting comparison of two studies focused on heterogeneity (high variety) challenges in historical data, Tuppen et al. (2016) advised cleaning and enhancing the bibliographic data, while Blaise et al. (2018) suggested that removing *imperfection* was not a good idea in minor cultural heritage. Rather, being allowed to remain imperfect was of great significance in minor cultural heritage. These results may help us understand the importance of subject-specific nature when visualizing different types of datasets. One visualization feature may be important for a particular dataset while another may not. An implication of this is the possibility that when visualizing heterogeneous CS data, a different strategy is required in relation to the subject-specifical nature of CS. Furthermore, another important finding was the incomparability challenge and its potential solutions. Gray et al. (2021) investigated the possibility of using an open-source application approach to allow end users to compare their own data to an accepted reference, which ensured the quality and reliability of the data before it was submitted (p. 8). Although attempts have been made, there is not enough evidence or discussion in the literature to assess the risks and benefits of to what extent an open-source software solution could address the incomparable challenge in the CS context.

Applicability

One of the most important results of this study was the opportunities for how the findings of best practices in the literature could be transferred to other use cases, and how. For some data types, both free and commercial tools could be used. One study demonstrated the case to use free visualization tools that were well suited to their requirements and easy to use, no programming knowledge was needed. However, more studies reported the requirement of developing their own applications based on their specific project or application goals and limitations. Several barriers were identified in the literature. Even though there has been a considerable amount of development of tools, many of them are not accessible due to cost, system/ program requirements, or are typically designed for a specific protocol or formatted input data that cannot be widely applied to other applications. Additionally, the high requirement of prerequisite skills to use an existing application was also seen as a barrier.

It was confirmed in the findings that there was a gap between the requirement and the available resources needed to successfully develop a new application or solution. Therefore, the findings suggested investigating whether relevant research was available, or whether existing solutions might fit or were adaptable to the design requirements in the early stage of the design process. However, based on the findings of RQ2, several reports have shown a lack of studies of proper usability evaluation of applications in this field, such as which functionality was useful for which target audience and in what context. An implication of this is the possibility that it will be difficult for designers or other stakeholders to assess the applicability of existing solutions if they lack evidence-based information. It also indicates that the evaluation could an important indication of applicability and may thus help identify applicability.

However, a noteworthy observation is that the current study found a few studies have considered and examined the possibility of applying their practices under certain conditions and they were grouped and reported as four frameworks. Besides the findings of the four frameworks, the result showed that adopting parts of visualization design strategies may work within the special requirements of the application goals even though the entire existing design solution or framework may not be applicable. One significant finding was adopting color theory techniques to investigate preserved and absolute color meaning in water monitoring observational citizen science. As presented in sections 4.1 and 4.3, to the best of my knowledge, the color meaning may remain underexplored or underresearched in other scientific fields regarding visualization. Perhaps little attention was paid to other scientific domains, the literature lacked sufficient information to determine if such a need for color meaning existed or not. However, interpreting visualization from a perceiver's perspective was highlighted in another study but simply focused on when communicating scientific findings to the public, instead of focusing on the data analysis. Nevertheless, the perceiver's perspective offered a chance to consider solving existing challenges by incorporating established theories i.e., color theory from other research fields.

Opportunities and Challenges in the Social Domain

Knowing the target audience is crucial

The literature has demonstrated that identifying the target audience and their needs is especially crucial. Based on the results, it is important to consider how a user's background, prior knowledge, motivation, and personal beliefs can affect the accuracy, reliability, and understanding of visualization. While only a few studies have attempted to identify a target audience for evaluation. This result may be due to a lack of reports, a lack of attention, or a lack of methods and resources available in academia. The lack of clarity in defining whom the target audiences are (in Section 4.2.3) may also be a factor. In spite of these limitations, the study has nonetheless yielded some insights. An empirical study suggested that it might not be necessary to consider reporting behavior of the participants as a relevant criterion when evaluating the observational CS visualization. The question of what criteria to use to identify the target audience for data visualization remains unanswered. This is an important issue for future research.

Data visualization as an aid for communication and learning

Based on Roche et al. (2020), studies such as Storksdieck et al. (2016), Roche and Davis (2017a), highlight that as citizen science develops as an emerging professional field, education and learning are critical issues. When it comes to communication in CS, in line with the discussion by Roche et al. (2020) in the Background section, one-way communication in citizen science involves communication of scientific findings, funding-specific public relations obligations, or attracting participants. Despite the high requirement for sufficient resources, two-way communication allows citizens and scientists to exchange information, build trust and relationships, and foster collaborations and learning (Mercer & Littleton, 2007, as cited in Roche et al., 2020). According to the current study, it was believed that data visualization could be used as an educational tool in both formal and informal education. Data visualization practices were found to support understanding data and communicating scientific findings and knowledge by citizen scientists, scientists, and educators.

According to Kloetzer et al. (2021), it was important to consider diverse audiences when planning and developing training to increase participants' scientific competence and support, as well as ensure the desired outcomes of citizen science projects (p. 302). According to Keim et al. (2008), meaningfully, different challenges and opportunities faced by scientists and citizens were found when it came to training or supporting scientists and citizens to understand and analyze/visualize data, as well as training or supporting them to understand the analyses/visualizations as mentioned by Blaise et al. in 2018 (p. 19). The plain language text support and various forms of visual support such as photographs integrated with multi-media were suggested to ensure data validity. In addition, it was an effective method of communicating with a broader audience. The results suggested that the data visualization application itself could be used in the training section by comparing collected data with reference data to ensure data validity and reliability.

Additionally, several studies also suggested video tutorials could be used to enhance understanding to aid efficient training. However, there was a need to notice that findings showed that video tutorials would have negative effects, such as being less mentally demanding and exerting less effort. Especially, it is crucial to realize a video tutorial can't resolve the transparency issue caused by the missing presence data. The finding suggested that visualizing the missing presence data requires not just a visual explanation of how the data is visualized and interpreted, but also explaining the meaning and scope of the datasets, and the limitations of the data validity.

A further important finding was that providing individualized solutions for a diverse audience supported the understanding of the visualization. In line with Daniel (1992), Sandhaus et al. (2019) stated that “safeguards against misrepresentation are essential to the validity of the data visualization field” (p. 68). The present study raises the possibility that individualized visualization requires careful consideration, including presenting and communicating relevant information in a way that meets the needs of a diverse audience while maintaining the validity of the data. While preliminary, these results have important implications for the design and the use of training and learning materials for data visualization. Along with who the target audience is, it indicates how these relevant factors should be considered in a subject-specific CS context.

The impact of collaborative design on the social domain

Several previous studies have noted the importance of collaboration in citizen science. As discussed in the Background section regarding to Arnstein, citizen science plays an important role in activism and advocacy when it comes to education and learning (Roche et al., 2020, p. 4). Regarding Haklay's (2013) typology (see Figure 3), he suggests that learning will be enhanced as citizens get more involved and engaged in citizen science. The collaborating level is the top level, knowing and acquiring knowledge as active and collective activities (p. 5).

The literature identified three collaborative approaches to Citizen Science data visualization practices. One co-design approach was discussed as a framework previously. Open-ended design study suggests that open-source solutions can be integrated with open-end design approaches. As described in the discussion on open source in section 5.2, however, it is crucial to consider what resources are needed in particular and what pre-skills are needed in order to implement this approach successfully. Another collaborative approach, co-authoring public visualization with polling systems, was identified in the literature and the findings proved the outcomes of promoting participatory engagement and actionable behavior. Unlike the open-ended design in developing an open-source application which focused on the application itself, this co-authoring approach focused on encouraging participants involved in the CS project by reflecting and acting on the readily collected data. The results suggest taking social dimensions into account in the context of CS data visualization. Despite their differences, both collaborative approaches promote action. Therefore, citizen science projects that promote action might benefit from a collaborative approach.

5.4 Data visualization in Higher Education: Recommendations

Based on Bouwma-Gearhart and Jennifer Collins (2015), data capacity and information literacy are becoming increasingly important for Higher Education, specifically around data visualization techniques (Zentner et al., 2019). According to the results of this thesis study, the best practices of data visualization technologies in Citizen Science were identified. Can visualization techniques be used in Higher Education based on lessons learned from citizen science? This section provides recommendations for using data visualization technology in Higher Education based on the best practices identified in the meta-synthesis.

5.4.1 Identifying Best Practices

Identifying learning goals and activities

- The informal education field used data visualizations to teach students about "the mechanics of data representation, pointing out specific visual features to flag potential errors or limitations" (Snyder, 2017, p. 2107).

- To facilitate long-term engagement, Oliver et al., (2020) recommended 1) Growth of knowledge with purpose, 2) Diversity in Task Difficulty, 3) Customizable Complexity, and 4) a Collaborative Puzzle Solving approach (p. 1694f).

- There have been several successful collaborative approaches applied to different stages of scientific research in various contexts. According to Sandhaus et al. (2019), a collaborative approach promoted self-efficacy and capacity, by making personalized decisions and developing personalized prevention and intervention strategies in the Gardenroots model (p. 56).

- With an open-source data visualization tool investigated by Gray et al. (2021), end users were able to independently analyze, visualize, and compare their data to reference datasets, which reinforced their understanding of the quality of their own data and how to improve it. When citizen scientists are involved directly in open-ended collaborative activities, they can ask their own questions, as well as learn about potential questions that they were not aware of. In addition to promoting individual development, these open-ended collaborative activities facilitate two-way communication between diverse audiences.

- Claes et al. (2018) discovered a co-design approach in public visualization with polling interaction promoting engagement, actionable behavior, and distribution potential. In contrast to data trends and patterns in visualizations themselves, the researcher pointed out that social relationships encourage more reflection (p. 8). Their recommendation was to consider social dimensions when distributing a public visualization.

Making the right choice of data visualization technology

- A variety of free and commercial tools suit well the visualization requirements and are easy to use since no programming skills are required. Tuppen et al. (2016) introduced three free tools in their practices, Google Fusion Tables, OpenHeatMap, and Palladio.

- The benefits of interactive web design include 1) reducing disk space and memory demands, 2) improving accessibility across operating systems, and 3) eliminating the need to compile or install software locally due to its massive disk space and memory requirements (Abramov et al., 2022).

- Using visualizations for data communication or sharing raises privacy concerns for end users (Gray et al., 2021). This problem can be addressed by de-identifying the data and preventing its re-identification, as suggested by Hsu et al. (2018, p. 3)

Training and learning materials to support analysis and communication

- In line with several studies in the literature, Kloetzer et al. (2021) stressed the need to consider a diverse audience when planning and designing citizen science training in order to increase participants' scientific competence, support, and achieve the intended outcomes (p. 302).

- There are several studies (e.g., Kloetzer et al., 2021; Tuppen et al., 2016; and Sandhaus et al., 2019) that suggest providing individualized solutions for training and supporting materials for diverse audiences. According to Sandhaus et al. (2019), CS project results should be displayed in personalized visualizations, and figures should be simplified to present only information that is relevant to participants and within their expectations.

- A broader approach to visualization was suggested to develop training and learning materials to support data visualization in citizen science settings. In order to ensure that the visual design and content were correct and consistent, Snyder (2017) reported that a hands-on visual was developed through collaborative vernacular design practice. Study results confirmed the importance of providing different forms of data visualization since these forms could provide different possibilities for presenting the information.

- Improving communication clarity by switching specific science language to plain language was highlighted by Snyder (2017), Gray et al. (2021), and Sandhaus et al. (2019). According to Gray et al. (2021), grassroots community members should be provided with informal explanations that differ from those provided by and for the wider scientific community.

- Video tutorials were shown to have a positive effect on enhancing the understanding and ease of the use of data visualization by Hoyer et al. (2020). It was worthwhile to consider there was also empirical evidence showing the challenges related to providing video tutorials, such as the support of video tutorials could exert less effort and be less mentally demanding.

5.4.2 Best practice applying to Higher Education

Higher Education institutions have begun to emphasize the importance of data visualization, which is regarded as a key component of information literacy (Zentner et al., 2019). As proved in this meta-synthesis study, thanks to its analytical and communicative nature, data visualization is part of the learning process itself in the learning environment of citizen science. Often, citizen science data visualization is based on scientific data, involving authentic scientific research processes, and communicating scientific findings, which also often occurs in higher education. As data visualization plays a large role and has strong potential benefits for the "from data to knowledge" flow in citizen science, it is not surprising that high education can also benefit from it. On the basis of the results of the best practices of this meta-synthesis study, data visualization in CS holds many lessons applicable to higher education to support students' learning. In an empirical study, Kandlikar et al. (2018) reported a collaborative course between researchers, educators, and students. This example gives a successful case showing Higher Education can benefit from data visualization when it is appropriately applied.

A collaborative use case on eDNA researchers and an undergraduate microbiology course

The goal of the twenty-week course was to provide undergraduate students with an authentic research experience in microbiology and microbial community ecology. As part of the course, students were introduced to the structure of eDNA sequencing results using the Ranacapa Shiny app. They were encouraged to explore data and perform statistical analyses most relevant to the hypotheses they had developed. A key benefit of using ranacapa was that despite having no prior bioinformatics experience, students could begin exploring their own collected data by using the online instance of the app. Through data visualization, Students can gain a better understanding of visualizing, and teachers can reduce the time and difficulty associated with troubleshooting informatics issues. Ranacapa is being integrated into the upcoming undergraduate curriculum module "Pipeline for Undergraduate Microbiome Analysis", which will be an open-source, comprehensive suite of analysis and data visualization tools for undergraduate researchers. (Kandlikar et al., 2018, p. 8)

Figure 16 illustrates recommendations for educators in Higher Education to use data visualization technology to support students' learning. The recommendations consist of four steps. First, the Initiation step considers three factors: setting educational goals, investigating available resources, and getting to know the student, namely target audiences. Identifying gaps between educational goals, resources available, and the background and skills of students are essential in order to consider to what extent specific data visualization tools can be used to achieve specific learning goals. As an example, for the target student group, it is vital to identify whether data visualization is needed to improve analytical or communicative skills or both. The reason is that a certain type of data visualization tool of a certain subject area may support analysis, but it may not be appropriate for communication needs. If the educational aim is for both analysis and communication, then it may need to search for other better options, etc. The second step involves the preparation of data visualization technology and training materials to support their use. Employing collaborative activities in implementation is highlighted in the third step. The inspiration from Roche et al. (2020), Snyder (2017), Gray et al. (2021), and Sandhaus et al. (2019), knowing and acquiring knowledge are integrated actively in various forms of collaboration and promotes real engagement of the target students. In addition, feedback plays an important role in learning. However, a difficult part of the process is getting feedback to assess the understanding of data visualization. Feedback on visualizations can be provided by educators, researchers, and peers, but other forms of feedback are also needed. Due to the complexity of evaluation in data visualization in citizen science, lessons are learned to find an easy-to-use evaluation solution to promote feedback and assess visualization. Conducting this literature review leads to the discovery of VisLab (see Figure 17). As a free and easy-to-use online platform, VisLab provides the opportunity to set up a personal experiment of visualization easily and outstretch to a broad audience, as well as gather feedback in order to enable assessing the understandability of visualization. It is hoped that in addition to VisLab, other solutions will be discovered or developed to help with the assessment process. Another assessment method is hosting a public visualization (Claes et al., 2018) to encourage reflective thinking. Either as a form of an assessment conducted between educators and students or as a co-design approach open and outstretched to local communities together with educators and students. In the end, as a stand-alone recommendation, detailed information is provided in Figure 16.

Initiation	Preparation	Implementation	Assessment
Goal, Resources and Audience	Techniques and Training materials	Collaborative Activities	Feedback and Assessment
Set data visualization's goals based on educational needs. Know the students' backgrounds, knowledge, and skills, as well as what they need to learn and achieve. Investigate the resource available, for example, there may be able to find ready-to-use examples of successful use of the curriculum by exploring the resources available. One example is the microbiology course with the model "Pipeline for Undergraduate Microbiome Analysis" using the data visualization application ranacapa (Kandlikar et al., 2018)	Identify the best application based on educational goals, available resources (technical issues and related learning materials), and the need of the target students. In terms of technical issues : • Is there any free and easy-to-use tool available that meets the goals? • Utilizing web-based/mobile apps that are easy to use and require fewer resources. • Using an open-source data visualization tool to allow independent analyzing, visualizing, and comparison among datasets, which reinforced understanding of the data quality issue. • Data Privacy issues should be considered. In terms of related learning materials for the use of technology: • Are there any existing materials? Explanatory text, video tutorials, additional links, hands-on visual documents, etc. Students may feel that their mental effort is reduced and that there is a lower level of demand if they are supported by video tutorials, which can be a challenge for video tutorial providers. • How well do the existing materials meet the learning needs? Where are the gaps? • Preparing the additional materials to fulfill the gap if there is one. • Providing students with a variety of media and forms of learning materials based on the resources available, lets them choose what works best for them.	It is recommended to use a collaborative approach as well as a variety of task difficulty levels when implementing learning activities. Collaborative activities facilitate two-way communication and promote self-efficacy and capacity, by making personalized decisions and developing personalized strategies. • By applying co-authoring data-driven narratives in the physical learning environment, deep involvement and engagement can be achieved. • Using open-source applications enables students directly in designing the application and thereby independently visualizing data depending on their preferences and concerns. In order to successfully implement this approach, it is important to consider what resources and pre-skills are required in particular. It is especially valuable when proving technical skills in data visualization is the main educational goal.	As a form of feedback, combining evaluation tools to enable students to assess the quality of their visualizations. VisLab, a free and easy-to-use online platform is presented as an example tool in Figure 17. The solutions like VisLab, provide the opportunity to set up a personal experiment of visualization easily and outstretch to a boarder audience and feedback. The hosting of a public visualization can encourage reflective thinking, either as a form of assessment conducted between educators and students or as a co-design approach open and outstretched to local communities. An example is the study of Claes et al. (2018), who examined a co-design approach in public visualization with polling interaction. They recommend considering social dimensions rather than data trends and patterns regarding public visualization, such as the target audience, the significance of the physical location to them, their perception and motivation, and their interactions caused by the visualization.

Figure 16. Recommendations for educators in Higher Education to use data visualization technology to support students' learning.

VisLab (Figure 17) is an online platform that allows users to design and deploy experiments to evaluate their visualizations. Easy-to-deploy experiments could be created using VisLab's online platform. Analytic dashboards and scaffold templates simplify experiment design and analysis. the platform supports also anonymous participation and personalized feedback, to motivate broad participation. Use case scenarios demonstrate the usability and usefulness of the platform in addressing the different needs of practitioners, researchers, and educators.

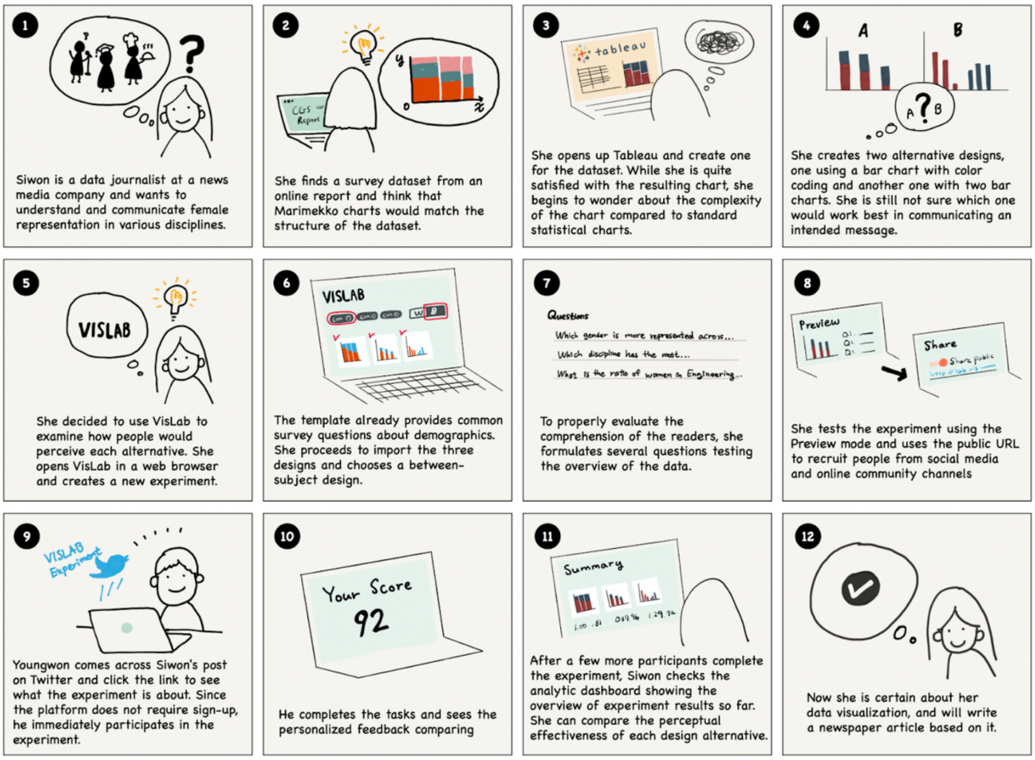


Figure 17: A user scenario showing how a visual data journalist would use VisLab to answer questions about the trade-off of her design alternatives and derive design feedback on how actual participants would react to her visualizations. (Source: Choi et al., 2021, p. 6)

5.5 Limitations

This master's thesis was constrained by time, so its findings must be considered within the context of some limitations. Generally, these results cannot be generalized due to the small sample size (16 articles). This study, however, is qualitative in nature, and the research questions were explained in a context, so validity claims may not be supported. Furthermore, only one coder was included in the meta-synthesis study, and given that the researcher had limited background knowledge, these limitations impacted the interpretation of the results, such as possible errors, subjective interpretations, and biases. In addition, a methodology limitation has been discussed in the Method section. Due to its contemporary scope, this topic has received little prior research. No existing framework could be used to analyze the data. The open coding method identifies what was visible but fails to reveal what was undercovered and what had not been discussed. Furthermore, the quality of the research design and evaluation design could affect the reliability of the findings. However, the research design and method of the study have not been examined, nor has the quality of the methodology of the articles been compared. A meta-analysis of the 16 samples is needed in the future.

6 Conclusion

6.1 Answering the Research Questions

Literature-based analyses provide a more comprehensive view of Citizen Science's data visualization feedback loop. Besides technical aspects of visualization, social context, viewer perception, individual experience, and pre-knowledge are also crucial factors to consider. There is a merging of social and technical aspects of data visualization in CS, which are considered to be interdependent.

From a technical perspective, there has been a focus on data heterogeneity (high variety), data incomparability, and uncertainty in the spatial dataset as challenges related to data quality. When interpreting data visualizations, explanations of dataset scope were just as important as explanations of visualization features, regarding the uncertainty challenges. Easy-to-access open-source solutions in CS data visualization were evidently emerging as a trend, it enabled wider participation in CS communication. Further research on the risks and benefits associated with open-source software solutions may be able to address the incomparable challenges in greater depth. Another important observation was the applicability. This study identified four frameworks and reported applicable visualization design strategies. Usability evaluation seemed to be lacking standardization and sufficient study hindering applicability. However, it was still possible to uncover valuable insights despite these limitations.

Importantly, from a social perspective, designing a technically effective visualization was insufficient; factoring in the user's perception is necessary. It was vital to consider how a user's background, prior knowledge, motivation, and personal beliefs can affect the accuracy, reliability, and understanding of visualization. The study also discovered three different collaborative CS data visualization practices that provided valuable insights into in-depth engagement through interactive collaboration. Additionally, visualizing data in CS was an activity that was geared toward learning, and it was believed that data visualization could be used as an educational tool in both formal and informal education. Meaningfully, the impact factors of training were identified and justified. For data validity and effective communication with a broader audience, plain language texts and visual media such as photographs and video tutorials were recommended. A data visualization application can be used as a training tool to ensure data quality. Individualizing solutions for a diverse audience can help to facilitate the understanding of the visualization as well.

Based on lessons learned from best practices, a set of recommendations has been developed to help educators in Higher Education to improve student learning using data visualization technology.

6.2 Study Contributions and Future Directions

The findings of this thesis contribute to the current literature in several ways by improving our understanding of data visualization in the context of Citizen Science. Firstly, the illustration of the "data visualization feedback loop in Citizen Science" provides a more comprehensive view of current understanding based on emerging trends and concepts evaluated in the literature. Additionally, valuable insights are gained, and the applicability is discussed. Meanwhile, unresolved, or under-studied problems are identified, which demand further research and attention. The theoretical and practical information in this thesis is intended to inspire designers or CS project managers to incite thought, or to provide a modest contribution to the design of applying data visualizations in CS. Besides providing unanswered or understudied problems to researchers in this field, this thesis will hopefully lay the groundwork and spur further research. Moreover, a study of this type hopes to, by providing actionable recommendations based on best practices, assist educators interested in using data visualization technology to support students' learning.

Based on recommendations, an empirical use case study in Higher Education is a logical next step. The next step in this research can also include developing recommendations for designing data visualization activities in CS and conducting an empirical study.

7 References

Antony Bryant & Kathy Charmaz. (2007). Grounded Theory and Reflexivity. In *The SAGE Handbook of Grounded Theory*. SAGE Publications Ltd. https://doi.org/10.4135/9781848607941.n24

Appelbaum, L. G., & Norcia, A. M. (2009). Attentive and pre-attentive aspects of figural processing. *Journal of Vision (Charlottesville, Va.)*, *9*(11), 18.1-18. https://doi.org/10.1167/9.11.18

Bonn, A., Allott, T., Evans, M., Joosten, H., & Stoneman, R. (Eds.). (2016). *Peatland Restoration and Ecosystem Services: Science, Policy and Practice*. Cambridge University Press. https://doi.org/10.1017/CBO9781139177788

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience*, *59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Bryant, A., & Charmaz, K. C. (2007). *The SAGE handbook of grounded theory*. SAGE.

Byrne, D. (2016). Reviewing the Literature. In D. Byrne, *Research Project Planner*. SAGE Publications, Inc. https://doi.org/10.4135/9781526408518

Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (Eighth ed.).

de Sherbinin, A., Bowser, A., Chuang, T.-R., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., Popescu, I., Shonowo, A., & Sivakumar, K. (2021). The Critical Importance of Citizen Science Data. *Frontiers in Climate*, *3*. https://www.frontiersin.org/articles/10.3389/fclim.2021.650760

de Vries, M., Land-Zandstra, A., & Smeets, I. (2019). Citizen Scientists' Preferences for Communication of Scientific Output: A Literature Review. *Citizen Science: Theory and Practice*, *4*(1), 2. https://doi.org/10.5334/cstp.136

Delbert Miller & Neil Salkind. (2002). Grounded Theory Research. In *Handbook of Research Design & Social Measurement* (6th ed.). SAGE Publications, Inc. https://doi.org/10.4135/9781412984386.n35

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., Phillips, T., & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297. https://doi.org/10.1890/110236

Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*.

Green, J., & Thorogood, N. (2009). Qualitative methods for health research. *Choice Reviews Online*, *47*(02), 47-0901-47–0901. https://doi.org/10.5860/CHOICE.47-0901

Haklay, M. (2013). Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice* (pp. 105–122). Springer Netherlands. https://doi.org/10.1007/978-94-007-4587-2_7

Hecker, S., Bonney, R., Haklay, M., Hölker, F., Hofer, H., Göbel, C., Gold, M., Makuch, Z., Ponti, M., Richter, A., Robinson, L., Iglesias, J., Owen, R., Peltola, T., Sforzi, A., Shirk, J., Vogel, J., Vohland, K., Witt, T., & Bonn, A. (2018). Innovation in Citizen Science – Perspectives on Science-Policy Advances. *Citizen Science: Theory and Practice*, *3*, Art. 4. https://doi.org/10.5334/cstp.114

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Communications of the ACM*, *53*(6), 59–67. https://doi.org/10.1145/1743546.1743567

Hsieh, H.-F., & Shannon, S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*, 1277–1288. https://doi.org/10.1177/1049732305276687

Kasperowski, D., & Hillman, T. (2018). The epistemic culture in an online citizen science project: Programs, antiprograms and epistemic subjects. *Social Studies of Science*, *48*(4), 564–588. https://doi.org/10.1177/0306312718778806

Kloetzer, L., Lorke, J., Roche, J., Golumbic, Y., Winter, S., & Jõgeva, A. (2021). Learning in Citizen Science. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, & K. Wagenknecht (Eds.), *The Science of Citizen Science* (pp. 283–308). Springer International Publishing. https://doi.org/10.1007/978-3-030-58278-4_15

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. https://doi.org/10.1016/j.tics.2006.10.012

Kollmuss, A., & Agyeman, J. (2002). Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research*, *8*(3), 239–260. https://doi.org/10.1080/13504620220145401

Mazza, R. (2009). *Introduction to Information Visualization*. Springer London. https://doi.org/10.1007/978-1-84800-219-7

Milliken, P. (2010). *Grounded Theory* (Vol. 1, pp. 548–553).

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (2012). The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, *10*(6), 298–304. https://doi.org/10.1890/110294

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Paleco, C., García Peter, S., Salas Seoane, N., Kaufmann, J., & Argyri, P. (2021). Inclusiveness and Diversity in Citizen Science. In K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson, & K. Wagenknecht (Eds.), *The Science of Citizen Science* (pp. 261–281). Springer International Publishing. https://doi.org/10.1007/978-3-030-58278-4_14

Pocock, M. J. O., Evans, D. M., Fontaine, C., Harvey, M., Julliard, R., McLaughlin, Ó., Silvertown, J., Tamaddoni-Nezhad, A., White, P. C. L., & Bohan, D. A. (2016). *The Visualisation of Ecological Networks, and Their Use as a Tool for Engagement, Advocacy and Management* (Vol. 54, p. 85). https://doi.org/10.1016/bs.aecr.2015.10.006

Popovich, N., Schwartz, J., & Schlossberg, T. (2017, March 21). How Americans Think About Climate Change, in Six Maps. *The New York Times*. https://www.nytimes.com/interactive/2017/03/21/climate/how-americans-think-about-climate-change-in-six-maps.html

Jacobson, R. (1999). *Information Design*. The MIT Press.

ResearchRabbit, *Version 2.0 Human Intelligence Technologies, Incorporated.* (n.d.). https://www.researchrabbitapp.com

Roche, J., Bell, L., Galvão, C., Golumbic, Y. N., Kloetzer, L., Knoben, N., Laakso, M., Lorke, J., Mannion, G., Massetti, L., Mauchline, A., Pata, K., Ruck, A., Taraba, P., & Winter, S. (2020). Citizen Science, Education, and Learning: Challenges and Opportunities. *Frontiers in Sociology*, *5*. https://www.frontiersin.org/article/10.3389/fsoc.2020.613814

Social Research. (2021, March 1). *Qualitative Content Analysis* [Video]. YouTube. https://www.youtube.com/watch?v=arXQ91uh9rY

Stamm, K. R., Clark, F., & Eblacas, P. R. (2000). Mass communication and public understanding of environmental problems: The case of global warming. *Public Understanding of Science*, *9*(3), 219–237. https://doi.org/10.1088/0963-6625/9/3/302

Strauss, A. L. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, Calif. Sage.

Sullivan, B., Phillips, T., Dayer, A., Wood, C., Farnsworth, A., Iliff, M., Davies, I., Wiggins, A., Fink, D., Hochachka, W., Rodewald, A., Rosenberg, K., Bonney, R., & Kelling, S. (2016). Using open access observational data for conservation action: A case study for birds. *Biological Conservation*, *208*. https://doi.org/10.1016/j.biocon.2016.04.031

Tory, M., & Moller, T. (2005). Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, *25*(5), 8–11. https://doi.org/10.1109/MCG.2005.102

Trumbo, J. (1999). Visual Literacy and Science Communication. *Science Communication*, *20*(4), 409–425. https://doi.org/10.1177/1075547099020004004

Tucker, T. (2016). Grounded Theory Generation: A Tool for Transparent Concept Development. International Studies Perspectives, 17(4), 426-438.

Tufte, E. R. (1983). *The visual display of quantitative information*. Graphic Press.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, *15*(3), 398–405. https://doi.org/10.1111/nhs.12048

Veeckman, C., & Temmerman, L. (2021). Urban living labs and citizen science: From innovation and science towards policy impacts. *Sustainability (Switzerland)*, *13*(2), 1–15. https://doi.org/10.3390/su13020526

Trumbo, J. (1999). Visual Literacy and Science Communication. Science Communication, 20(4), 409–425. https://doi.org/10.1177/1075547099020004004

Vries, R.D. (2018). *The importance of user analysis before the technical design of an instrument, which presents information to users from a different discipline.*

Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., Bhavsar, A., Eyers, J., Koehlmoos, T. P., Petticrew, M., Valentine, J. C., & Tugwell, P. (2012). How to do a good systematic review of effects in international development: A tool kit. *Journal of Development Effectiveness*, *4*(3), 359–387. https://doi.org/10.1080/19439342.2012.711765

Patterson, L. (2018, December 6). *What are data, information, and knowledge?* Internet of Water - Better Water Data for Better Water Management. https://internetofwater.org/valuing-data/what-are-data-information-and-knowledge/

Wolfswinkel, J., Furtmueller, E., & Wilderom, C. (2013). Using Grounded Theory as a Method for Rigorously Reviewing Literature. *European Journal of Information Systems*, *22*. https://doi.org/10.1057/ejis.2011.51

Zentner, A., Covit, R., & Guevarra, D. (2019). Exploring Effective Data Visualization Strategies in Higher Education. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3322856

8 Appendices

8.1 Reference (Articles)

Abramov, D., Burchett, J. N., Elek, O., Hummels, C., Prochaska, J. X., & Forbes, A. G. (2022). CosmoVis: An Interactive Visual Analysis Tool for Exploring Hydrodynamic Cosmological Simulations. *IEEE Transactions on Visualization and Computer Graphics*, 2909–2925. Scopus. https://doi.org/10.1109/TVCG.2022.3159630

Biraghi, C. A., Pessina, E., Carrion, D., & Brovelli, M. A. (2020). VGI Visualisation to support participatory lake monitoring: The case study of simile project. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *43*(B4), 237–244. Scopus. https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-237-2020

Blaise, J.-Y., Dudek, I., & Saygi, G. (2018). Citizen Contributions and Minor Heritage: Feedback on Modeling and Visualising an Information Mash-up. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 11–20. https://doi.org/10.1109/DSAA.2018.00011

Claes, S., Coenen, J., & Vande Moere, A. (2018). Conveying a civic issue through data via spatially distributed public visualization and polling displays. *Proceedings of the 10thNordic Conference on Human-Computer Interaction 2018*, 597–608. Scopus. https://doi.org/10.1145/3240167.3240206

Clark, G. F., Gacutan, J., Lawther, R., Johnston, E. L., Tait, H., & Bednarz, T. (2021). A visualization tool for citizen-science marine debris big data. *Water International*, *46*(2), 211–223. https://doi.org/10.1080/02508060.2021.1888495

Gray, A., Robertson, C., & Feick, R. (2021). CWDAT—An Open-Source Tool for the Visualization and Analysis of Community-Generated Water Quality Data. *ISPRS International Journal of Geo-Information*, *10*(4), Article 4. https://doi.org/10.3390/ijgi10040207

Hoyer, T., Moritz, J., & Moser, J. (2021). Visualization and Perception of Data Gaps in the Context of Citizen Science Projects. *KN - Journal of Cartography and Geographic Information*, *71*(3), 155–172. https://doi.org/10.1007/s42489-021-00083-2

Hsu, Y.-C., Cross, J., Dille, P., Nourbakhsh, I., Leiter, L., & Grode, R. (2018). *Visualization tool for environmental sensing and public health data*. 99–104. Scopus. https://doi.org/10.1145/3197391.3205419

Kandlikar, G. S., Gold, Z. J., Cowen, M. C., Meyer, R. S., Freise, A. C., Kraft, N. J. B., Moberg-Parker, J., Sprague, J., Kushner, D. J., & Curd, E. E. (2018). Ranacapa: An R package and shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations [version 1; referees: 1 approved, 2 approved with reservations]. *F1000Research*, *7*. https://doi.org/10.12688/f1000research.16680.1

Navarra, C., Opach, T., Vrotsou, K., Joling, A., Wilk, J., & Neset, T. S. (2020). Visual Exploration of Climate-Related Volunteered Geographic Information. *Workshop on Visualisation in Environmental Sciences (EnvirVis)*, 7 pages. https://doi.org/10.2312/ENVIRVIS.20201092

Nicholson-Cole, S. A. (2005). Representing climate change futures: A critique on the use of images for visual communication. *Computers, Environment and Urban Systems*, *29*(3), 255–273. https://doi.org/10.1016/j.compenvurbsys.2004.05.002

Offenhuber, D. (2020). Data by Proxy—Material Traces as Autographic Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 98–108. https://doi.org/10.1109/TVCG.2019.2934788

Oliver, J. L., Brereton, M., Turkay, S., Watson, D. M., & Roe, P. (2020). Exploration of Aural & Visual Media About Birds Informs Lessons for Citizen Science Design. *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 1687–1700. https://doi.org/10.1145/3357236.3395478

Sandhaus, S., Kaufmann, D., & Ramirez-Andreotta, M. (2019). Public participation, trust and data sharing: Gardens as hubs for citizen science and environmental health literacy efforts. *International Journal of Science Education Part B-Communication and Public Engagement*, *9*(1), 54–71. https://doi.org/10.1080/21548455.2018.1542752

Snyder, J. (2017). *Vernacular visualization practices in a citizen science project*. 2097–2111. https://doi.org/10.1145/2998181.2998239

Tuppen, S., Rose, S., & Drosopoulou, L. (2016). LIBRARY CATALOGUE RECORDS AS A RESEARCH RESOURCE: INTRODUCING 'A BIG DATA HISTORY OF MUSIC.' *Fontes Artis Musicae*, *63*(2), 67–88. https://doi.org/10.1353/FAM.2016.0011