Data Speaks, But How?

The Role of the Medium in Data Storytelling

Master Thesis in Information Studies

Written by
Tamara Nagel

Supervised by
Toine Bogers

May 2025

Master Thesis in Information Studies

Department of Communication and Psychology

Aalborg University, Copenhagen

Information Studies MSc Program

4th Semester

May 31, 2025

Approximately 23,566 words and 153,656 characters

Tamara Nagel Student in Information Studies

AALBORG UNIVERSITY
STUDENT REPORT

Department of Communication and Psychology

http://www.aau.dk

Title:
Data Speaks, But How?
The Role of Medium in Data Storytelling.

Theme:
User Studies and Information Behaviour

Project Period:
Spring 4. Semester

Project Group:
9

Participant(s):
Tamara Nagel

Supervisor(s):
Toine Bogers

Page Numbers: 123

Date of Completion:
May 31, 2025

Abstract:

Despite growing interest in data storytelling, it remains unclear which communication medium, verbal or written, is most effective in helping people recall and act on data, particularly for those with limited data literacy. This matters especially in times of crisis, where poor data interpretation can lead to harmful real-world consequences. This thesis presents an experimental, longitudinal study comparing the effects of verbal, written, and no storytelling, each using simple graphs and an author-driven narrative on recall and attitude change over time. Results show that while storytelling did not enhance recall, verbal storytelling and no storytelling led to long-term attitude change, unlike written storytelling. Higher data literacy improved long-term recall but initially reduced attitude change, an effect that diminished over time. These findings challenge assumptions about the universal benefits of storytelling, highlighting that author-driven narratives are not equally effective across mediums. The study contributes empirical evidence to the field and calls for further research into how narrative formats, graph complexity, and real-world settings influence the effectiveness of data storytelling.

Keywords: Data Storytelling, Author-driven Narrative, Communication Medium, Attitude Change, Recall, Longitudinal Experiment

Acknowledgements

I want to thank my supervisor, Toine, for the guidance and support throughout the research process, especially during our regular supervision meetings. I'm also thankful to all the participants who took part in my study through Prolific. A big thanks to my boyfriend, Paul, for helping me get through some stressful moments and for proofreading my thesis. I'm also grateful to my parents for their support and for taking the time to read through my work. Finally, thanks to my study friends who made the last two years of this master's journey more fun, collaborative, and motivating.

Contents

Abbreviations

AI Artificial Intelligence

ANOVA Analysis of Variance

ICC Intraclass Correlation

IQR Interquartile Range

ELM Elaboration Likelihood Model

PQRS Preview, Question, Review, Summarize

1 Introduction

Today, people are surrounded by data more than ever. Numbers, charts, and dashboards appear in the news, on social media, and in everyday conversations. They help us understand things like the economy, health risks, climate change, and social issues (Engebretsen and Kennedy, 2020, chapter 1). In the past, data was mostly used by experts, but now it affects many decisions we make in daily life from understanding inflation statistics to interpreting political polling (Kennedy and Engebretsen, 2020). At the same time, the majority of people struggle to access and interpret raw data due to limited data literacy (Frank et al., 2016). This highlights a critical need to make data more accessible and engaging for the public. Data storytelling is one approach which combines data visualization with narrative techniques to make it more appealing to the public by visualizing and explaining complex data in a way that makes it more tangible (Zhang et al., 2022).

Data storytelling has proven to be particularly important in times of crisis. During the Covid-19 pandemic, people around the world watched the news for numbers and graphs of infections or R-values, "[b]ut more than the numbers, we hang on the lips of the interpreters of these data" (Eckert, 2022, Preface). The importance of effective data communication became evident during this time. Misunderstandings about terms such as "flattening the curve" had real-world consequences influencing people's attitudes toward public health measures (Mat Daud, 2022). This demonstrates that the narrative used to contextualize the information is important for the intended message to reach its audience.

Previous research has questioned whether data storytelling enhances the ability to recall information. A study comparing traditional data visualizations with written storytelling-enhanced versions found no significant differences in recall performance, despite storytelling visuals being perceived as clearer and more engaging by some participants (Zdanovic et al., 2022). This raises the question of whether the format in which a data story is delivered — for example, written versus spoken in addition to the data visualization — may influence how well it is received and remembered. Based on this, the current study investigates the role of communication medium in data storytelling, with a focus on how verbal and written formats affect attitudes and information recall. The goal is to better understand whether the medium itself contributes to the impact of storytelling in public communication.

1.1 Problem Definition and Research Questions

Recognizing the importance of understanding data, this study investigates the impact of data storytelling. It focuses on different mediums, such as written and verbal, as well as data visualization. The goal is to understand how the medium of data storytelling affects engagement with information. To that end, the study explores how mediums influence attitude change and recall. Since engaging with data is especially important for individuals with low data literacy to understand data better, the study also explores how these effects may vary depending on a person's level of data literacy. By the end of the study, the following problem statement should be answered:

"**How do different mediums of data storytelling - specifically written and verbal - differ in their effects on audience attitudes and information recall, and how might these effects vary based on audiences' data literacy?**"

To answer this, the research is guided by the following questions:

RQ1: How might audiences' attitudes change after being presented with data storytelling communicated through different mediums?

RQ2: How might audiences' recall differ when being presented with data storytelling communicated through different mediums?

RQ3: How might audiences' level of data literacy influence their ability to recall information they were presented with?

RQ4: How might audiences' level of data literacy influence their attitude change towards information they were presented with?

1.2 Problem Delimitation

This study focuses on verbal and written data storytelling presented alongside identical data visualizations in the context of public communication. The study investigates how these formats influence the recall of information and attitudes toward the content. The storytelling is presented in a short video format inspired by common social media content.

The study **does not** examine different types of media, such as newspapers or TV news, because online video is becoming an increasingly important source of information nowadays (Newman et al., 2024). Including multiple media types would introduce additional complexity, especially since certain formats inherently limit or dictate the way stories are told. Furthermore, the content is presented in a controlled environment and is not published on actual social media platforms. Therefore, platform-specific features, such as scrolling behavior, algorithmic curation, and comment sections, are not considered. The study also does not compare the effects of storytelling across demographic groups or geographic regions. Although some demographic data was collected and analyzed for descriptive purposes, the goal of the study is to provide a general understanding of how the format of stories affects public communication rather than to identify effects specific to subgroups. Finally, the study does not evaluate the factual accuracy, credibility, or ethical framing of the content. All stories are assumed to be neutral and consistent across conditions.

These limitations allow for a deep focus on the core research question: How does the communication medium (verbal versus written) of a data story, alongside data visualization, affect recall and attitude? Narrowing the scope allows the study to avoid unnecessary complexity and focus on the effects of the medium. This would not be feasible if additional variables, such as media type, platform behavior, or demographic differences, were included.

1.3 Thesis Structure

The following Figure 1 shows the structure of this thesis. It begins with the problem statement and the research questions that guide the research and help answer the problem statement. The literature review examines literature related to the research questions to find fitting designs and methods for the study and derive research hypotheses based on current theories.

After collecting the data, the results of the experimental study are analyzed to test the research hypotheses. The research questions are then answered in the discussion that follows. Lastly, the problem statement is addressed in the conclusion.

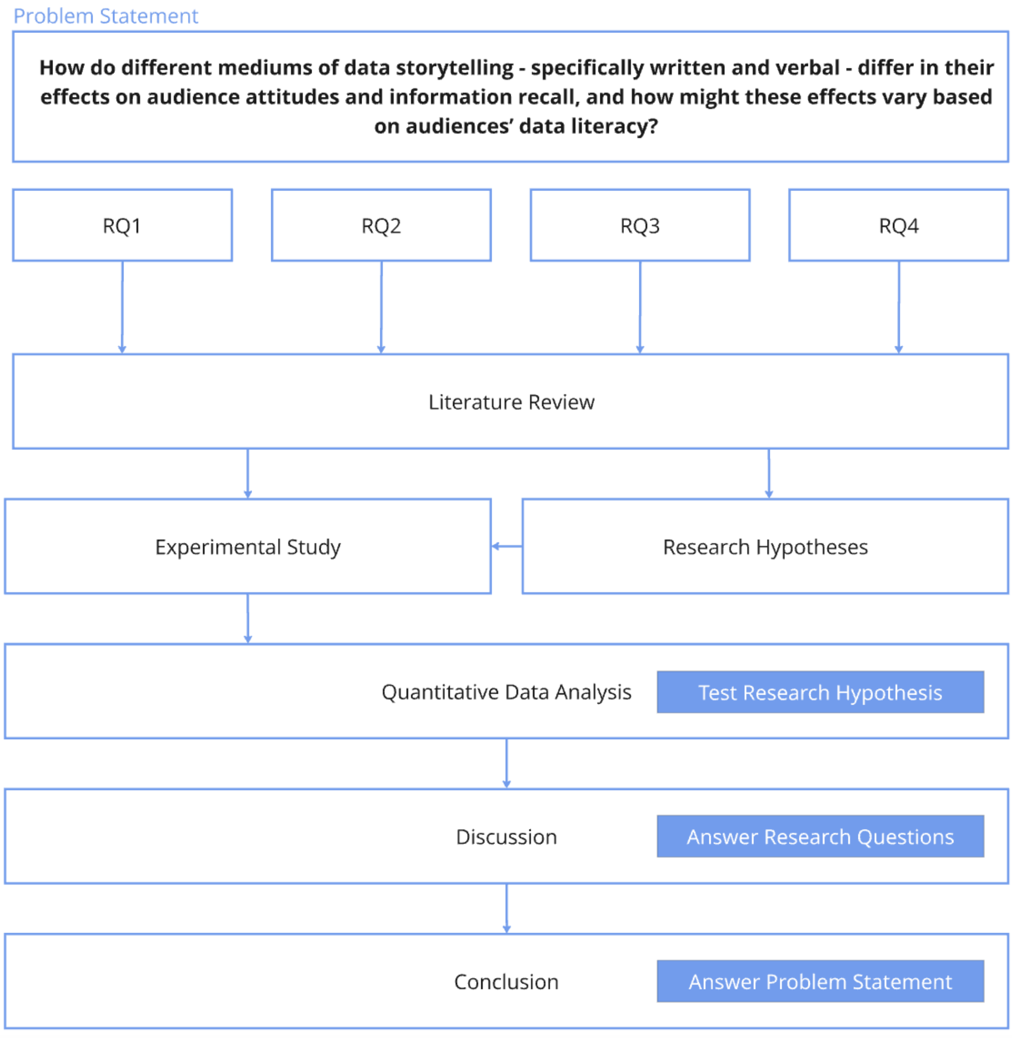


Figure 1: Structure of the Thesis based on the Problem Statement and Research Questions.

2 Literature Review

The goal of this literature review is to examine the current state of knowledge on the topics related to the research questions (Baumeister and Leary, 1997).

A narrative literature review is conducted for this purpose (Ferrari, 2015). The research questions provide a structure for the narrative literature review by identifying relevant topics (Table 1).

Research Question	**Associated Topics**
RQ1: How might audiences' attitudes change after being presented with different data storytelling mediums?	Attitude, Data Storytelling, Data Visualization, Medium
RQ2: How might audiences' recall differ when presented with different data storytelling mediums?	Recall, Data Storytelling, Data Visualization, Medium
RQ3: How might the audiences' level of data literacy influence their ability to recall information they were presented with?	Data Literacy, Recall
RQ4: How might the audiences' level of data literacy influence their attitude change towards the information they were presented with?	Data Literacy, Attitude

Table 1: Topics derived from Research Questions for the Literature Review

2.1 Literature Search Process

Primo, a search engine from Aalborg University is used to find relevant literature. To structure the search process, a building block search is applied (Lund University, 2025). The following Figure 2 shows the building blocks, one for each topic defined in Table 1.

<table>
<tr><td colspan="2">RQ4</td><td></td><td>RQ4</td><td></td><td>RQ4</td></tr>
<tr><td></td><td colspan="2">RQ3</td><td colspan="3">RQ3</td></tr>
<tr><td></td><td colspan="4">RQ 2</td><td></td></tr>
<tr><td colspan="4">RQ 1</td><td></td><td></td></tr>
<tr><th>Attitude</th><th>Data Storytelling</th><th>Medium</th><th>Data Visualization</th><th>Recall</th><th>Data Literacy</th></tr>
<tr><td>Attitude Change</td><td>Storytelling</td><td>Multi-Media</td><td>Data Presentation</td><td>Measure Recall</td><td>Information Literacy</td></tr>
<tr><td>Measure Attitude</td><td>Narrative</td><td>Verbal</td><td>Data Analysis</td><td>Retrieval</td><td>Measure Data Literacy</td></tr>
<tr><td>Attitude Behavior</td><td>Data Narrative</td><td>Written</td><td></td><td>Recall Test</td><td></td></tr>
<tr><td></td><td></td><td>Communication Mode</td><td></td><td></td><td></td></tr>
</table>

Figure 2: Building Block Search

First, one search per block is conducted to get an overview of the topics. For example, for RQ1, "attitude," the Boolean search is "**attitude**" **OR** "**attitude change**" **OR** "**measure attitude**" **OR** "**attitude behaviour**". Then, to get more specific results, a Boolean string of all the building blocks per research question is applied. For RQ1, for example, it is: **("attitude" OR "attitude change" OR "measure attitude" OR "attitude behavior") AND ("data Storytelling" OR "storytelling" OR "narrative" OR "data Narrative") AND ("medium" OR "multi-Media" OR "verbal" OR "written" OR "communication mode") AND ("data visualization" OR "data presentation" OR "data analysis")**. The same logic is applied to the building blocks of the other research questions.

The Preview, Question, Review, Summarize (PQRS) system is followed to critically review the content. In the preview stage, relevant articles are identified. All relevant articles are mentioned in the following summaries. Questions are asked, and the content is reviewed to answer them. During this process, literature cited in the articles that is considered relevant is added to the list of articles to be reviewed. Finally, the relevant information is summarized, and any contradictory findings or different approaches across articles are discussed. (Cronin et al., 2008)

The summaries are structured according to single building blocks. Articles found using a search string that combines multiple blocks are sorted into the most fitting topic.

2.2 Data Visualization

Data are facts or information used to understand something or make decisions (Olson, 2021). "Data visualization is the process of representing data in a graphical or pictorial way in a clear and effective manner. It has emerged as a powerful and widely applicable tool for analyzing and interpreting large and complex data" (Sadiku et al., 2016). Data visualization is a multidisciplinary field that draws on psychology to understand how people perceive shapes and colors, computer science and statistics to develop data mining techniques, and graphic design to design infographics (Aparicio and Costa, 2015). Data visualization can change the way information is perceived (Aparicio and Costa, 2015). Furthermore, Eckert (2022) (Section 2.3) also mentions that appropriately visualized data helps one understand it more quickly and intuitively. He draws attention to the Gestalt principles that should be considered when visualizing data (Eckert, 2022). Sadiku et al. (2016) mention data visualization as a tool for analyzing data.

Munzner (2014) defines data visualizations as "visual representations of datasets designed to help people carry out tasks more effectively" (Munzner, 2014, chapter 1). Designing visualizations requires considering many trade-offs and limitations of computers, humans, and displays. Munzner (2014) suggests analyzing visualizations by answering questions about the audience's intentions, the data they see, and the design of the visualization. Her framework, guided by these questions, offers a structured guideline for using shapes, colors, graph types, and positions to build data visualizations. (Munzner, 2014)

Segel and Heer (2010) combine data storytelling and visualization in their book "Narrative Visualization: Telling Stories with Data". Their definition of a narrative aligns with that of Eckert (2022), who says that a narrative is present if data is explained. They also argue that Gestalt theory influences how visualizations are perceived. Outliers among visual features (e.g., color and size) attract the audience's attention. (Segel and Heer, 2010)

In summary, the literature shows that data visualization is an interdisciplinary field combining knowledge from psychology, computer science, statistics, and design to create clear and intuitive visual representations of data. Munzner (2014) framework provides a structured approach to designing visualizations by focusing on why they are needed, what data is being shown, and how they are designed. Segel and Heer (2010) adds to this by exploring how storytelling techniques, such as narrative structures and visual cues, can make data more engaging and easier to understand. Together, these works demonstrate how data visualization can effectively support analysis and storytelling.

2.3 Data Storytelling

Dykes (2019) describes data storytelling as the bridge between logic and emotion, with three core pillars: data, narratives, and visuals. These pillars complement each other in different ways to drive change (see Figure 3). "Essentially, data storytelling is a form of persuasion. It employs data, narrative, and visuals to help an audience see something in a new light and to convince them to act." (Dykes, 2019, p.32)

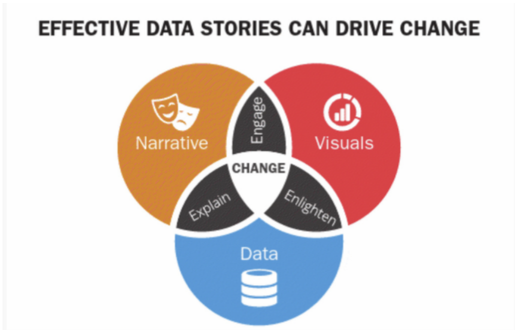


Figure 3: Complementing core pillars of data storytelling can drive change (Dykes, 2019)

Since data storytelling is a form of persuasion it depends on all of Aristoteles' modes of persuasion:

- Ethos - appeal to credibility
- Logos - appeal to logic or reason
- Pathos - appeal to emotion
- Telos - appeal to purpose
- Kairos - appeal to timeliness (Dykes, 2019, p.32)

Dykes (2019) argues that merely presenting information is insufficient to impress an audience. If someone already has a certain belief, simply showing them data that supports a different idea usually won't work. People prefer to stick with their current way of thinking because it feels complete and makes sense to them. This phenomenon is also known as confirmation bias (Nickerson, 1998). Without a clear, meaningful narrative to connect the new information, the data can feel disconnected and unconvincing. A narrative makes information easier to understand and more likely to influence perspective because emotion significantly influences decision-making. (Dykes, 2019, p.72)

According to Dykes (2019), the strength of a data story directly influences the likelihood of prompting the audience to take action. This process begins by capturing the audience's attention and helping them understand and remember the message. Ultimately, it inspires them to act (Dykes, 2019, p.36). Furthermore, Dykes (2019) introduces the concept of the "story zone," which identifies types of data for which storytelling is particularly effective. Data outside the story zone is either simple enough to understand without a narrative or requires an unreasonable amount of effort to create a story. The story zone encompasses medium-to-high-value insights that are also characterized as unpleasant, disruptive, unexpected, complex, risky, costly, or counterintuitive (Dykes, 2019, p.117).

A story is defined as a linear and connected series of events. Dykes (2019) argues for a narrative structure instead of the inverted pyramid structure often used in journalism. A narrative structure begins with an introduction, builds to a climax, and ends with a resolution that has emotional power. In general, Dykes (2019) argues that there are six essential elements to a data story (Figure 4).

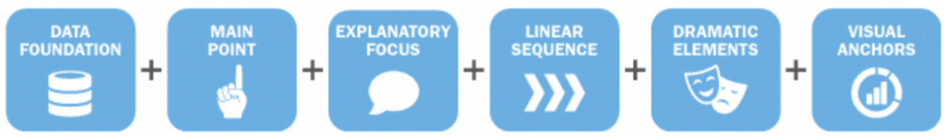


Figure 4: Six elements of a data story (Dykes, 2019)

In line with this, Ojo and Heravi (2018) explain that the goal of data storytelling is to inform, explain, persuade, or engage the audience with data using a narrative structure. Furthermore, Ojo and Heravi (2018) categorize data stories into seven distinct types, based on an analysis of award-winning data stories, according to their purpose and narrative focus. These types demonstrate the diverse applications of data storytelling, ranging from investigative journalism to public education. Each type aligns with specific communication goals, ensuring that data-driven narratives inform and resonate with their intended audiences. (Ojo and Heravi, 2018)

Additionally, the study by Zhang et al. (2022) shows that data storytelling is based on a narrative process that combines data visualization and entertainment techniques. The study points out that the data visualization narrative process identifies a target message (the main point in Figure 4) and emphasizes what needs to be communicated, rather than focusing purely on the data itself.

The book "Storytelling with Data" from Eckert (2022) aligns with the above and highlights that "Storytelling is so effective because it is deeply embedded in the structures of our brain" (Eckert, 2022, p.8) and has positive effects on memorizing the message, engaging and persuading the audience.

Furthermore, Eckert (2022) clearly defines that if data provides explanations, then it has a narrative structure. This is contrary to data being used for exploratory purposes, such as providing insight or being the topic. Additionally, Eckert points out that literature about storytelling focuses on the hero's journey, in which the hero embarks on a quest, wins battles, and discovers themselves in the process. This pattern is used as a model for storytelling. (Eckert, 2022, chapter 2)

Eckert (2022) argues that data storytelling needs a multidisciplinary skill set of analysts, designers and journalists. Knaflic (2015) in his book also called "Storytelling with data", underlines this multidisciplinarity further.

Weber et al. (2018) identify seven key features of data storytelling in journalism. Data storytelling in journalism uses data as the foundation for creating narratives, focusing on patterns, outliers, and correlations to uncover meaningful stories. Visualizations play a key role, either standing alone as complete stories integrated with text and multimedia to clarify complex information. Stories can follow different structures. One format is straightforward and linear, with an author-driven narrative in which the audience is presented with a story and cannot explore it independently. On the other hand, there are reader-driven, nonlinear, interactive formats that allow audiences to freely explore data. A combination of these two approaches is the hybrid approach, such as the Martini Glass Structure, where the story starts author-driven and then opens up, allowing the audience to interact with the data further. Lastly, there is scrollytelling, where the story is discovered as the audience scrolls down. Figure 5 shows a visualization of these four different storytelling structures. (Weber et al., 2018)

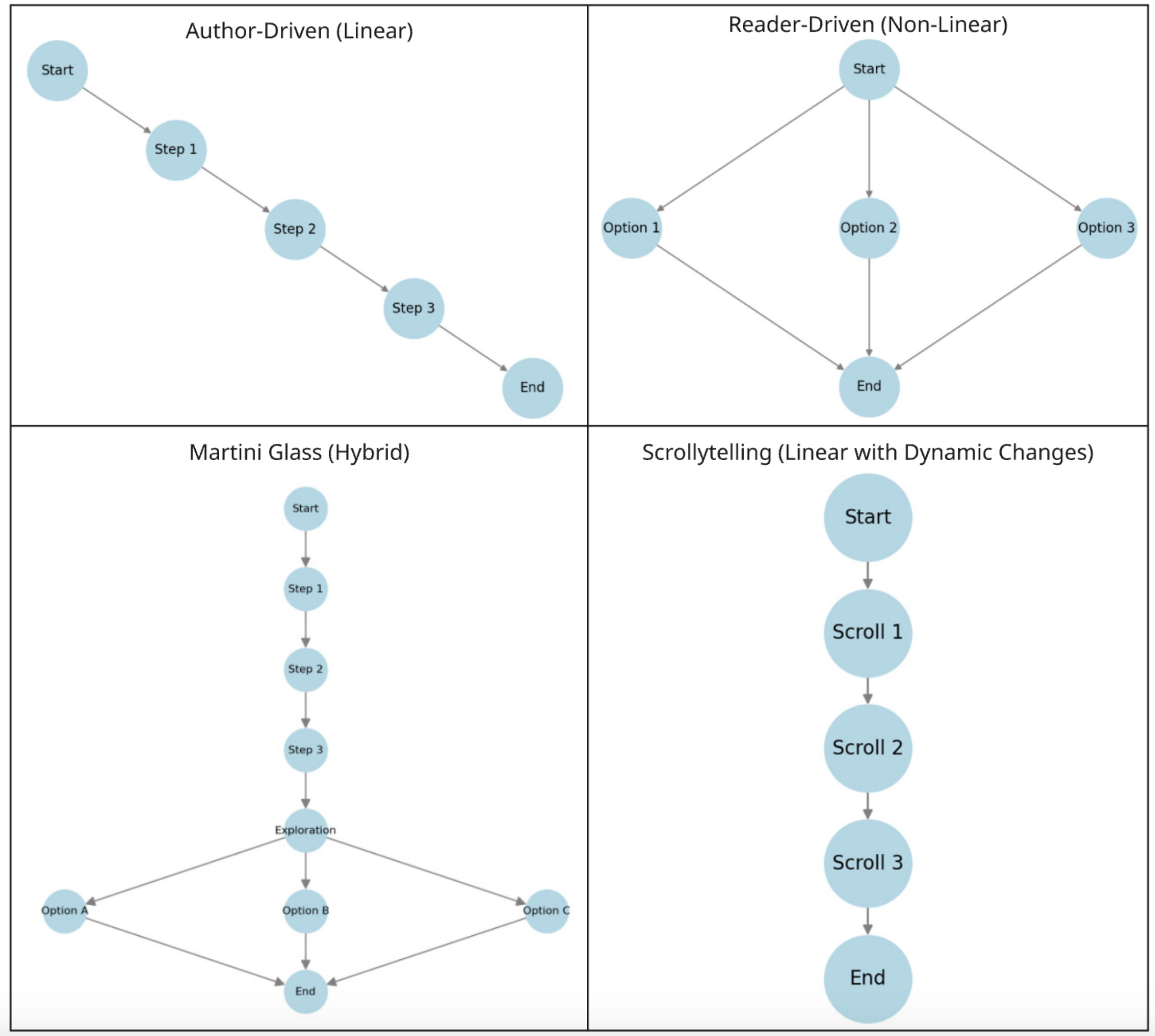


Figure 5: Four different storytelling structures (Weber et al., 2018)

A clear and attractive design is essential for drawing in audiences and building trust. Transparency is also critical, and explanations of data sources and methods help ensure credibility. Together, these elements reshape traditional journalism by making stories engaging and informative through the effective combination of data, visuals, and narrative. (Weber et al., 2018)

In general, to be able to tell stories with data, Knaflic (2015) shares six key lessons:

1. Understand the context (who, what, how)
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus attention where you want it (size, colour, position)
5. Think like a designer
6. Tell a story (Knaflic, 2015, Introduction)

Although Knaflic (2015) focuses on design principles and storytelling techniques, and Dykes (2019) focuses on six structural storytelling elements, the two topics are closely related. In order to lay the data foundation and identify the main point one needs to understand the context. One must choose an appropriate visual display without clutter to create the visual anchor. Focusing attention where desired ensures a linear story sequence. Thinking like a designer also impacts the visual anchor. Lastly, telling a story ensures the explanatory focus and dramatic elements of a data story.

To sum up, data storytelling is a powerful way to connect with audiences by combining data, narrative, and visuals. It helps people understand, remember, and act on complex information. Clear storytelling structures, good visual design, and transparency about data and methods are key to building trust and creating engaging stories. Since data storytelling requires skills from analysts, designers, and communicators, it is a multidisciplinary effort. These principles demonstrate how data storytelling transforms complex information into accessible and impactful narratives, making it a valuable tool in fields such as journalism and education.

2.4 Attitude

The paper by Albarracin and Shavitt (2018) is a review of different research on attitude and attitude change from 2010 to 2019. Since this study aims to measure attitudes and changes in attitudes, the papers mentioned in Albarracin and Shavitt (2018) that measure attitudes are reviewed in detail.

"An attitude is a predisposition to respond in a favo[u]rable or unfavo[u]rable manner with respect to a given attitude object" (Dijkstra and Goedhart, 2012). Attitudes refer to a specific target. Thus, they appear in various contexts, such as politics (attitude toward political parties), marketing (attitude toward products), and health (attitude toward medications) (Albarracin and Shavitt, 2018). "Attitudes are of particular concern in the area of climate change, where scholars and practitioners are investigating the potential for reducing climate change denial" (Albarracin and Shavitt, 2018). Attitudes can be specific or general and can be positive or negative. Some attitudes are implicit and based on memory and experience, while others are constructed in a situation, e.g., when receiving information about an object (Sheets et al., 2011). Attitudes can be stable, but they can also change, e.g., from favor to disfavor over time.

Ehret et al. (2015) state that attitudes can be measured by asking participants to self-report their attitudes (which is only possible for explicit attitudes) or by observing their reactions to attitude objects. Sheets et al. (2011) presents a indirect way of measuring attitudes using Implicit Association Tests, which measure mental associations between pairs of concepts. The study found that implicit attitudes played a significant role, even influencing behavior outside of participants' awareness (Sheets et al., 2011). Implicit measures might be particularly insightful for capturing attitude change when unconscious biases are at play, whereas explicit measures are better for deliberate, reflective changes (Sheets et al., 2011). Thus, the best approach often combines both methods to gain a comprehensive understanding of attitudes and changes in attitudes. Dijkstra and Goedhart (2012) use statements with Likert-scale answering options to measure attitudes toward a topic. For instance, attitudes toward school science are measured using a statement such as, "I would like to do less science at school," which respondents rate on a Likert scale ranging from "strongly disagree" to "strongly agree" (Dijkstra and Goedhart, 2012).

Albarracin and Shavitt (2018) mention that changes in implicit attitudes can influence behavior, though this is not always the case. "Since the early 1900s, psychologists have been interested in the attitude–behavior relationship, in understanding whether and how attitudes guide behavior" (Bechler et al., 2021). Views on the strength and nature of this relationship vary among researchers. While the majority found a linear relationship, others have different perspectives (Figure 6) (Bechler et al., 2021).

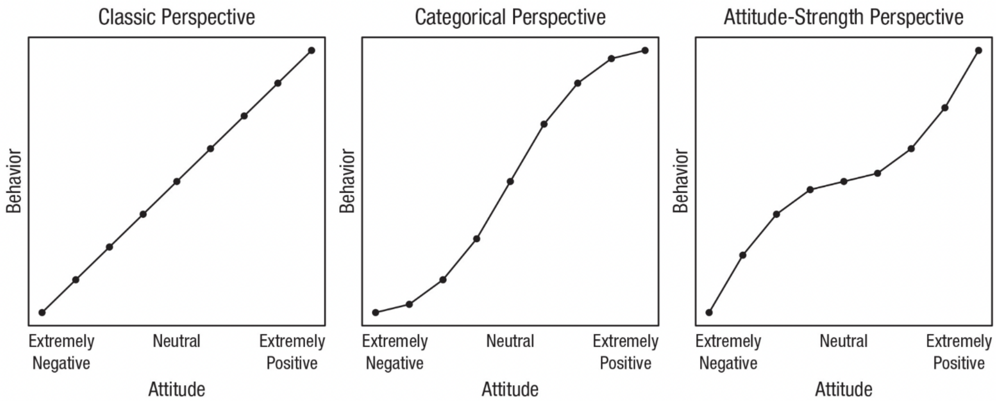


Figure 6: Classic, categorical, and attitude-strength perspectives on the attitude–behavior relationship (Bechler et al., 2021)

Figure 6 shows the classic perspective of a linear relationship between attitude and behavior. This means that if behavior becomes more positive, attitude becomes more positive to the same extent. Additionally, Bechler et al. (2021) introduce the categorical perspective, which reflects how people naturally group and perceive information into distinct categories, and its influence on attitudes and behaviors. This perspective assumes that changes in attitude and behavior are greater when moving from positive to negative than when moving from positive to more positive. The Attitude-Strength perspective proposes that the impact of attitudes on behavior increases as attitudes become more extreme, forming a cubic relationship. However, there is limited direct evidence supporting this nonlinear relationship. The findings of Bechler et al. (2021) study show that the relationship between behavior and attitude is categorical. This may explain why past studies often found weak connections between the two. The study also shows that shifts between negative and positive attitudes have a greater impact on behavior, with positive attitudes influencing behavior more strongly than negative ones. (Bechler et al., 2021)

Furthermore, Glasman and Albarracín (2006) conducted a meta-analysis of 128 studies and found that attitudes influence behavior more when they can be easily recalled and are stable over time. Additionally, the relationship between attitude and behavior strengthens with repeated exposure to the attitude object. The researchers also found that individuals who are motivated to think deeply about an issue are more likely to focus on information that strongly supports a single perspective related to the issue. This information directly relates to the behavior they might exhibit. Consequently, their attitude becomes more focused and solid, making it more likely to influence their actions. (Glasman and Albarracín, 2006)

According to the Theory of Planned Behavior, the stronger one's intention to behave in a specific way, the more likely one is to act accordingly. Intention encompasses motivational factors and is influenced by attitude, subjective norms, and perceived behavioral control. Attitude is one factor that can lead to behavior, but other factors influence behavior as well. (Ajzen, 1991)

The paper of Briñol et al. (2019) summarizes past research on attitude change and persuasion. It mentions the Elaboration Likelihood Model (ELM) by Petty and Briñol (2011), a dual-processing model that explains how attitudes change and form (Figure 7).

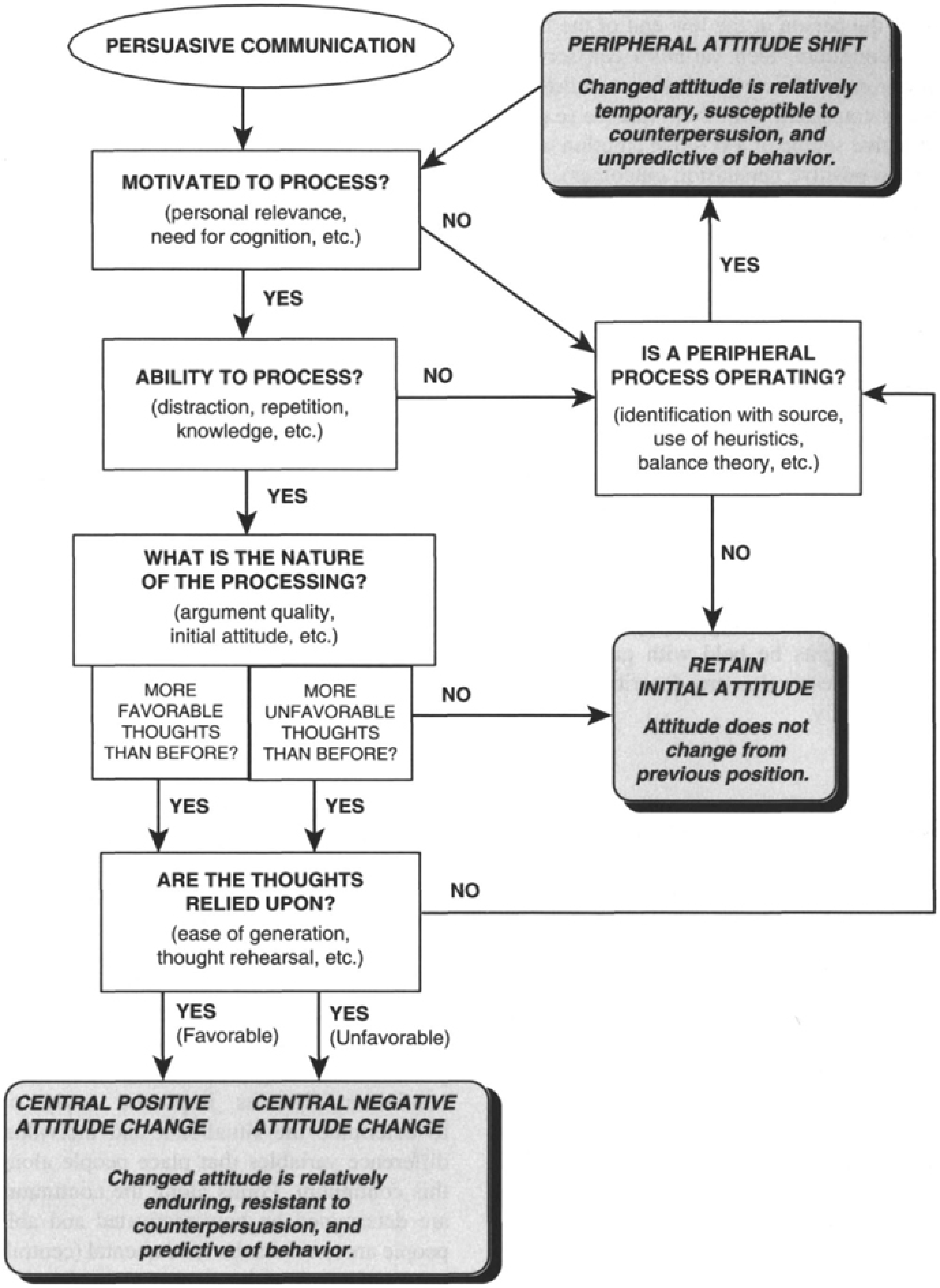


Figure 7: Dual process of the Elaboration Likelihood Model (Petty and Briñol, 2011)

The model incorporates four main ideas. First, changing people's attitudes can be achieved with either a low or high level of thought (elaboration). Second, as can be seen from Figure 7, if elaboration is low, it takes the peripheral route; if elaboration is high, it takes the central route. In reality, persuasion is achieved through a combination of peripheral and central processes. Furthermore, attitudes created through high thought processes last longer. Finally, certain factors influence how people think about a message. The same factors (e.g., attractiveness, emotions, and credibility) can have different effects depending on how deeply people think. When people think deeply, persuasion works in a more complicated way. Factors such as emotions and credibility not only act as quick signals, but also influence how people analyze information and how strongly they believe in their opinions. These factors are sometimes theories of psychology in themselves. (Petty and Briñol, 2011)

A study by Liem et al. (2020) compares the effectiveness of different visual data storytelling techniques in changing attitudes. The study examined an interactive flow map, in which participants explored the data independently (control condition); a visual-narrative design, in which different personas displayed on the screen told their stories in the first person until the flow map was shown; and another visual-narrative design, in which a third-person narrative explained the flow map. There is no evidence that the visual-narrative designs evoke more positive attitudes than the control condition.

Another study examined the impact of different data visualizations (tables versus graphs) on attitudes. The study concludes that the persuasiveness of a visualization depends on the strength of the initial attitude. Charts lead to greater attitude change when initial attitudes about the topic are weak, while tables are more persuasive when initial attitudes are strong. The study mentions that the results need further validation. (Pandey et al., 2014)

In summary, attitudes are tendencies to respond positively or negatively to specific objects or ideas. They appear in various contexts, such as politics, health, and education. Attitudes can be either explicit, meaning they are consciously held, or implicit, meaning they are unconscious. They may also remain stable or change over time. Methods to measure attitudes include self-reported surveys for explicit attitudes and implicit association tests for implicit biases. Depending on different variables of the attitude, attitudes can influence behaviors to different degrees. This is important to understand when measuring and assessing attitudes. The ELM provides an overall picture of how attitudes change and how this process can be influenced, which makes it possible to conduct a more detailed analysis of individual steps and theories. In relation to data storytelling, to the best of current knowledge, no impact of visual-narrative designs on attitudes has yet been shown. Data visualization seems to influence attitude change based on the strength of the initial attitude.

2.5 Recall

Recall is, "in psychology, the act of retrieving information or events from the past while lacking a specific cue to help in retrieving the information" (Britannica, 2024).

Buckland and Gey (1994) further specify recall as the proportion of retrieved events from all relevant events ($N_{\text{ret}\cap\text{rel}}/N_{\text{rel}}$). His research focuses on the relationship between recall and precision, which is a trade-off. Thus, the higher the recall rate, the lower the precision, and vice versa. (Buckland and Gey, 1994)

Recall is also mentioned in a learning context (Abbott, 1909; Schöner, 1989; Carpenter et al., 2006). Early research by Abbott (1909) concludes that recall is beneficial for the learning process. Findings of Carpenter et al. (2006) underline this by showing that actively testing oneself in the learning process is beneficial, suggesting that the retrieval practice strengthens the memory process in a way simple restudying can not.

Hogan and Kintsch (1971) define recall in line with the definition in Britannica (2024). The study compares the effects of study trials (passive exposure to information; participants are shown 40 words, one at a time, for two seconds each) and test trials (participants are given 100 seconds to remember as many words as possible) on long-term recognition and recall performance. The latter is a recall test in which participants must retrieve the words they saw from memory without receiving any prompts (e.g., multiple choice). The study validates the two-stage theory of recall proposed by Kintsch (1970). This theory suggests that, in order to recall information, it must first be retrieved, and then recognized as correct or relevant, based on how often an item has been seen before. The study shows that items that were not recognized were rarely recalled, and items that were recalled were also recognized. (Hogan and Kintsch, 1971)

In a more recent study from Laming (2021) participants were shown advertisements consisting of a brand, picture, and slogan. After different time periods, the participants were shown one of the three components and asked to match the other two components to it. They are given a complete list of all three categories because the goal of the experiment is to test recall of the associations, not the exact brand or slogan. The research shows that recall performance declined with increasing time. Additionally, some ad types were recalled significantly better than others, suggesting that prior exposure or personal relevance influences recall. (Laming, 2021)

A study by Zdanovic et al. (2022) investigated the impact of data storytelling visualizations on recall. No significant differences were found between data visualizations and data storytelling. However, the study identified indicators of chart types and prior knowledge about the topic that affect recall (Zdanovic et al., 2022). In contrast, a study by Shao et al. (2024) shows that data storytelling leads to more accurate recall of insights. This discrepancy may be due to the different study designs (between-group vs. within-group) and visualization techniques employed. Shao et al. (2024) also found out that data literacy is key to getting the most out of the data. Nevertheless, data storytelling helps all people, regardless of their data literacy levels.

In conclusion, recall is defined as the ability to retrieve information without specific cues. Its measurement is often framed as the proportion of retrieved events among all relevant events. Research highlights the crucial role of recall in learning, demonstrating that active retrieval practice enhances memory retention more effectively than passive review. Recall is usually measured in experimental setups where participants retrieve information from memory without prompts, as in studies comparing test trials to simple exposure. Furthermore, findings suggest that prior exposure and personal relevance significantly influence recall performance. This reinforces the idea that familiarity and experience shape an individual's ability to retrieve information. However, the results of the impact of data storytelling on recall differ across studies and require further investigation. However, the type of chart and prior knowledge seem to affect recall.

2.6 Data Literacy

Frank et al. (2016) define data literacy as the ability of non-experts to make use of data especially in the context of the internet. The research of Gummer and Mandinach (2015) focuses on developing a data literacy framework for teaching and clearly specifies the abilities involved in using data, including collection, examination, analysis, and interpretation, with the goal of enabling decision-making in an educational context. Bonikowska et al. (2019) reviews more definitions of data literacy and concludes that "a data literate individual would, at minimum, be able to understand information extracted from data and summarized into simple statistics, make further calculations using those statistics, and use the statistics to inform decisions. However, this definition is context-dependent" (Bonikowska et al., 2019).

The paper of Bonikowska et al. (2019) summarizes different data literacy frameworks and approaches to measuring data literacy. These frameworks cover the self-assessment of individuals' data literacy, objective measurement of data literacy, and the self-assessment of organizational data literacy (Bonikowska et al., 2019). Cui et al. (2023) also review different ways of assessing data literacy for professionals, K-12 students, researchers, and data librarians, since existing assessments are domain-specific. Most of the reviewed assessments are self-assessments with Likert-scale answers (Cui et al., 2023). The questions touch on different abilities, ranging from graph literacy and knowledge skill to advanced skills, such as data transformation and correlation (Cui et al., 2023).

Since data literacy is assessed differently across domains with varying data usage requirements, the various assessments and frameworks reviewed in Bonikowska et al. (2019) and Cui et al. (2023) can inspire the development of a tailored data literacy assessment suited to the research needs that focus on the general population.

2.7 Medium

Mediums can be modalities, channels, or devices through which people communicate. Oba and Berger (2024) suggest that they have an impact on what is communicated. Although there are many different types of mediums, their effects can be simplified into two categories, deliberation and audience salience. The amount of thought a communicator puts into a message directly affects the message. Additionally, the presence of the communicator and synchronicity directly influence the salience of the communicator and, consequently, the message's content. For example, a medium like writing gives the communicator more time to deliberate, resulting in a more organized, thoughtful, and normative message. Conversely, a medium such as speaking leads to greater audience salience and a more concrete, honest, and emotional content style. (Oba and Berger, 2024)

A study by Song et al. (2023) found that people prefer different media depending on their level of expertise on a given topic. Experts are more likely to use text content, whereas novices prefer multimedia content, which enhances usability, reading comprehension, and time on task. (Song et al., 2023)

A cognitive theory of multimedia learning by Mayer (2005) explains why individuals learn more effectively through a combination of words and pictures than through words alone. The theory is based on three assumptions:

1. Dual-channel assumption: visual and auditorial information is processed in two separate channels
2. Limited capacity assumption: the amount of information that can be processed in each channel at the same time is limited
3. Active processing: Learning is an active process where individuals engage in selecting, organizing, and integrating information based on prior knowledge (Mayer, 2005)

Following these assumptions, twelve multimedia principles are introduced (Figure 2).

Principle	Explanation
Multimedia Principle	People learn better from a combination of words and pictures rather than words alone.
Coherence Principle	Extraneous information should be minimized to avoid cognitive overload.
Signaling Principle	Cues (e.g., highlights, arrows) help direct attention to key content.
Redundancy Principle	Narration with visuals is better than visuals with added redundant text.
Spatial Contiguity Principle	Related text and images should be placed close together to aid learning.
Temporal Contiguity Principle	Words and images should be presented simultaneously rather than separately.
Segmenting Principle	Information should be broken into smaller, learner-controlled segments.
Pre-training Principle	Teaching key terms and concepts beforehand improves comprehension.
Modality Principle	Presenting information via both auditory and visual channels enhances learning.
Personalization Principle	Conversational, informal language fosters better engagement and understanding.
Voice Principle	A human voice in narration is more effective than a robotic or synthetic voice.
Image Principle	Adding an instructor's image on the screen does not necessarily improve learning.

Table 2: Principles of Multimedia Learning (Mayer, 2005)

In summary, different communication mediums influence how messages are shaped and understood. Some mediums allow for more careful thought and organization, while others encourage more immediate and emotional responses. Research also shows that experts tend to prefer text-based content, while multimedia helps beginners understand information better. According to the theory of Mayer (2005), combining words and visuals makes learning more effective. These findings underscore the importance of selecting the appropriate medium to enhance clarity and engagement in communication. These findings are important for effectively designing the medium of data storytelling.

2.8 Research Gap

While data storytelling is widely recognized as a powerful method to engage audiences and communicate complex information (Dykes, 2019; Eckert, 2022; Knaflic, 2015), existing research primarily focuses on its structural components such as narrative frameworks and visual design (Segel and Heer, 2010; Weber et al., 2018), rather than the medium through which the story is communicated. Yet, as Oba and Berger (2024) note, the medium fundamentally shapes the message by influencing both the deliberation of the communicator and the salience of the audience.

Recent studies suggest mixed results on the actual impact of data storytelling on recall. Zdanovic et al. (2022), for example, found no significant difference in recall between data storytelling visualizations and traditional ones. Meanwhile, Shao et al. (2024) indicate storytelling may help users recall insights more accurately, but these results vary depending on user characteristics like data literacy. These contradictory findings point to a need for deeper exploration on if the medium influences outcomes like recall and attitude change.

Moreover, data literacy has emerged as a moderator in understanding how users interact with data (Frank et al., 2016; Bonikowska et al., 2019; Cui et al., 2023). Yet, despite its importance, few studies systematically examine how data literacy affects the effectiveness of different storytelling mediums. While Song et al. (2023) suggest people with low previous knowledge benefit more from multimedia content and experts prefer text, the role of data literacy in shaping attitude and recall of data storytelling remains unexplored. Although multimedia learning theory by Mayer (2005) outlines how people process verbal and visual information, it has not yet been fully applied to the context of data storytelling.

By addressing these gaps, this study contributes to a better understanding of how different data storytelling formats affect audiences, particularly in terms of attitude and recall. The findings offer practical implications for designing more effective data-driven communication.

3 Methodology

In this section, the methodology of the research is described thoroughly. This includes the approach, design and implementation of the research process.

3.1 Philosophy of Science

"Our assumptions and views about how knowledge should be produced and about the nature of the social world heavily influence the research process." (Clark et al., 2021, chapter 1) Therefore, this section discusses the stance on the philosophy of science of this study.
The ontological stance of the research is that a social reality exists which can be measured, though limitations occur due to methodological and contextual factors. Although knowledge is created by aiming for objectivity, interpretations of the researcher (e.g., survey design) influence the research design. Thus, the epistemological stance is modified objectivism. Methodologically, the study uses a modified experimental approach that combines hypothesis-driven quantitative methods with attention to contextual variables in order to explain and generalize insights rather than to understand individual cases (Figure 8). (Pickard, 2013, chapter 1)

	Postpositivism
Ontological stance	Critical realism
Epistemological stance	Modified objectivist
Methodological stance	Modified experimental
Purpose	Prediction/control/explanation

Figure 8: This studies research paradigms (Pickard, 2013, chapter 1)

The experimental procedure begins with the identification of a research hypothesis based on existing literature and theories. Then, the research takes a deductive approach. (Clark et al., 2021, chapter 7)

3.2 Research Design

This study uses an experimental design to explore how different storytelling formats affect changes in people's attitudes and recall (Berger et al., 2002, chapter 1). This approach allows for the systematic manipulation of the independent variable while holding other factors constant. It increases the study's internal validity, which is one of the three key criteria in social research design. The study randomly assigns participants to different groups and measures the dependent variables to isolate the specific impact of the storytelling format. The experimental setup also makes the study easier to replicate and helps to compare results across groups. This supports the goal of understanding not only whether, but also how, storytelling influences attitudes and recall over time. (Clark et al., 2021, chapter 3)

To structure the methods applied in the experiment and its design, the methods are structured in line with the experimental procedure by Lazar et al. (2017), Chapter 3: "

1. Identify a research hypothesis.
2. Specify the design of the study.
3. Run a pilot study to test the design, the system, and the study instruments.
4. Recruit participants.
5. Run the actual data collection sessions.
6. Analyse the data.
7. Report the results."

The last step, "Report the results" won't be covered in this section, a separate section is dedicated to results and their analysis (Section 4).

3.3 Research Hypotheses

To examine if and how different modes of communicating data effect a change in attitude and recall of information, this section derives research hypotheses from the research questions.

RQ1: How might audiences' attitudes change after being presented with data storytelling communicated through different mediums?

Data storytelling is a form of persuasion (Dykes, 2019, p.32). Persuasive communication can be the trigger to a changing attitude of the audience of the communication. Therefore, the first hypothesis is:

H1: audiences' attitude changes after being presented with data storytelling.

There is no existing research on the different impacts of written versus verbal data storytelling. However, the modality principle by Mayer (2009) suggests that learning is more effective from pictures and spoken words than from pictures and text because the latter overloads the visual channel. Learning increases the probability of an attitude change (Petty and Briñol, 2011). Additionally, a medium like speaking supports a more emotional content style, which leads to higher persuasiveness and thus a higher probability of attitude change (Oba and Berger, 2024; Dykes, 2019; Petty and Briñol, 2011). Therefore, another hypothesis is:

H1.1: Audiences' attitudes are more likely to change if they are presented with verbal data storytelling than written one.

RQ2: How might audiences' recall differ when being presented with data storytelling communicated through different mediums?

"When multiple areas of the brain are engaged, the hippocampus—which stores short-term memories—is more likely to convert the experience of hearing a story into a long-term memory" (Cote, 2021). Thus, the research hypothesis is:

H2: Audiences' recall improves after being presented with data storytelling.

In addition to that, an experiment by Moreno and Mayer (1999) reveals that students who listened to narration alongside animations scored higher on retention tests compared to those who read on-screen text with animations. The following research hypothesis is:

H2.1: Audiences' recall is more likely to improve when presented with verbal data storytelling compared to written one.

RQ3: How might the audiences' level of data literacy influence their ability to recall information they were presented with?

Schema Theory suggests that audiences with higher data literacy have more developed schemas for understanding data, meaning they can process data more effectively, leading to stronger recall (Pankin, 2013). Following:

H3: Audiences with a high data literacy are more likely to recall information they were presented with across all modes of data communication.

RQ4: How might the audiences' level of data literacy influence their attitude change towards the information they were presented with?

Applying the ELM people with higher data literacy are more likely to engage in central processing, deeply analyzing and integrating the data-driven message that leads to long-term attitude change (Petty and Briñol, 2011). Those with lower data literacy may rely more on peripheral cues, leading to a temporary attitude change (Petty and Briñol, 2011). The hypothesis is:

H4: Audiences with high data literacy are more likely to change their attitude in the long term compared to audiences with low data literacy.

Subsequently, the following hypothesis model can be drawn:

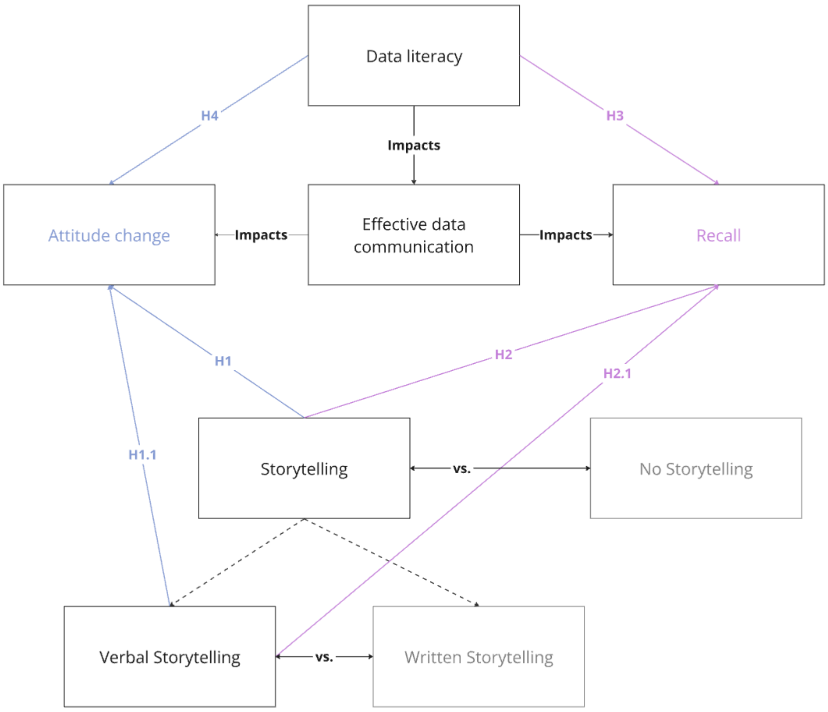


Figure 9: Hypothesis model for effective data communication

3.4 Additional Hypotheses

In addition to the hypotheses related to the research questions, further hypotheses are summarized. The research is expected to show relationships between the following attributes (Figure 10).

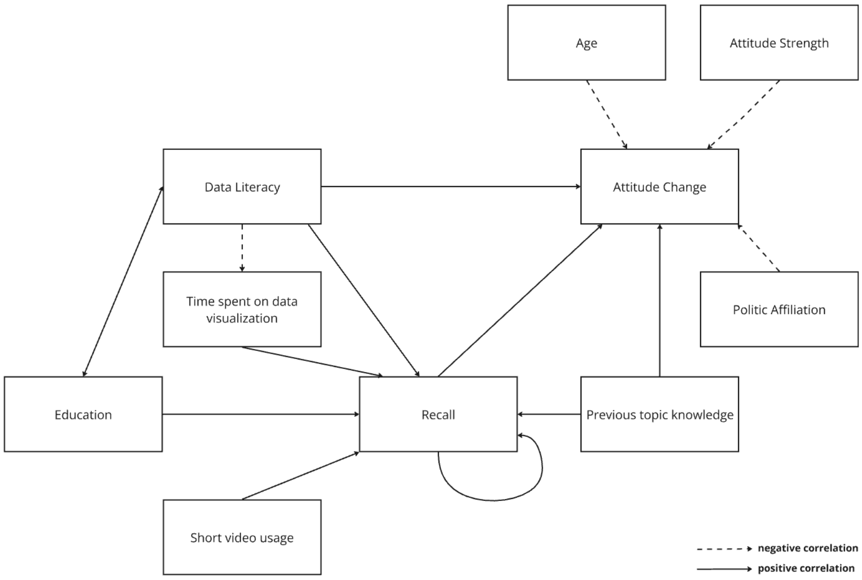


Figure 10: Expectations of the relationship between different attributes that are measured. The Relationships are visualized with an arrow.

The country, gender, and topic are not expected to influence any of the other measured attributes and are therefore not shown in Figure 10. Attitude change is expected to be greater for younger participants because they are more likely to change their point of view than older participants, who tend to become more resistant to change over the course of their lives (Ksiazkiewicz et al., 2020). Additionally, politic affiliation can also have an influence on the resistance to change, as people with strong political views are less likely to change their attitudes due to a higher likelihood of cognitive rigidity (Zmigrod, 2020). Extreme attitudes from the beginning are expected to be less likely to change due to confirmation bias (Nickerson, 1998).

Recall is expected to be influenced by previous topic knowledge, as for people with previous knowledge it is easier to recall information because they can put it in their existing schema and retrieve it more easily (Pankin, 2013). The argument is the same for previous experience with watching short videos. It is also expected that recall will decrease over time due to the psychological principle of decay, which states that memories disappear unless they are actively maintained over time (NeuroLaunch editorial team, 2024; Laming, 2021). This is also why a better recall is expected to increase attitude change, as a change in long-term attitude requires actively maintaining the information. Furthermore, higher education and higher data literacy are expected to be related because people with higher education are more likely to be educated about data. Additionally, the more time the audience spends on the data visualization, the higher their recall because they have more time to memorize the information (Goldstein et al., 2011).

3.5 Experimental Design

There is one independent variable: storytelling medium. Two conditions of storytelling medium will be examined: verbal and written, additionally, No storytelling as a control condition is added (Lazar et al., 2017, Chapter 2). To minimize the probability that the results depend on the chosen topic for data storytelling, another independent variable, topic, is introduced. Two topics are prepared in the same way. In conclusion, the experiment designed for this study contains six conditions (three conditions (verbal, written, No storytelling) x two topics). Furthermore, randomization is used to assign participants to one of the six conditions. This makes the following experiment a true experiment. (Lazar et al., 2017, Chapter 3)

There are two independent variables in the experiment, making it a factorial design. A between-group, within-group, or split-plot design can be applied (Lazar et al., 2017, Chapter 3). In this case, a between-group design is most suitable to avoid the learning effect on participants (Charness et al., 2012). If participants are first presented with Written storytelling and then Verbal storytelling about the same subject, the first exposure could influence the second in terms of previous knowledge. Additionally, repeated exposure to the same information can influence the ability to process information (Petty and Briñol, 2011). One could argue that the participants can be exposed to two or three different conditions about different topics (factorial experimental design). This would mean a large amount of information for participants to process, as they would be exposed to three different pieces of information, which could lead to fatigue during the experiment (Lazar et al., 2017, Chapter 3). Therefore, six groups of participants are required. To account for potential differences among participants in a between-group design, a pre-questionnaire is included to gather information that may influence the results. (Lazar et al., 2017, Chapter 3).

3.5.1 Structure

The structure of the experiment is described to ensure transparency, thus the possibility to replicate the experiment which is a criterion of quality in social research (Clark et al., 2021, chapter 3).

Participants who pass the pre-screening and are eligible for the experiment are randomly assigned to one of two experimental topics (Figure 11). Every step of the process shown in Figure 11 is implemented in Gorilla, a tool that helps build an experimental workflow where rules for automatically assigning participants to conditions and tasks, as well as other rules (e.g., interventions, delays), can be specified (Gorilla, 2024).

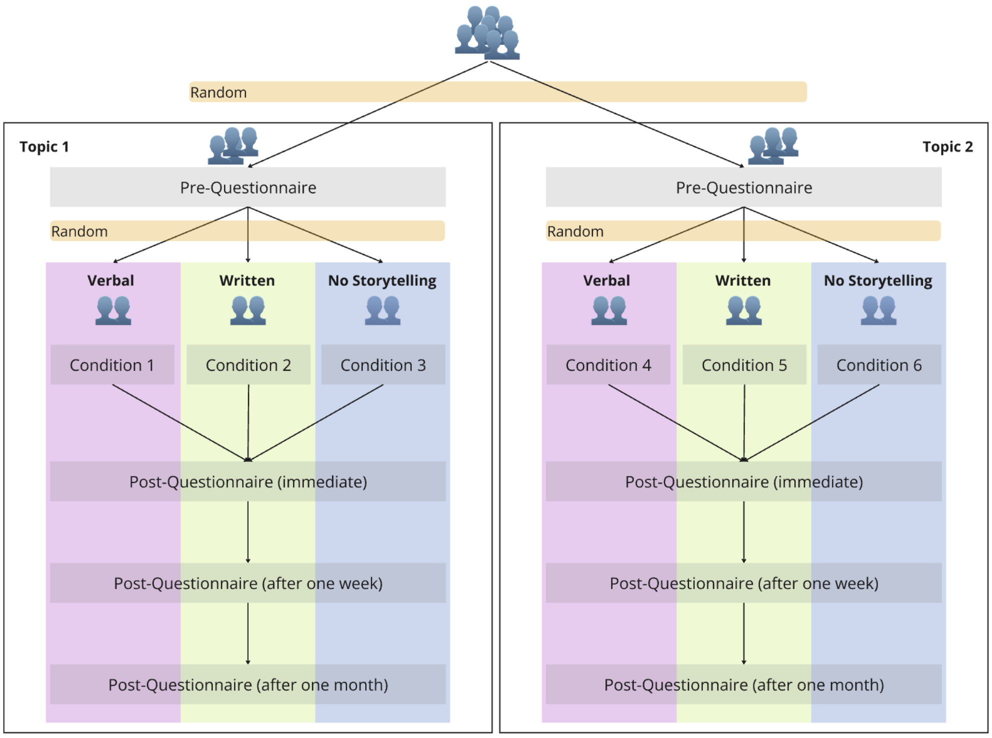


Figure 11: Experimental design structure. It is a between-group design across six different conditions.

After being randomly assigned to one of two topics, participants receive a pre-questionnaire containing demographic questions, topic-related questions, and questions examining their attitudes toward the topic and their data literacy. The reasoning behind the questions and their design are explained in detail in Section 3.8. After completing the pre-questionnaire, the participants are further randomized into three conditions per topic. At this stage, they are exposed to information about the topic in either a verbal or written story, or no story format. Immediately following the exposure, they are asked to complete another post-questionnaire to assess recall and attitude changes. Since recall and attitude can change over time, the same post-questionnaire is sent to participants again one week and one month after exposure (Laming, 2021; Petty and Briñol, 2011). This is a longitudinal study in which data are captured at least twice over time (Clark et al., 2021, chapter 7.3).

3.5.2 Design of Conditions

First, the two topics with existing data to focus on must be defined to design the six conditions of the experiment. Then, the visualization must be designed and a narrative developed. The following describes these steps.

Topic

To find a topic that is suitable for data storytelling it needs to meet certain requirements. First, the topic should fit into the "story zone" defined by Dykes (2019) (Figure 12).

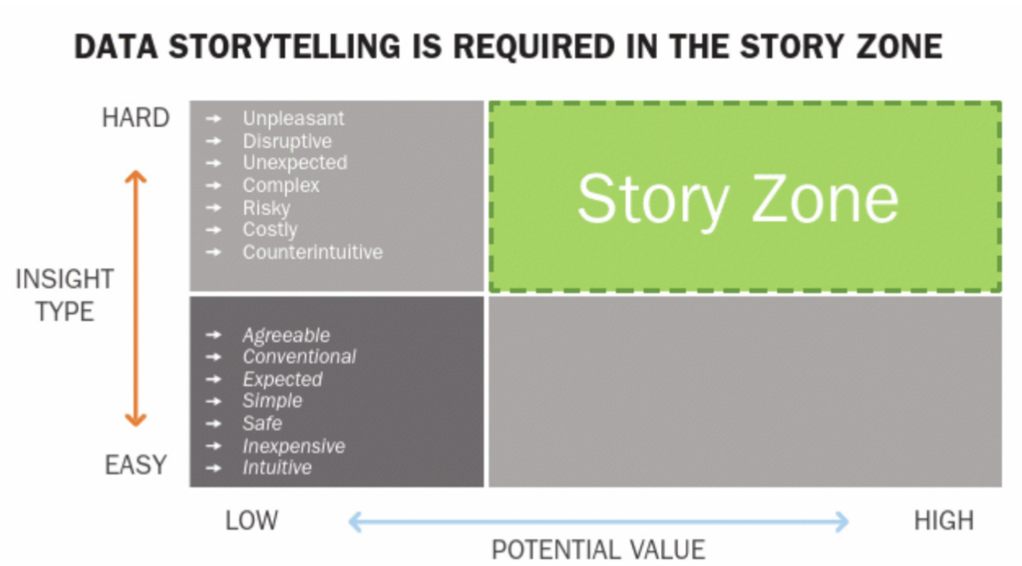


Figure 12: Story zone that describes which data is suitable for storytelling (Dykes, 2019)

Second, there needs to be available and reliable data on the topic of storytelling from which visualizations can be built. Additionally, since attitudes and changes in attitudes toward the topic should be measured, the topic should be something that can elicit an attitude (Dijkstra and Goedhart, 2012). Finally, to prevent results that are too dependent on one topic, two different topics that do not fall into the same category are chosen for the experiment.

To find reliable and available data, Kaggle, a website offering high-quality public datasets, is used (Kaggle, 2024). Additionally, Statistics Netherlands and similar government institutions have reliable, documented data on various topics of interest to their citizens (CBS, 2025). The latter is excluded because it is limited to one country. On Kaggle, five quality datasets suitable for data visualization are identified:

Topic	Description	Story Zone	Attitude Object	Source
Global Health Indicators	Dataset includes key health indicators for over 200 countries	Unpleasant	/	Qurban (2024)
Fish and Overfishing	Dataset includes data on how the Global Fish production is done and managed throught the years	Unpleasant	Fish consumption	Geukjian (2025)
Plastic Pollution	Dataset includes data on global plastic pollution and waste management	Unpleasant	Plastic waste management	Ritchie et al. (2023)
Social Media Usage and Emotional Wellbeing	Dataset includes data about different Social Media platforms and their impact on emotions	Unexpected	Social Media usage	BULUT (2024)
Titanic Passenger deaths	Dataset holds data about titanic passenger attributes and if they died or not	Unpleasant	/	Yasser (2022)

Table 3: Set of five quality datasets for visualization evaluated against topic requirements

Table 3 shows that Titanic passenger data is not suitable for the experiment. It is data from a century-old tragedy, and, although unpleasant, it is unlikely to affect people's attitudes today. Additionally, global health indicator data shows information about various health indicators, but it is too abstract to influence attitudes. On the other hand, data about fish and overfishing, plastic pollution, and social media usage and emotional well-being fall into the Story Zone and can be an attitude object. Since fish and overfishing and plastic pollution are both environmental topics, only plastic pollution is chosen for the experiment, along with social media usage and emotional well-being.

Data Visualization

To design the visualization that is used in the control condition as well as the basis for the two storytelling visualizations, the framework from Munzner (2014) is applied. In the framework, the visualization process is guided by three questions that are asked iteratively: "

- **What** data is shown in the view?
- **Why** is the task being performed?
- **How** is the viz idom contructed in terms of design choices?" (Munzner, 2014, chapter 1)

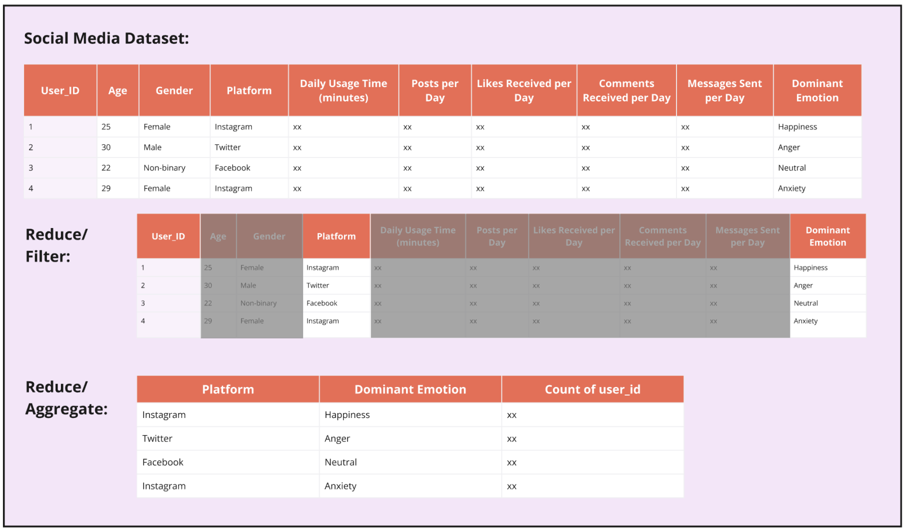
What is the data that is used to build the visualizations. In the following table the dimensions Munzner (2014) describes are listed and the characteristics of each of the two datasets are defined. This is done because the type of data can have an influence on the design of the visualization (Munzner, 2014, chapter 2).

What	Category	Plastic Waste Dataset	Social Media Dataset
Datasets	Data Types	Items	Items
	Dataset Types	Tables	Tables
	Dataset Availability	Static	Static
Attributes	Attribute Types	Categorical, Quantitative	Categorical, Quantitative
	Ordering Direction	Sequential	Sequential

Table 4: Description of full datasets according to "What" by Munzner (2014)

Table 4 describes the full raw data (BULUT, 2024; Ritchie et al., 2023). However, the goal (**Why**) of the visualization is not to visualize every aspect of the raw data, but rather to identify one interesting aspect suitable for data storytelling that has the potential to influence audiences attitudes. The **Why** consisting of actions and targets, is the same for both datasets. Both datasets show attributes that are categorical and quantitative. To tell a story about the data, different categories are compared, and the goal is to identify and display extremes. The dataset on plastic waste management should show the audience how plastic waste is managed in different regions and compare the regions with each other. The social media data should provide insight into which emotions dominate on which platforms and compare them. (Munzner, 2014, chapter 3)

This leads to the **How**, meaning how to visualize the data in the end. Figure 13 shows how the social media data is filtered and reduced to end up with the necessary data for visualization (Munzner, 2014, chapter 3). The according visualization for the plastic waste dataset can be found in Appendix A.



Social Media Dataset:

User_ID	Age	Gender	Platform	Daily Usage Time (minutes)	Posts per Day	Likes Received per Day	Comments Received per Day	Messages Sent per Day	Dominant Emotion
1	25	Female	Instagram	xx	xx	xx	xx	xx	Happiness
2	30	Male	Twitter	xx	xx	xx	xx	xx	Anger
3	22	Non-binary	Facebook	xx	xx	xx	xx	xx	Neutral
4	29	Female	Instagram	xx	xx	xx	xx	xx	Anxiety

Reduce/ Filter:

User_ID	Age	Gender	Platform	Daily Usage Time (minutes)	Posts per Day	Likes Received per Day	Comments Received per Day	Messages Sent per Day	Dominant Emotion
1	25	Female	Instagram	xx	xx	xx	xx	xx	Happiness
2	30	Male	Twitter	xx	xx	xx	xx	xx	Anger
3	22	Non-binary	Facebook	xx	xx	xx	xx	xx	Neutral
4	29	Female	Instagram	xx	xx	xx	xx	xx	Anxiety

Reduce/ Aggregate:

Platform	Dominant Emotion	Count of user_id
Instagram	Happiness	xx
Twitter	Anger	xx
Facebook	Neutral	xx
Instagram	Anxiety	xx

Figure 13: Reduction of the social media dataset to prepare them for the visualization

Since the goal is to compare data points, the data is encoded further by aligning it and using color and size to distinguish between different categories and make comparisons easier. To find the most intuitive way to visualize the data and show the insights, the guide of Wilke (2019) is used in the following.

The two datasets consist of "numerical values shown for some set of categories" (Wilke, 2019, chapter 5). Bars are used to visualize them. If there are multiple sets of categories the bars can be grouped or stacked to show different proportions (Wilke, 2019, chapter 5). According to Cleveland and McGill (1984), grouped bar charts have the lowest error rate because they share a common baseline and facilitate comparison. However, grouped bars display a lot of information and can be difficult to read, particularly when there are many categories. For data including many conditions, stacked bars are more appropriate, even though they have a higher error rate (Wilke, 2019, chapter 5, 6). The higher error rate means the different proportions within the bar chart do not share a common baseline (except for the bottom category), as they rely on the segment below them. This requires the viewer to mentally align elements (Cleveland and McGill, 1984). To mitigate this mental effort, color is used to help viewers distinguish between categories and compare sizes (Wilke, 2019, chapter 4). For that the principle of proportional ink, "the sizes of shaded areas in a visualization need to be proportional to the data values they represent" (Wilke, 2019, chapter 17) is followed. The result is one visualization for each dataset:

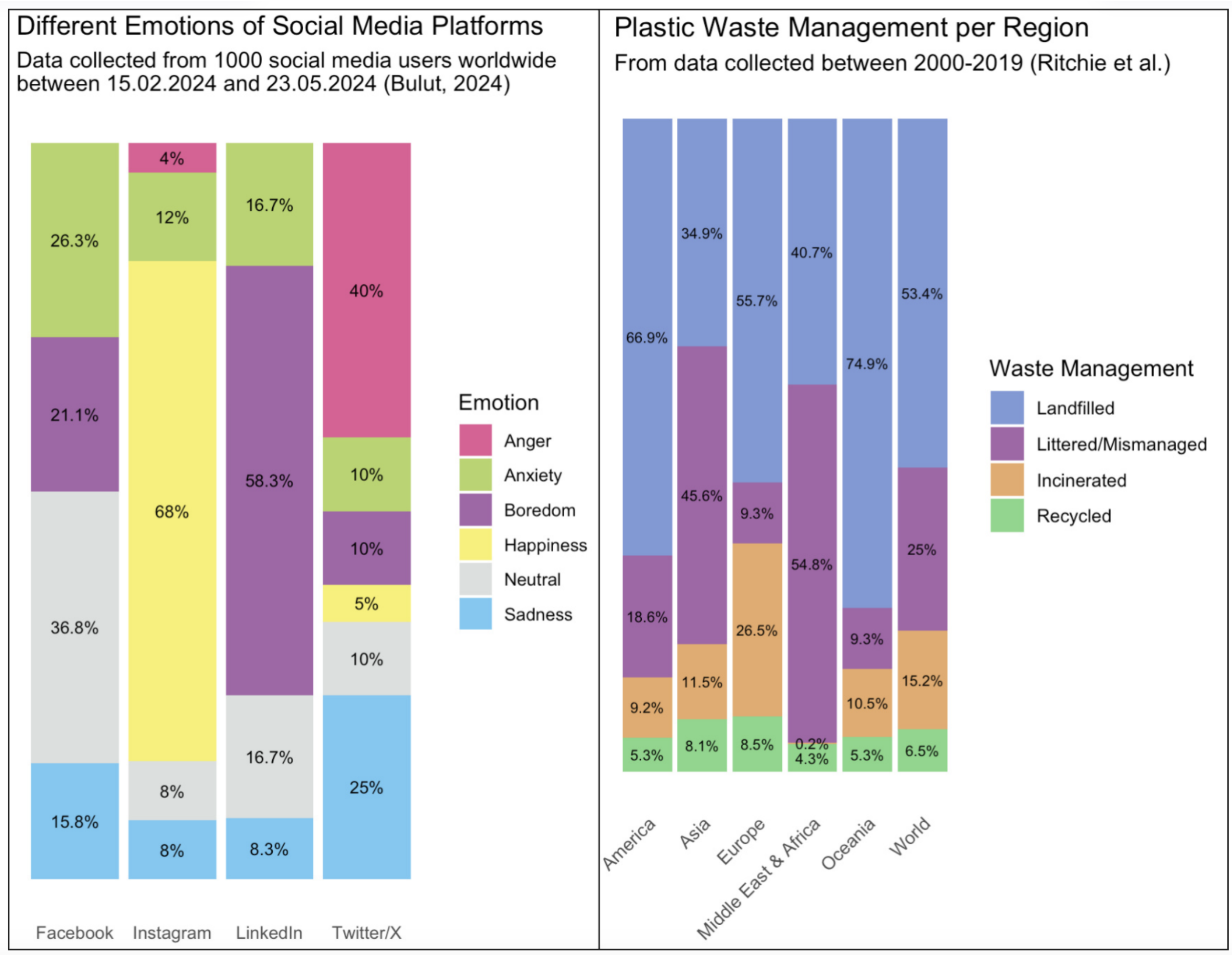


Figure 14: Visualization of the two datasets which is the control condition and serves as a baseline for the storytelling conditions

Instead of showing the y-axis of the graphs, the values are shown within the stacked bars (Figure 14). "This substantially increases the amount of information conveyed by the plot without adding much visual noise" (Wilke, 2019, chapter 6). The categories on the x-axis are self-explanatory, so they don't need labels. For example, "Social Media Platforms" is not labeled (Wilke, 2019, chapter 22). Adding the source of the data to the visualization makes the data more trustworthy because factors like credibility influence how people perceive the data (Petty and Briñol, 2011). The choice of color should represent the categories. For emotions, it represents the colors on the color wheel of emotions (health_blogger, 2023). The different types of waste management show recycling in green, representing the green color of the recycling symbol (WRAP, 2025). The other waste management types could not be assigned to a specific color. Incineration is orange to represent fire when burning trash. Red is not used because of color deficiency. Landfill could be represented by brown because of the earth where the trash is buried, but brown is too similar to orange. Therefore, blue was chosen because landfilling can also impact the sea (Silverman, 2023). Lastly, purple was chosen for littered and mismanaged waste because it can be distinguished from the other colors. Grid lines in the background are not used because they provide no additional visual help and would only add visual clutter (Wilke, 2019, chapter 23). The R code for creating the visualizations can be found in Appendix B.

Data Storytelling

As mentioned by Dykes (2019), p.32, effective data stories need to combine narrative, visuals, and data. The visual based on the data is designed in the previous section (Figure 14). Next, a narrative is developed in this section. A classical narrative structure consists of an introduction that builds to a climax and ends with a resolution (Dykes, 2019, p.87). To develop the data stories, the six elements of a data story (Figure 4) are used to guide the process (Dykes, 2019). Since a data foundation is already in place, the remaining elements are the main point, explanatory focus, linear sequence, dramatic elements, and visual anchors (Dykes, 2019). The main point identifies the target message that the story delivers. For social media platform data, the target message is that social media affects emotions in ways that are not always apparent and one should be mindful while using particular platforms. The main point of the plastic waste management data is that actions are needed to increase more sustainable ways of managing plastic waste. A data story must have an explanatory focus, meaning all insights pointed out during the story must be explained (Dykes, 2019, chapter 4). Furthermore, a linear sequence is required. This means the data is revealed in stages, with new insights building on each other to support the main point (Dykes, 2019, chapter 6). The data storytelling arc provides guidance on structuring the linear sequence (Figure 15).

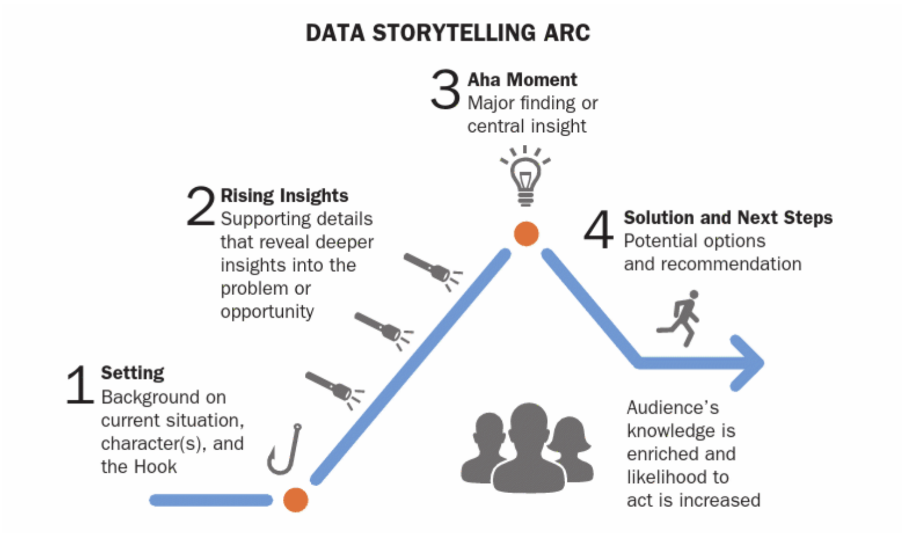


Figure 15: Data Storytelling Arc providing a structure for telling a data story in four steps (Dykes, 2019, chapter 6)

According to this arc, the two data stories are structured. The data storytelling arc of the social media platform dataset is displayed in the following (Figure 16). The arc for plastic waste management can be found in Appendix C.

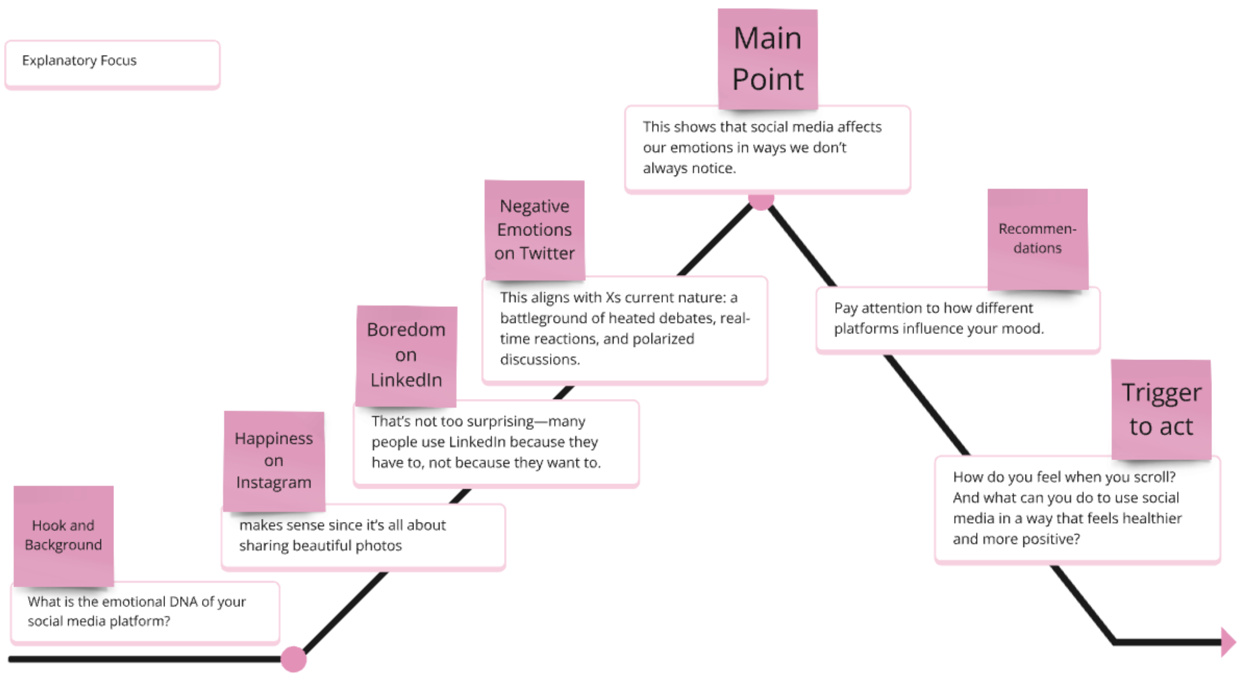


Figure 16: Data Storytelling Arc for the social media platform data (Dykes, 2019, chapter 6)

Figure 16 shows a hook and three rising actions, including an explanatory focus and main point, as well as a recommendation and trigger to take action. Dramatic elements, which are also part of the six elements of a data story, are used in the story. The data storytelling arc itself is a dramatic element (Dykes, 2019, chapter 6).

Furthermore, a visual anchor is needed, which is a visual representation of the data that helps the audience to understand the story better (Dykes, 2019, chapter 4). As discussed in the previous section, the data is visualized intuitively. This visualization can also be used for data storytelling. To support the linear sequence, only the parts of the data touched on in the story will be highlighted, one after another, while the rest fades out. This aligns with the signaling principle, which states that highlights direct attention to key content, and the segmenting principle, which states that information should be broken down into smaller segments (Mayer, 2005).

Finally, the data story should incorporate all five of Aristotle's modes of persuasion. Mentioning the source of the data and picking a trustworthy source ensures the appeal to credibility is met (Dykes, 2019, chapter 2). Explaining possible reasons for the data and basing the story on credible data meets the appeal to logic or reason. The appeal to emotion, or pathos, is ensured through the dramatic elements and structure of the story. The purpose is delivered through the main point, recommendations, and call to action. The appeal to timeliness is difficult to achieve because, even though the two chosen topics are of public interest, the exact time at which the data is presented to participants cannot be influenced, and good timing can vary from person to person (Dykes, 2019, chapter 2).

Communication Medium

Two different videos are created to put the story and the visuals together. The videos are created in the format of a TikTok video because "video is becoming a more important source of online news, especially with younger groups" (Newman et al., 2024) as two-thirds of the sample of the digital news report reported that they access short news videos weekly (Newman et al., 2024). One video includes the visualization that is highlighted throughout the story, while text is presented in the video. One of TikTok's standard fonts is used for the videos (TikTok, 2025). Depending on the text's length, it appears and disappears after six to eight seconds. There is a one-second break between every newly presented insight. The second video presents the story in audio format without text, except at the beginning and end. The voice presenting the insights is TikTok's standard Artificial Intelligence (AI)-generated voice (TikTok, 2025). Although the voice principle suggests that a human voice is more effective for narration, using a human voice would introduce hidden variables, such as accent and tone, that are difficult to control (Mayer, 2005). This study aims to compare Written and Verbal storytelling in video format. Other than that, the two videos are as similar as possible so that the focus can be on the different modes of communication when comparing the results. The control condition is just a visualization the same size as the video, with no storytelling narrative applied. The four data storytelling videos can be found in Appendix D.

3.6 Participants

The sample for the experiment consists of 166 people between the ages of 18 and 60 with a sufficient level of English proficiency to understand the information presented. A non-probability, self-selection sampling method is applied since the probability is unknown (Vehovar et al., 2016). Participants can decide for themselves whether or not to participate in the study. They are recruited through Prolific, a platform that connects researchers with participants, meaning that participants are paid for their participation, which increases extrinsic motivation. (Prolific, 2025; Ryan and Deci, 2000).

3.6.1 Legal concerns

Data collected from participants is stored and protected in accordance with the General Data Protection Regulation (Proton AG, 2025). For that, participants are asked for their consent to take part in the experiment. Additionally, at the beginning of the experiment, participants receive information about the data collected and how it is used to make the process as transparent as possible (Appendix E). Only general personal data is collected. Therefore, the collected data is anonymous.

3.6.2 Ethical concerns

The danish national center for ethics defines four scientific ethical principles: Autonomy, Beneficence, Non-maleficence, Justice (The Danish National Center for Ethics, 2024).

Autonomy means respecting participants' right to make their own decisions at all times.

As mentioned above, participants are asked to give their consent to participate in the study. Additionally, they have the possibility to leave the experiment at any point during the experiment. The researcher's contact details are provided at the beginning and end of the study so participants can reach out for more information or to see the results. (The Danish National Center for Ethics, 2024)

Beneficence describes whether the outcomes of the research justify the amount of work participants need to undertake. More specifically, the research must create value for the participants or for broader society. The motivation behind the research is explained in Section 1 and justifies the time participants invest.(The Danish National Center for Ethics, 2024)

Furthermore, the research should not cause any harm. No amount of value that the research brings justifies harm. In this experiment, participants are given information and asked questions. Since the experiment takes place online, participants can take part from any location. The information they are given represents facts. There is no psychological or physical harm done to the participants during the experiment. (The Danish National Center for Ethics, 2024)

Lastly, the principle of justice must be observed. In research, justice means fairness in who bears the risks and who benefits. Trials should include those who will benefit, avoid exploiting vulnerable groups, offer fair compensation, and ensure diverse participation. To ensure diverse participation, Profilic is used to recruit participants from different backgrounds. Anyone who wants to participate and is signed up on Profilic can do so and get compensated. (The Danish National Center for Ethics, 2024)

3.7 Materials

In this section, the tools that are used to develop the experiment are described.

For recruitment, Prolific is used (Prolific, 2025). It is a platform that connects researchers to a pool of participants and allows them to set up a specific screener to find the right participants for the research (Prolific, 2025). The experiment is set up in Gorilla (Gorilla, 2024). Gorilla allows setting up complex experiments that include surveys, reaction time tasks and cognitive tests providing a drag-and-drop interface (Gorilla, 2024). Additionally, it supports Prolific as a tool to recruit participants around the world (Gorilla, 2024).

For transforming and visualizing the two datasets R Studio and Miro are used (The R Foundation, 2025; Miro, 2025). R is a language for statistical computing and graphics that makes it possible to transform, analyze and visualize data (The R Foundation, 2025). The adjustments made to the visualization (highlighting different parts) for the video were made after the visualization was created in R in Miro, an extensive online whiteboard tool (Miro, 2025). Additionally, R Studio is also used for the final statistical and exploratory analysis of survey results. Last but not least, TikTok, a social media platform, offers the functionality to cut short videos with AI-generated voice and the possibility to incorporate extra text (TikTok, 2025).

3.8 Data Collection

In the experiment, quantitative data is collected through a survey. The survey is chosen because it allows for efficient outreach to a variety of participants (Clark et al., 2021, chapter 10). Due to the longitudinal study format, it is more convenient for participants to answer the survey at their leisure than to schedule three meetings with each participant. Additionally, presenting data insights to participants in their familiar surroundings when they open their laptop more closely resembles reality than presenting data insights during an online interview with an observer. The choice and design of the questions are explained in detail below. (Lazar et al., 2017, chapter 5)

3.8.1 Demographic Questions and Data Literacy

In order to provide context for the study and understand how demographics may influence the participants, the pre-questionnaire includes some demographic questions. These include questions about age, gender, highest level of education, and country of residence. Additionally, a question about political affiliation is included because strong political views can impact attitude and likelihood of change. To answer **RQ3**, the data literacy of the participants must be measured using a set of questions designed to assess data literacy. Since the goal is to assess the overall data literacy of individuals, and not data literacy in a professional context, the survey from QlikTech International AB (2025) is used as a survey to assess individuals' data literacy by Bonikowska et al. (2019). Only questions one to four are included as the other questions are too detailed and specific for people with a profession that has touchpoints with data (QlikTech International AB, 2025). All questions in the demographic and data literacy sections are taken from previous surveys, recommended by different platforms, or based on existing definitions. Another question about the frequency of watching short videos is asked using a Likert scale, since some participants will be shown short videos, and their prior experience with this format could impact the results. The following table shows the questions and their sources (Table 5).

Question	Type	Source
What is your age in years?	Text field	SurveyMonkey (2025)
How do you identify?	Multiple Choice (Likert-scale)	Matsumoto (2025)
What is your country of residence?	Text field	SurveyMonkey (2025)
What is the highest level of education you completed?	Multiple Choice (Likert-scale)	SurveyMonkey (2025)
Which of the following best describes your political affiliation?	Rating scale	Eldridge (2025)
Data Literacy Questions (1-4)	Multiple Choice (Likert-scale)	QlikTech International AB (2025)
How much time do you typically spend watching short videos (e.g., TikTok, Instagram Reels, YouTube Shorts) per day?	Multiple Choice (Likert-scale)	Self-developed

Table 5: Demographic and data literacy questions used in the survey and their sources

The guide from SurveyMonkey (2025) recommends placing demographic questions at the end of a survey to make them feel less invasive. However, since this experiment involves multiple surveys at different times and participants are compensated for their participation, the demographic questions are placed at the beginning. This ensures that the demographic data of each participant is collected and that their answers are put into context, even if they only participate in two of the three parts of the experiment.

3.8.2 Measuring Attitude

To measure attitudes toward plastic waste management and mental health on social media, an examination of existing literature is conducted. A study by Dowarah et al. (2022) measured attitudes toward plastic waste using a Likert-scale questionnaire. This method of measuring attitude is also recommended by Dijkstra and Goedhart (2012). The six questions fit the presented data and are used to measure attitude in this study (Dijkstra and Goedhart, 2012). No fitting existing questionnaire could be found for the data about social media. Therefore, six questions similar to those about plastic waste management were developed to measure attitudes toward social media and mental health. Both sets of questions can be found in Appendix F.

This study requires measuring changes in attitude to answer the research questions. Therefore, the same set of questions is asked before and after exposure to the insights, as well as one week and one month afterwards. The order of the questions is randomized to reduce the risk of participants memorizing the answers. Additionally, one attention question is added to ensure that participants read the questions carefully.

The pre-questionnaire asks one extra question about previous knowledge of plastic waste management and social media's impact on mental health because previous knowledge can impact attitude change (Petty and Briñol, 2011). This question is in a Likert-scale format.

3.8.3 Measuring Recall

Recall is measured by asking participants questions about the insights presented to them. They are given multiple-choice answer options to make the analysis of the results easier. Although the definition of recall is to retrieve information without specific cues, the answer options do not provide a cue because the participants' task is not to select words that were previously presented (recognition), but rather to select the correct answer to the question (Britannica, 2024).

Recall is measured over time (immediately, one week and one month after) to distinguish between short-term and long-term recall. Questions are randomized again to avoid recalling response patterns rather than the actual information. The recall questions can be found in Appendix F.

3.9 Data Analysis

For analyzing the data in R Studio, the process of Wickham and Grolemund (2016) is followed (Figure 17).

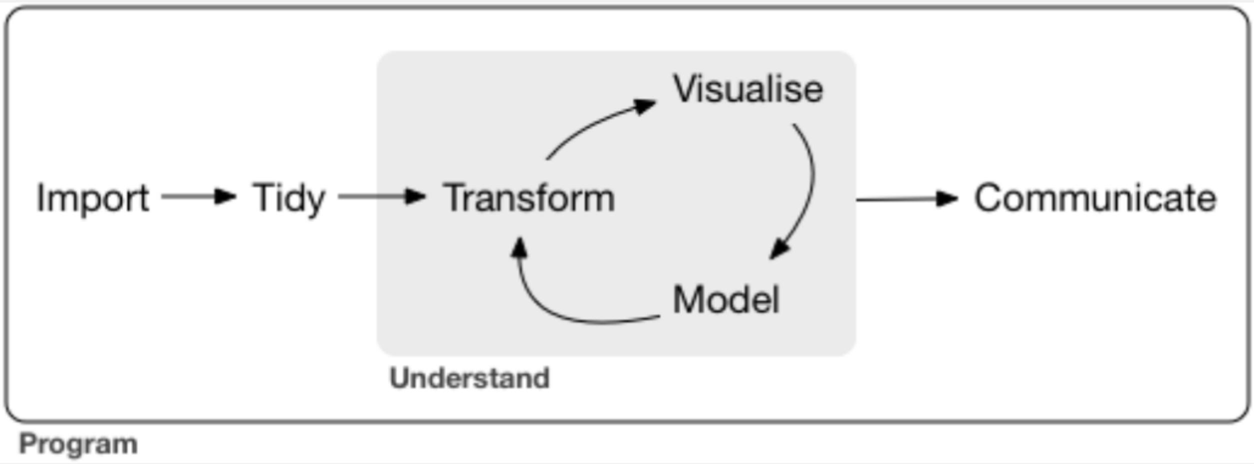


Figure 17: Data Analysis Process (Wickham and Grolemund, 2016)

First, the data is imported into R Studio. Then, it is cleaned so that each column is a variable and each row is an observation. In order to start working with the data, it needs to be transformed so that the variables of interest can be visualized and models can be used. This often includes calculations with existing variables, as well as converting Likert scale answers into numbers for easier analysis later on. Transformation is part of an iterative process, and if specific variables are missing during visualization or modeling, transformation can be performed on demand. Visualizations provide an overview of the data and reveal patterns that can be investigated further through modeling. They also help find the right model for the data. (Wickham and Grolemund, 2016)

3.10 Biases

Different biases caused by measurement instruments, experimental procedures, participants, environmental factors, and experimenter behavior can affect an experiment (Lazar et al., 2017). The latter is not applicable to this experiment because there is no experimenter present. To mitigate their impact as much as possible, these biases are considered and described below.

Bias caused by Measurement Instrument

This experiment uses questionnaires to measure outcomes. Questionnaire biases can arise from the design of the entire questionnaire, individual questions, or the administration process (Choi and Pak, 2005).

A potential bias that could appear in this experiment is social desirability bias when asking questions about attitudes toward an object. Participants who do not change their attitudes after being exposed to information may adjust their answers to align with social norms and expectations (Clark et al., 2021, chapter 10.3). To minimize this bias, the questionnaire is anonymous and conducted online, so participants do not feel observed or pressured (Clark et al., 2021, chapter 10.3). Additionally, all data in the questionnaires is self-reported, meaning the quality of the results depends on participants' honesty and motivation to complete the questionnaire. Therefore, the purpose of the study is explained at the beginning, and the experiment is kept as short as possible to prevent fatigue.

Bias caused by Experimental Procedures

To mitigate the risk of bias in the experimental procedure, clear instructions about what is expected are provided to the participant for every part of the experiment. Additionally, attitude and recall questions are randomized to mitigate learning effects. Finally, a pilot test with three participants is conducted to test the experimental procedures before the experiment is distributed to all participants. (Lazar et al., 2017, chapter 3)

Bias caused by Participants

Because the characteristics of participants can introduce biases into an experiment, participants must be carefully recruited. Since this is an online experiment, participants can decide when and where to participate, which reduces stress. Additionally, a between-group design was chosen to prevent participants from becoming overwhelmed with too many tasks. Additionally, an attention question is included to verify that participants are reading and answering the questions carefully. (Lazar et al., 2017, chapter 3)

Bias caused by Environmental Factors

Because the participants' environment cannot be controlled, biases may occur. To prepare participants for the experiment's tasks and allow them to find suitable surroundings in which to complete them, the type of task is disclosed at the beginning of the experiment. Ultimately, the experiment should feel as realistic as possible to the participants, and any resulting biases are taken into account. (Lazar et al., 2017, chapter 3)

4 Results and Analysis

This section covers the analysis of the collected data starting with a general analysis of the sample, followed by an analysis of the data relevant to answering the research questions and validating the research hypotheses. Finally, some additional data analysis is included to examine the expectations collected in Figure10. The full anonymized raw survey data can be found in Appendix G, the R code for the analysis in Appendix H.

4.1 Participant Sample

190 participants that consented to participating in the study and finished the pre-questionnaire are recruited through prolific in total. From them, participants that are not reading through the questions carefully, are not watching the full storytelling video, or are not looking at the visualization for more than 30 seconds are filtered out. This is done by filtering out all participants who answered the attention question wrongly and the ones who spent less time than the duration of the video on the video page or less than 30 seconds looking at the visualization. The breakdown of the number of participants per condition before and after the threshold and filter throughout the experiment can be seen in the following Table 6.

Condition	Initial	Threshold	Included	2nd Survey	3rd Survey
Social Media					
Verbal Storytelling	37	> 81 sec	36 (97.3%)	33 (89.2%)	30 (81.1%)
Written Storytelling	25	> 42 sec	25 (100%)	23 (92.0%)	22 (88.0%)
No Storytelling	35	> 30 sec	29 (82.9%)	27 (93.1%)	25 (86.2%)
Plastic Waste					
Verbal Storytelling	24	> 107 sec	21 (87.5%)	19 (90.5%)	17 (89.5%)
Written Storytelling	33	> 63 sec	28 (84.8%)	26 (92.9%)	24 (85.7%)
No Storytelling	36	> 30 sec	27 (75.0%)	25 (92.6%)	18 (72.0%)

Table 6: Participant counts, time thresholds, and retention rates across experimental conditions and survey rounds.

All participants who are above the time threshold in their experimental condition and answer the first attention questions correctly are included in the study, no matter if they drop out at some point in the longitudinal study. The study set-up in Gorilla is a random but equal distribution between the conditions. This did not work out too well as the social media, written storytelling condition has approx. 50% less participants than the plastic waste, no storytelling condition. The 166 included participants are from 28 different countries in

Europe (n = 78), Asia (n = 6), Africa (n = 78), Oceania (n = 3) and South America (n = 1). The majority of participants are either from European or African countries. The age distribution can be seen in the following (Figure 18).

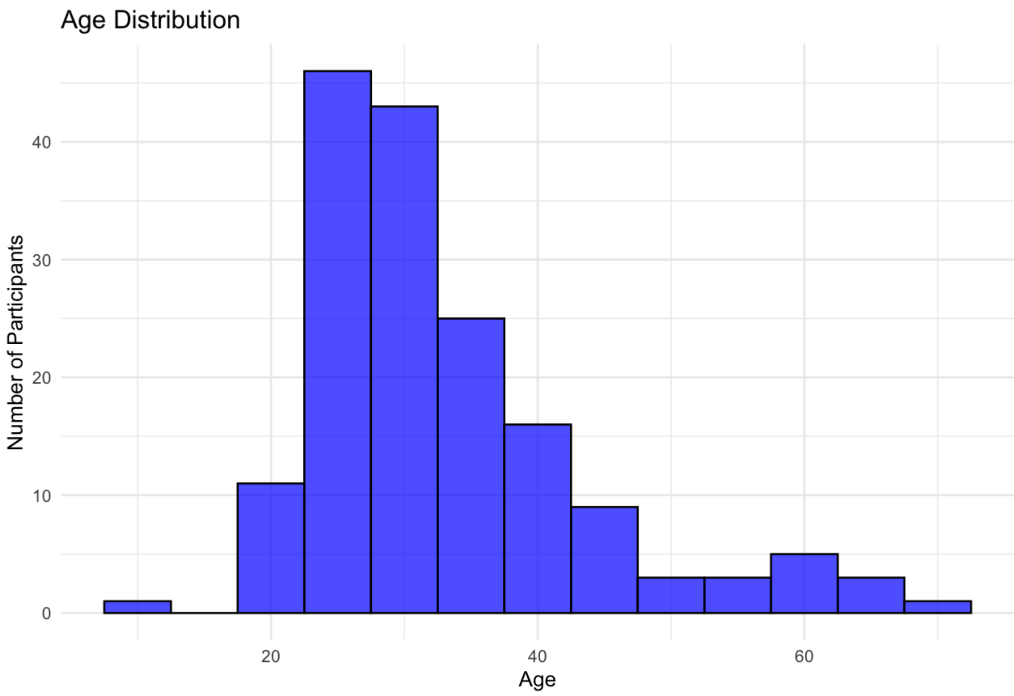


Figure 18: Age Distribution across sample. The majority of participants are between 20 and 40 years old.

Figure 18 shows that the majority of participants are between 20 and 40 years old. The median age is 30 years. 56.4% (n = 93) participants are women, 43.6% (n = 72) are men and 0.61% (n = 1) of participants prefer not to say their gender. There are no participants who identify as non-binary in this sample.

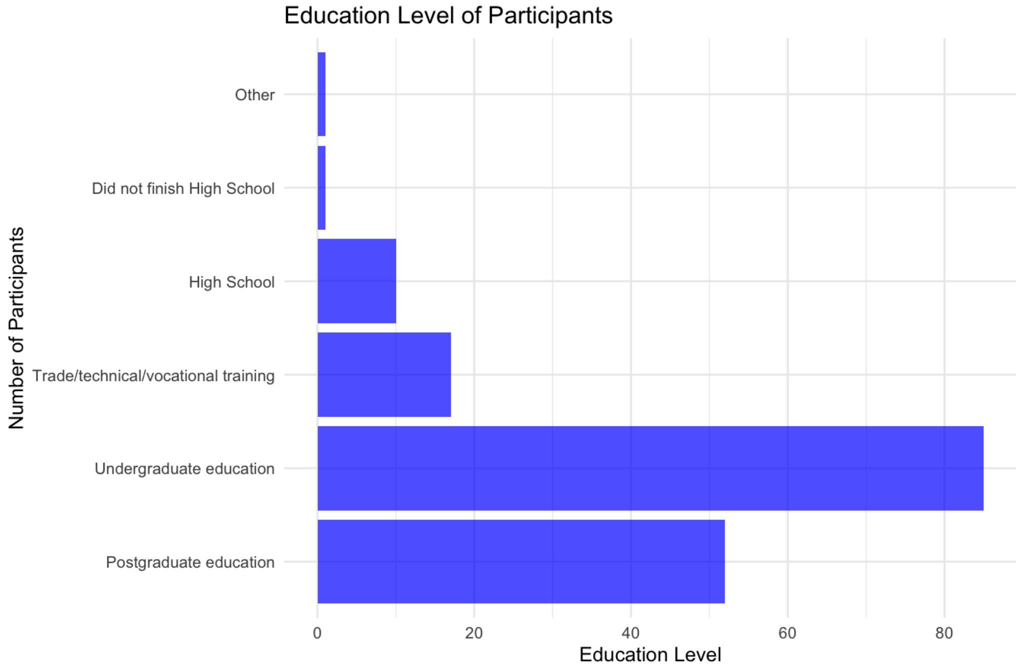


Figure 19: Number of participants per highest completed education. The majority of participants have completed a higher education.

Figure 19 shows the number of participants per highest completed education. The majority of participants have at least completed an undergraduate education. Following, this study is representative of rather educated people. 38% of participants have a political affiliation of centre (Figure 20). There are 6.6% more left-centre than right-centre participants. The extreme left and right positions differ by just 1.2% (Figure 20).

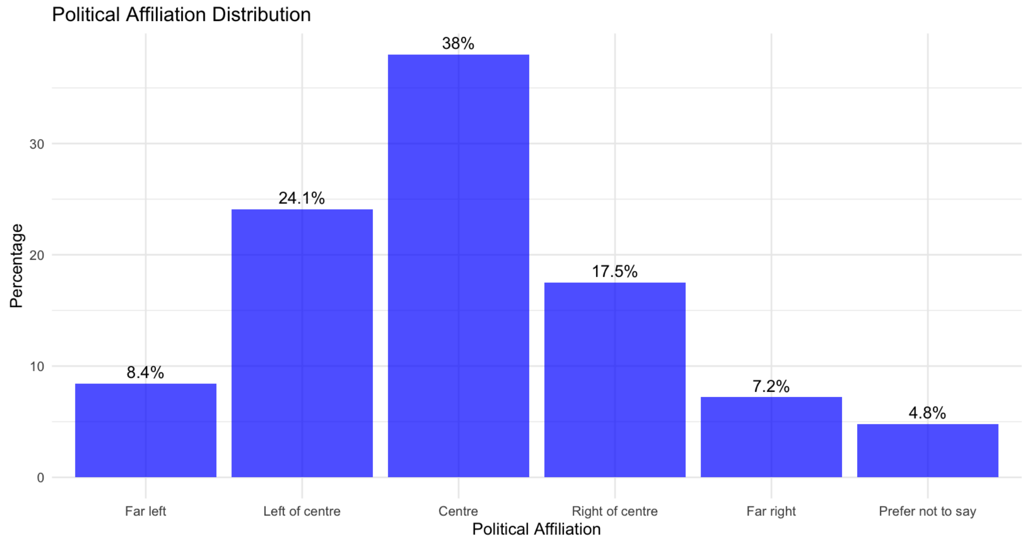


Figure 20: Politic Affiliation of participants. 32% of participants see themselves in the political centre.

The Interquartile Range (IQR) of previous knowledge about Social Media is lower than the one about Plastic Waste (Figure 21). To see if the previous topic knowledge differs statistically significantly from each other between topics, a Mann–Whitney U test is conducted (Field et al., 2012, chapter 15). The results show that the previous knowledge differs significantly between topics ($W = 4153.5$, $p = .010$). This suggests that participants that get exposed to the social media topic have more previous knowledge which could have an influence on attitudes and their ability to recall the information.

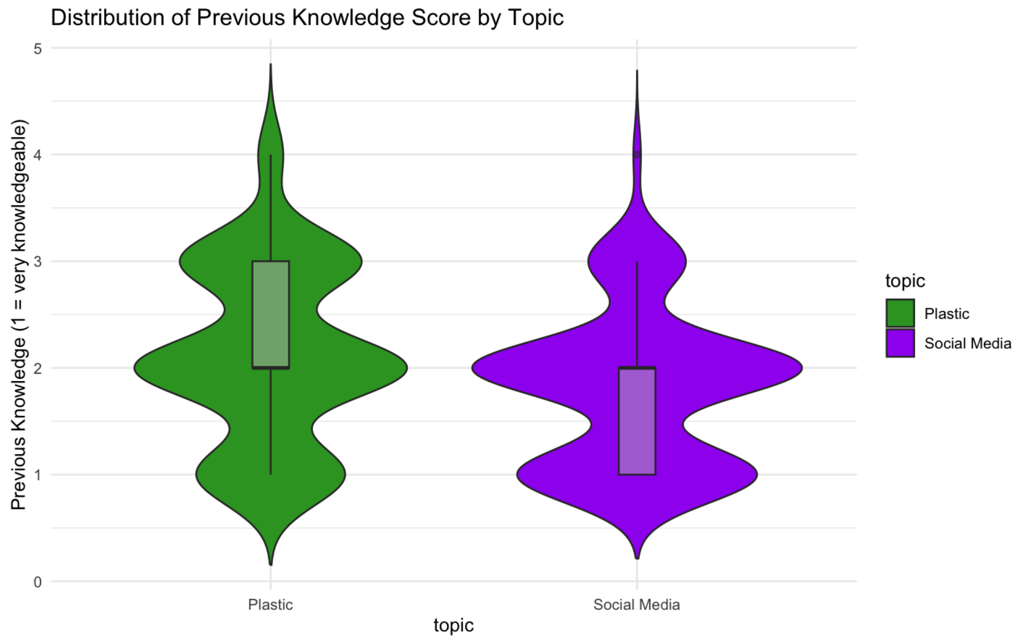


Figure 21: Previous topic knowledge by topic. The previous knowledge about the social media topic is higher than the plastic waste topic.

Only 6% of the participants report that they barely watch any short videos. Three-quarters of all participants watch short videos at least a few hours a day (Figure 22). The distinction between "I lose several hours per day" and "A few hours per day" might not be completely clear to the participants. Nevertheless, it can be seen that short videos are watched at least a few hours a day by the majority.

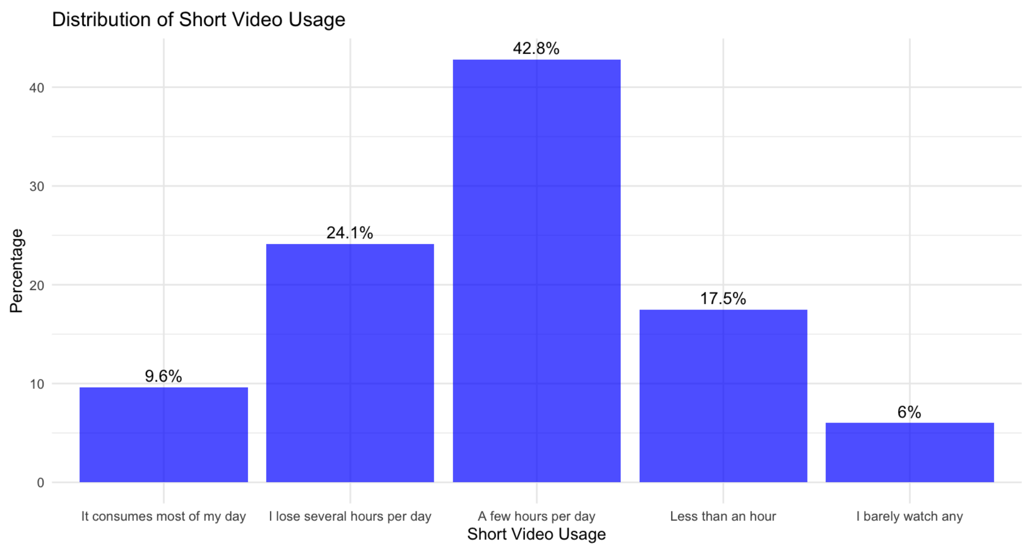


Figure 22: Short video usage of participants. Only 6% barely watch any short videos.

The best data literacy score a participant can achieve is four, representing a high data literacy. From seven on it is a moderate, ten a basic and 13 a low data literacy.

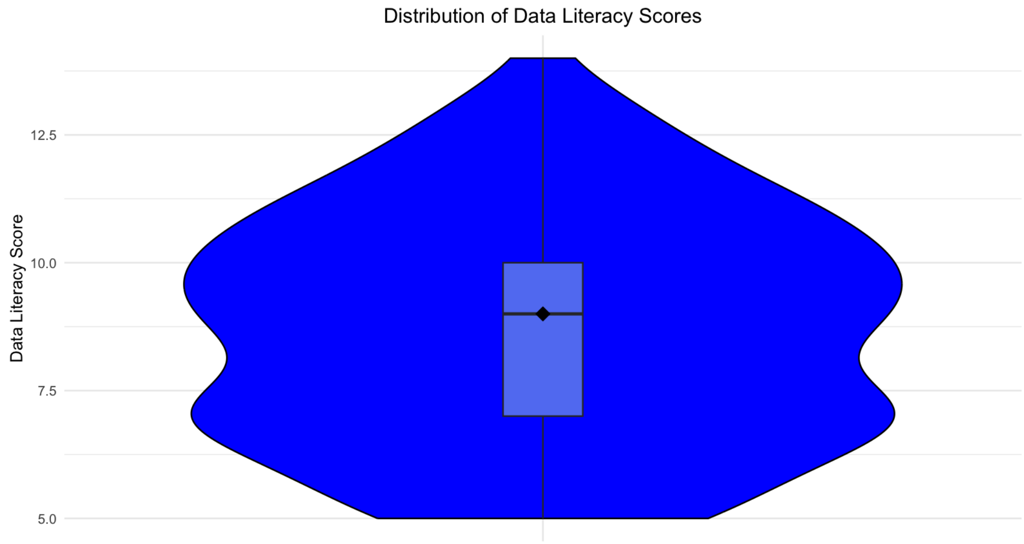


Figure 23: Distribution of data literacy score across participants. (4,5,6 $\rightarrow$ high; 7,8,9 $\rightarrow$ moderate; 10,11,12 $\rightarrow$ basic; 13,14,15 $\rightarrow$ low data literacy).

It can be seen from Figure 23 that the median of participants data literacy score is 9, meaning a moderate data literacy. Additionally, the middle 50% of participants have a moderate to basic data literacy score. Figure 24 shows the relation between data literacy and age.

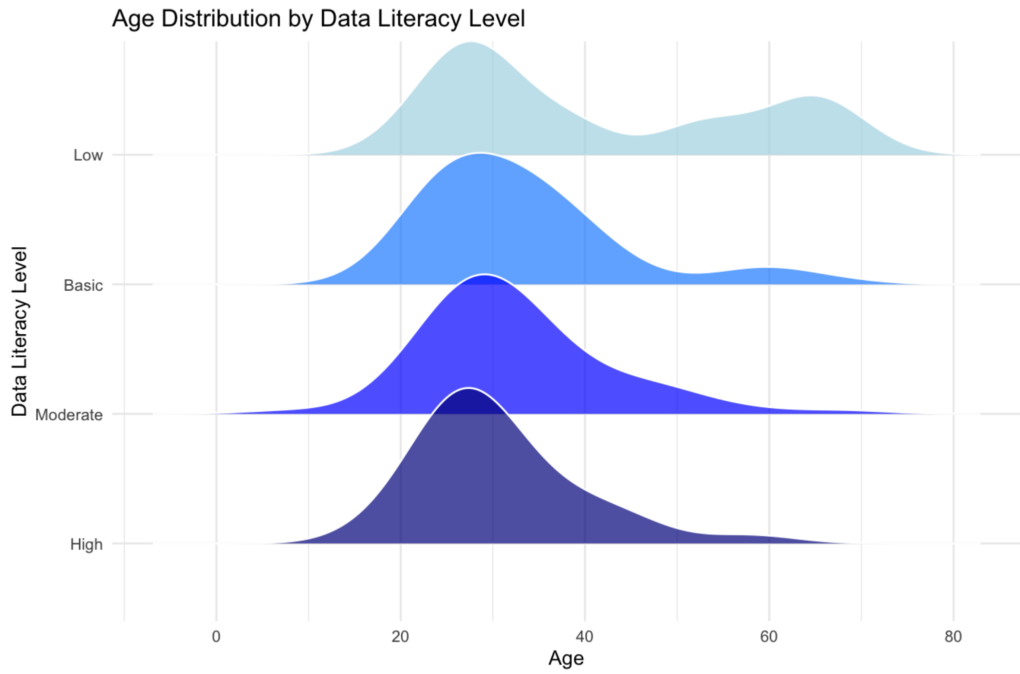


Figure 24: Age distribution by level of data literacy. Most participants with a higher age have a low data literacy.

The age of participants with low data literacy has two peaks, one at around 28 and one at 65. Additionally, low data literacy participants show more variability in age than participants of other data literacy levels. Participants with high data literacy on the other hand show the highest peak at 28 without having another peak, meaning that younger participants have a higher data literacy. To see if the correlation between age and data literacy is statistically significant Kendall's correlation with the exact data literacy scores and age is conducted, since the data is not normally distributed after a Shapiro-Wilk test and QQ-plot and shows many ties (data literacy: $W = 0.958, p < .010$; age: $W = .889, p < .010$) (Field et al., 2012, chapter 5,6) (Appendix I). The test fails to reject the null hypothesis that there is no monotonic relationship between age and data literacy ($\tau = .066, p = .239$).

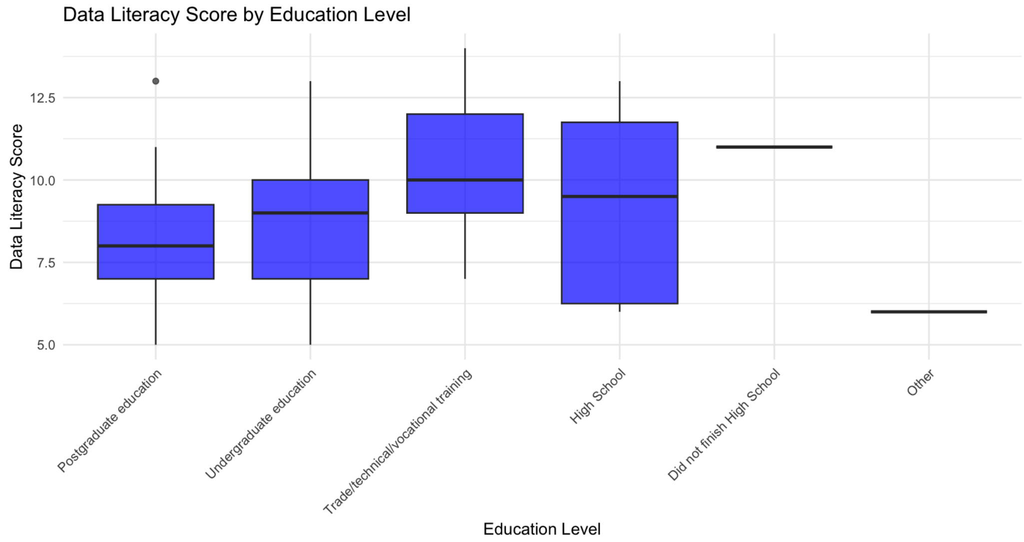


Figure 25: Data literacy score of participants per highest completed education. Postgraduate education has the lowest median data literacy score (highest data literacy)

Another variable that could influence data literacy is education. Figure 25 shows that people with a postgraduate education have the lowest data literacy scores, meaning the highest data literacy. Followed by undergraduate education. Participants whose highest completed education is high school have a higher data literacy than those who completed trade/technical/vocational training. Nevertheless, the IQR for participants who completed high school as their highest education is also wider suggesting more variability in data literacy scores across the participants. The IQR for "Did not finish High School" and "Other" are not visible due to a very small sample size (compare Figure 19). Since data literacy is not normally distributed (Appendix I), a correlation test (Kendall) is conducted to examine if the trend of higher educated participants having a higher data literacy is statistically significant or not (Field et al., 2012, chapter 5,6). The result shows a statistically significant (p = .002) positive correlation (τ = .195), meaning participants with a higher completed education (lower education score) tend to have a higher data literacy (lower data literacy score). (Figure 25)

4.2 Attitude Change over Time

To analyze attitude change over time, the answers to the six attitude questions per time point are summed up into a total attitude score. "Strongly agree" is coded with 1 and "Strongly disagree" with 5 and "Don't know" is coded as NA. Thus, in the end, there are four different attitude scores, the lower they are (closer to 0), the more they align with the information the participants are exposed to in the experiment.

Before analyzing attitude change over time, the initial attitudes of participants for each topic are visualized. This shows how much the initial attitudes vary from each other and if the topics can be analyzed together or should be treated separately from each other (Figure 26).

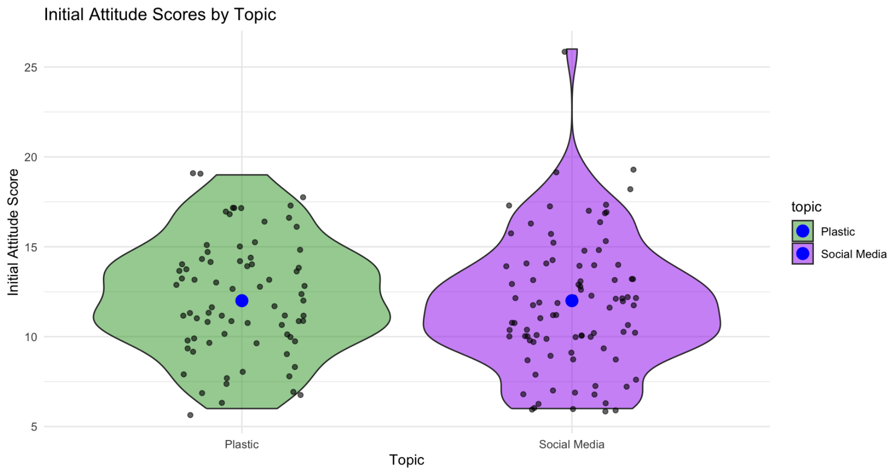


Figure 26: Initial attitude score by topic. The distribution looks similar across topics. A lower attitude score is better (aligns with the information they will be presented with).

The median of the initial attitude scores for both topics is 12. In addition to that, except for some outliers, the distribution of attitude scores per topic looks similar. A comparison analysis is run to see if the two groups differ statistically significantly from each other. The Social Media data turns out to be not normally distributed after visualizing it with a QQ-plot and doing a Shapiro-Wilk test (Social Media: $W = .951$ $p = .002$; Plastic: $W = .975$ $p = .141$) (Appendix J) (Field et al., 2012, chapter 5). Due to non-normality and the fact that two groups are compared with each other, a Mann–Whitney test is chosen to do the comparison (Field et al., 2012, chapter 15). The results show no statistically significant difference in initial attitude scores between the two topics ($W = 3777.5$, $p = .245$). Therefore, it is assumed that the two topic groups can be meaningfully compared and analyzed together in subsequent analyses.

The mean attitude scores over time per condition are visualized in the following (Figure 27).

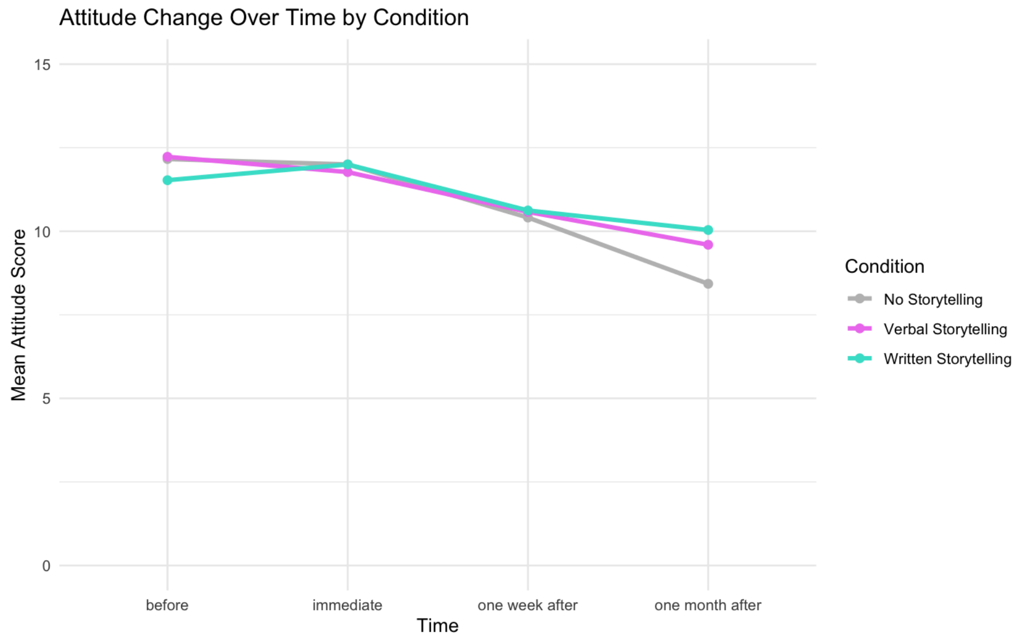


Figure 27: Mean attitude scores over time per condition. The scores decline over time.

It can be seen that the score declines over time for all conditions, meaning that the attitudes change in favor of the information they were presented with. For the no storytelling condition, the strongest decline happens between one week and one month. For written storytelling, the mean attitude score goes up between before and immediately after the exposure of information before it starts declining. Mean attitude scores of the Verbal storytelling condition show a continuous decline over time.

To analyze if changes in attitudes within the different conditions over time are statistically significant, a linear mixed model regression is applied. This regression is chosen because the data has repeated measures (four time points). At this point, also a repeated measure Analysis of Variance (ANOVA) could be used for this. Nevertheless, the test should also account for participants having individual baseline attitude levels which violates the independence assumption of ANOVA. A likelihood ratio test comparing a baseline model without participants as a random effect to a random intercept model including the participants indicated a significant improvement in fit ($\chi^2 = 199.04$, $p < .001$). Additionally, an initial Intraclass Correlation (ICC) analysis showed that approximately 46% of the variance in attitude scores was attributable to differences between participants (ICC $\approx$ 0.46), supporting the use of a multilevel model to account for the nested structure of repeated measurements within individuals. To start with, a simple model is used. More interactions and fixed effects are added gradually to compare the fit of the model. First, time is treated as fixed effects to understand if the attitude changes over time. As mentioned before, the participant is treated as a random effect (model 1). The second model adds the condition (Verbal, Written, No storytelling) as an interaction (model 2). Comparing both models (1,2) by calculating the likelihood ratio shows that there is an improvement when adding the condition (χ^2 = 8.800). Nevertheless, this improvement is not statistically significant (p = .360). Since extreme attitudes from the beginning are expected to be less likely to change, the initial attitude is added to the not significantly improved model including the conditions (model 3) (Nickerson, 1998). Comparing model 2 and model 3 shows that model 3 significantly improves model fit ($\chi^2 = 152.74$, p < .001). The higher the initial attitude score (the more it does not align with the data the participants are presented with), the higher the attitude change. This makes sense, as people with an attitude that aligns fully with the information can not change their attitude to the same extent, as their attitudes already align with the information. Another variable that could have an influence is previous knowledge about the topic presented (Petty and Briñol, 2011). Adding this (model 4) and comparing it to model 3 results in no significant improvement in fit, ($\chi^2 = 2.37$, p = .499). Therefore, previous knowledge is not retained in the final model. The performance of the final model 3 is evaluated using marginal and conditional R^2 values (Nakagawa and Schielzeth, 2013). The marginal R^2 indicates that the fixed effects (time, condition, and initial attitude) explain 42.1% of the variance in attitude scores. The conditional R^2 indicates that the full model, including random intercepts for participants, explains 54.4% of the variance. These results suggest that both fixed and participant-level random effects contribute to explaining attitude change. (Field et al., 2012, chapter 19).

In summary, attitude change over time is significantly influenced by the initial attitude score and the time point, while the storytelling condition does not have a significant effect. Participants with initial attitudes that are misaligned with the presented information show greater changes in attitude.

When viewing the regression model, only the mean differences from the baseline (before) can be seen. To be able to see a pairwise comparison of the different time points, we, therefore, calculated estimated marginal means, which gets the predicted mean attitude score for each time point, separately for each condition (Rdocumentation, 2025). Last but not least, the time points within each condition are compared and shown whether the differences are significant. For that, Bonferroni correction for multiple comparisons is used to protect from false positives (Field et al., 2012, chapter 10.5). The results are shown in the following table. Significant changes (after Bonferroni correction: $p < .0028$) are highlighted (Table 7).

Time Comparison	**No Storytelling**		**Verbal Storytelling**		**Written Storytelling**	
	Estimate	**p-value**	**Estimate**	**p-value**	**Estimate**	**p-value**
Before → Immediate	-.161	1.000	-.456	1.000	.472	1.000
Before → One week	-1.750	.019	-1.649	.030	-.906	.818
Immediate → One week	-1.589	.044	-1.193	.252	-1.377	.142
Before → One month	**-3.732**	**<.001**	**-2.632**	**< .001**	-1.491	.086
Immediate → One month	**-3.571**	**< .001**	**-2.175**	**.001**	-1.962	.008
One week → One month	1.982	.005	.982	.563	.585	1.000

Table 7: Estimated attitude change across time points by condition. Statistically significant changes are highlighted.

The table shows significant changes in attitude for the before - one month and immediate - one-month comparison for both the No storytelling and Verbal storytelling conditions. In the Written storytelling condition, no change in attitude is statistically significant. In the No storytelling and Verbal storytelling condition participants' attitudes significantly decrease one month after exposure compared to before and immediately after, suggesting a delayed but strong negative shift. The shift in attitude is slightly bigger in the No storytelling condition than in the Verbal storytelling condition. (Table 7)

The difference of the statistically significant attitude changes between No storytelling and Verbal storytelling is tested in the following. First, the attitude change from before to one month and immediate to one month is tested for normality for both, Verbal and No storytelling, with a QQ-plot and Shapiro-Wilk test (Field et al., 2012, chapter 5.6) (Appendix K). Attitude change from before to one month is not normally distributed (No storytelling: $W = .892$, $p < 0.001$; Verbal storytelling: $W = .879$, $p < 0.01$). Additionally, attitude change from immediate to one month is neither normally distributed (No storytelling: $W = .884$, $p < .001$; Verbal storytelling: $W = .859$, $p < .001$). Therefore, a Mann–Whitney U test is used to do the comparison analysis (Field et al., 2012, chapter 15.4). The test reveals no significant difference in attitude change from before to one month between the Verbal and No storytelling conditions ($W = 1722.5$, $p = .468$). Also for attitude change from immediate to one month between the two storytelling conditions, no significant differences are revealed ($W = 1801.5$, $p = .236$).

To sum up, attitudes change most clearly in the No storytelling and Verbal storytelling conditions. In both, participants' attitudes became significantly more aligned with the information one month after they saw it. This suggests that the biggest shift in attitude happens over time. In the Written storytelling condition, attitude scores also go down, but the changes are not statistically significant. A statistical model that takes into account each participant's individual starting point shows that the attitude change can be mainly explained by time and people's initial attitudes. Finally, while Verbal and No storytelling both lead to significant attitude change, they do not differ significantly from each other, meaning that Verbal storytelling does not lead to a stronger shift than No storytelling. Following, **H1** is partially true because attitude changes not for all storytelling conditions, just for the Verbal and No storytelling condition. However, **H1.1**, saying that Verbal storytelling is more likely to change attitudes than Written storytelling is true in this experiment.

4.3 Recall of Information

The answers to the recall questions are coded into 1 for a right answer, -1 for a wrong answer and 0 for "I don't know" answers. For every point in time, a total recall score is calculated. Following, the higher the score, the better the information is recalled (max. score 5).

To start with, a look at the immediate recall scores per topic is taken to see if recall scores differ across topics and need to be separated in the following analysis (Figure 28).

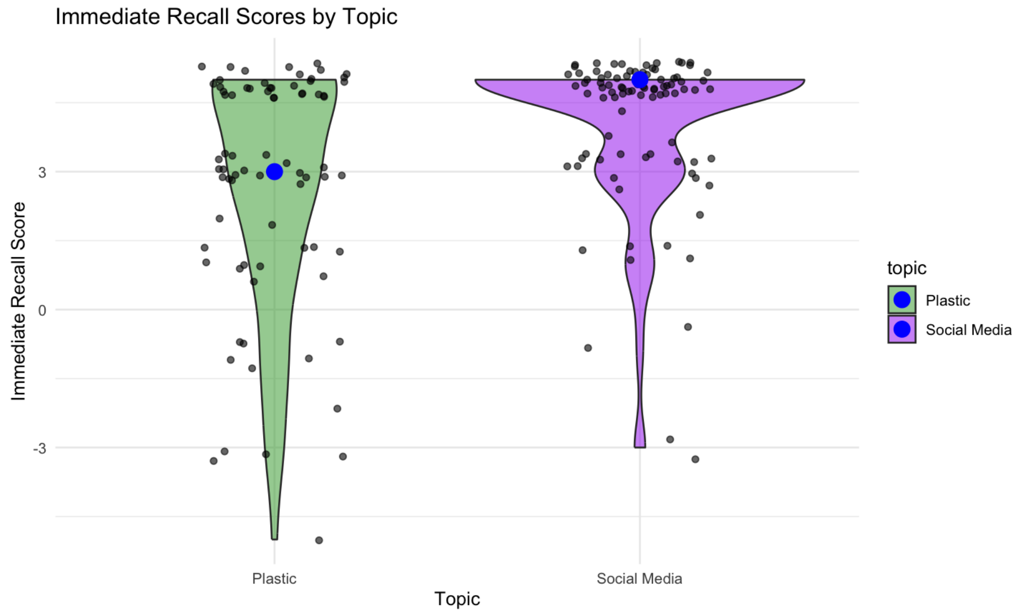


Figure 28: Immediate recall scores by topic. The median recall score is marked with a blue dot. For social media, the median recall score is higher than for information about plastic waste management.

The distribution shows that recall scores for the social media topic are higher than scores for information about plastic waste management. The medians also differ from each other. To see if the scores across topics differ statistically significantly from each other the scores are first checked for normality. The results of the QQ-plots as well as the Shapiro-Wilk test show no normal distribution (Plastic: $W = .818$, $p < .001$; Social Media: $W = .612$), $p < .001$) (Appendix L). Following, a Mann–Whitney-U test is conducted (Field et al., 2012, chapter 15.4). The results show that there is a significant difference in recall scores between the topics ($W = 2369$, $p < .001$). Therefore, the topics will be analyzed separately from each other in the following. Figure 29 shows the immediate recall scores per topic for every storytelling condition.

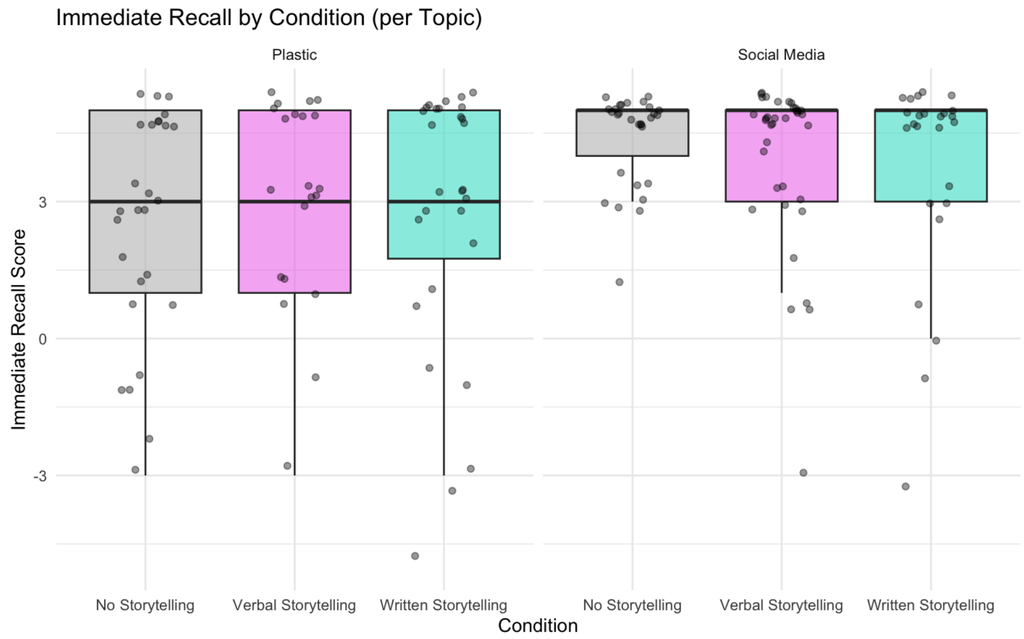


Figure 29: Immediate recall scores by condition per topic. Recall scores of No storytelling for the social media topic are the highest.

For Plastic, the immediate median recall scores of all three conditions look similar, the IQR for Written storytelling is the highest. Also for Social Media, the medians look similar, here the IQR of the No storytelling condition is the highest (Fig 29). A Kruskal-Wallis test is performed to investigate if the three groups differ from each other significantly within each topic. This test is chosen because the data is not normally distributed (ANOVA not possible) and the comparison is between three groups (Field et al., 2012, chapter 15.6). The test shows no statistical significance in either of the topics (Social Media: χ^2 = .824, p = .662; Plastic: χ^2 = .493, p = .782). To see if there are significant differences in recall scores between conditions in long-term recall, the same steps as before are gone through with recall after one week and one month. The recall after one week split by condition is visualized in the following (Figure 30).

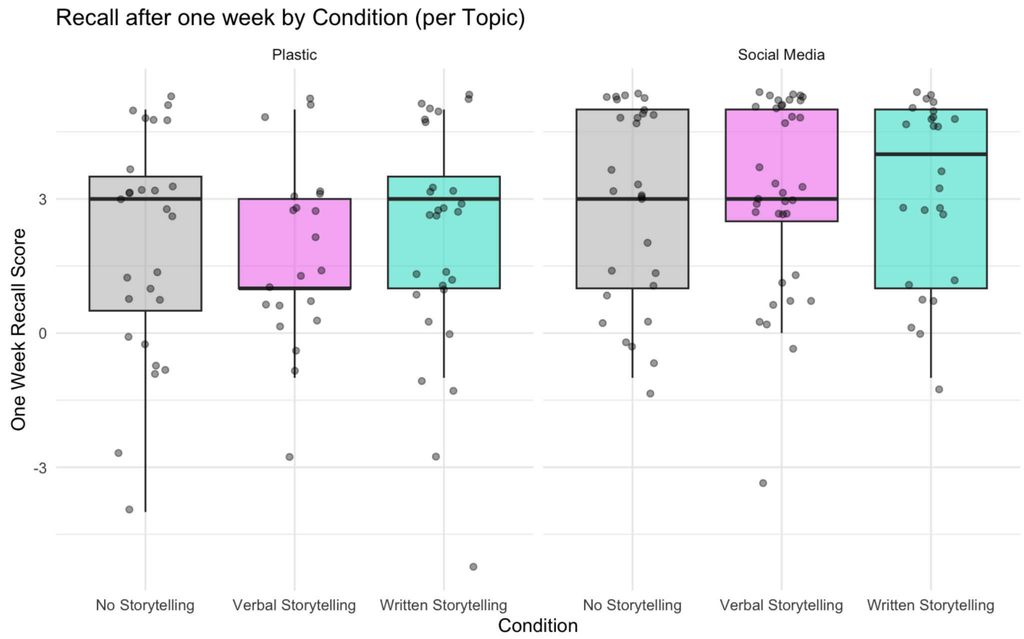


Figure 30: Recall scores after one week by condition per topic. Verbal storytelling has the lowest median for topic plastic. Written storytelling has the highest median in the social media topic.

For the topic plastic, the median recall score is lowest for Verbal storytelling and the data also seems to have less variation. For Social Media, Written storytelling shows the highest median recall score and tighter upper distribution. Verbal and No storytelling are similar in medians but the No storytelling condition has more lower outliers. A Kruskal-Wallis test, again, doesn't show any statistically significant differences in one-week recall scores between the storytelling conditions (Social Media: $\chi^2 = .569$, $p = .752$; Plastic: $\chi^2 = .549$, p = .760).

Figure 31 shows recall scores after one month. The median recall scores across all storytelling conditions for the plastic topic are lower than for the social media topic. For the topic plastic the median recall score of the No storytelling condition is the lowest. Verbal and Written storytelling show a similar median. For the social media topic, Written storytelling shows the highest median. Verbal and No storytelling have a similar median but in the No storytelling condition recall scores are more spread.

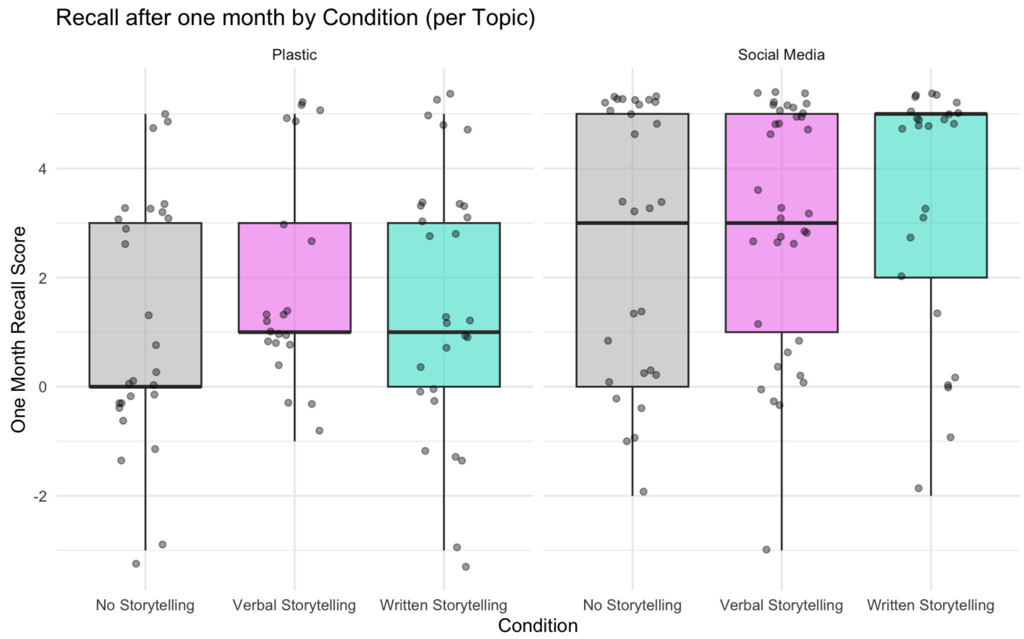


Figure 31: Recall scores after one month by condition per topic. Verbal storytelling has the lowest median for topic plastic. Written storytelling has the highest median in the social media topic.

The results of the Kruskal-Wallis test do not show a statistically significant difference in one-month recall scores between the storytelling conditions within each topic (Social Media: $\chi^2 = 1.424$, $p = .491$, Recall: $\chi^2 = 1.839$, $p = .399$).

In summary, participants are able to recall information better immediately after the exposure, when it is about social media compared to plastic waste management. This difference is statistically significant. Because of that, the two topics are analyzed separately. When comparing recall scores across the three storytelling conditions, no significant differences are found, neither right after the exposure, nor after one week and one month. Overall, the type of storytelling does not lead to significantly different recall scores within either topic over time. Concluding, **H2** and with that **H2.1** are both rejected in this experiment.

4.4 Influence of Data Literacy on Attitude

To understand if data literacy impacts attitude change after being exposed to data, the relationship of the two variables is investigated closely. To start with, the previously defined data literacy levels are visualized with the attitude score over time to make trends over time visible (Figure 32).

Figure 32: Attitude change over time split by data literacy levels. The change in attitude from before to immediate for participants with low and basic data literacy is the only one showing a different trend.

The visualization shows that, except for low and medium data literacy participants and their attitude change from before to immediate, all data literacy levels show the same downward trend in attitude change. This means that the attitude changes towards the information the participants were presented with. Nevertheless, some changes are steeper than others (low data literacy between immediate and one week) (Figure 32). It should be mentioned that the sample size for participants with low data literacy (8 participants) is low and might not be representative.

To see if, regardless of condition or baseline attitude, people with higher data literacy are more or less likely to change their attitude short- and long-term, a simple linear regression is run for three time comparisons (before - immediate, before - one week, before - one month). The results show a statistically significant negative relationship between attitude change from before to immediate ($b = -.143$, $p = .040$), suggesting that people with higher data literacy show less immediate attitude change. No statistically significant relationships for before - one week ($b = -.023$, $p = .870$) and before - one month ($b = .073$, $p = .705$) can be found.

To see if data literacy has different effects on attitude dependent on what condition the participant is exposed to, the previously analyzed data is further split into the different conditions. The visualization is displayed in Figure 33.

Figure 33: Attitude change over time split by data literacy levels and conditions. Attitude scores decline over time.

The visualization shows that attitude scores decline over time. In all conditions low and basic data literacy participants attitude score is rising from before to immediate, before they start declining, suggesting a delayed effect of attitude change here. The attitude score rises between one week and one month for high data literacy participants in the Verbal storytelling condition and basic data literacy participants in the Written storytelling condition. The highest drop in attitude score show low data literacy participants in the Written storytelling condition from one week to one month. Also, the mean attitude score for low data literacy participants after one month in the No storytelling condition is the lowest. This could be again, due to the small sample size of participants with low data literacy.

An interaction term between time and data literacy is added to the attitude change linear mixed model (model 3) to assess whether individuals with varying data literacy levels experienced different trajectories of attitude change (Field et al., 2012, chapter 15). To get more accurate results in the analysis, the accurate data literacy score instead of the data literacy levels are used in the model. Adding an interaction between time and data literacy to the model predicting attitude scores does not significantly improve model fit ($\chi^2 = 2.2623$, p = .688). Data literacy is also added as a main effect instead to see whether, on average, higher data literacy is associated with higher or lower attitude scores. This also does not significantly improve model fit ($\chi^2 = 0.7239$, p = .395), indicating that participants' level of data literacy does not significantly influence their overall attitude scores across time and conditions.

Concluding, participants' attitudes tend to change in the direction of the information they were shown, regardless of their data literacy level. Although people with lower data literacy show a slightly different pattern right after exposure, and some groups show steeper changes than others, these differences are small and may be influenced by the low number of participants with low data literacy. A linear regression shows that people with higher data literacy change their attitudes less immediately after exposure compared to people with lower data literacy, this difference disappears over time. This suggests that people with higher data literacy need more time to change their attitudes than people with lower data literacy. When data literacy is added to a more complex model, both as an interaction with time and as a general predictor, it does not improve the model. This means that, overall, data literacy does not effect how much people's attitudes change in this study. Thus, **H4**, suggesting that high data literacy leads to attitude change in the long term can not be confirmed.

4.5 Influence of Data Literacy on Recall

The following visualization shows an overview of recall at every point in time by level of data literacy (Figure 34). It shows that participants with basic data literacy levels show the lowest recall scores for every time point. For immediate recall, participants with moderate data literacy have the highest mean recall score, followed by low data literacy but low data literacy shows a high error rate due to the small sample size. For recall after one week, participants with high data literacy show the highest mean recall together with moderate data literacy participants. All recall scores decrease over time. (Figure 34)

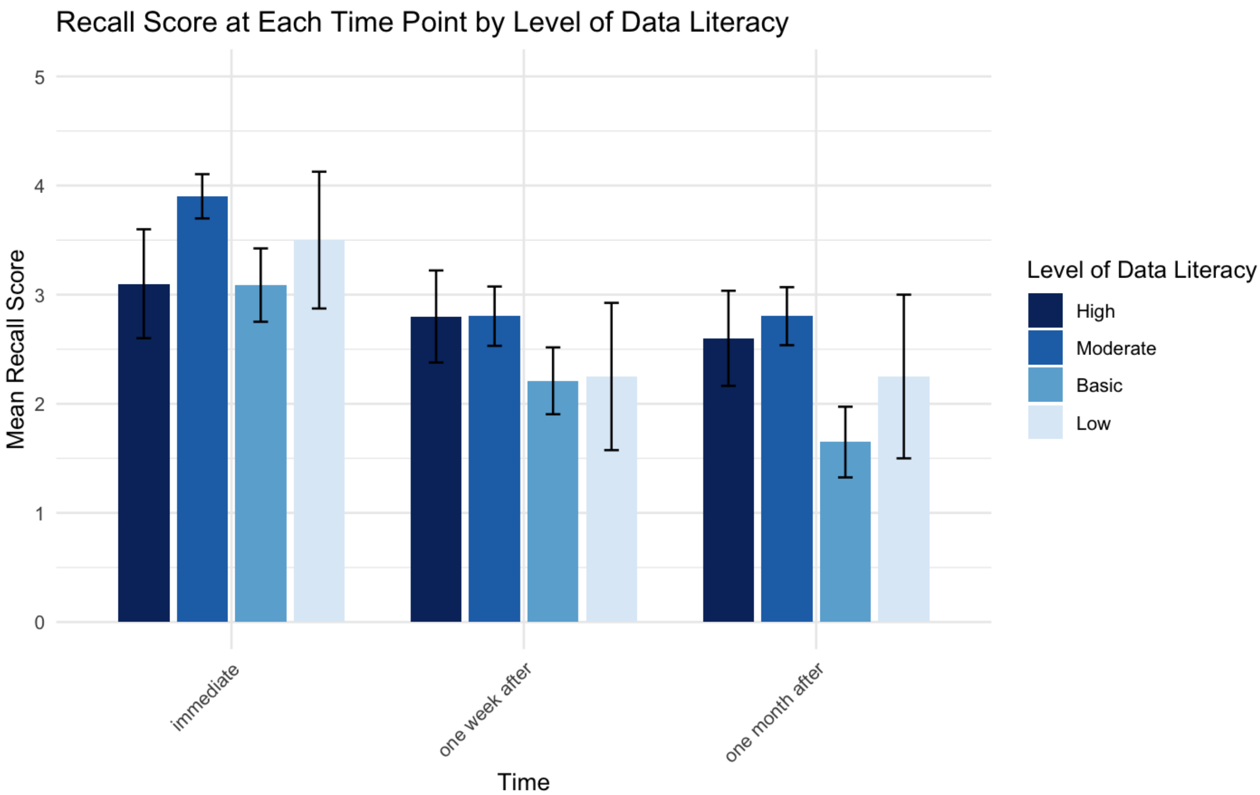


Figure 34: Mean recall score by level of data literacy for each time point. Participants with moderate data literacy have the highest recall scores.

Since a statistically significant difference between immediate recall and the two topics exists, a linear regression model is run for every point in time separately to see if the effect of data literacy on recall depends on the topic (Section 4.3). None of the models returns a statistically significant difference in the relationship between data literacy and recall by topic (Table 8).

Predictor	**Immediate Recall**	**One Week Recall**	**One Month Recall**
Data Literacy × Topic	.209 (p = .164)	.027 (p = .861)	.049 (p = .753)

Table 8: Estimates and p-values from linear models predicting recall score at each time point.

Therefore, for the following statistical tests, no split into topics is necessary. A Kendall's tau correlation is conducted since the recall scores for every time point are not normally distributed after a Shapiro-Wilk test (immediate recall: $W = .720$, $p < .001$; one-week recall: $W = .876$, $p < .001$; one-month recall: $W = .871$, $p < .001$) (DATAtab Team, 2025). All three taus for each time point are negative suggesting that higher data literacy (lower data literacy score) is associated with higher recall scores (immediate: $\tau = -.03$, $p = .633$; one week: $\tau = -.073$, p = .231; one month: $\tau = -.130$, p = 0.031). Nevertheless, just the recall scores after one month show a statistically significant correlation with data literacy. These results suggest that participants with higher data literacy show better recall scores in the long term.

To see if there is a relationship between recall and data literacy within different conditions in general, a total recall score (the sum of all recall scores at each time point) is calculated and the relationship visualized in Figure 35.

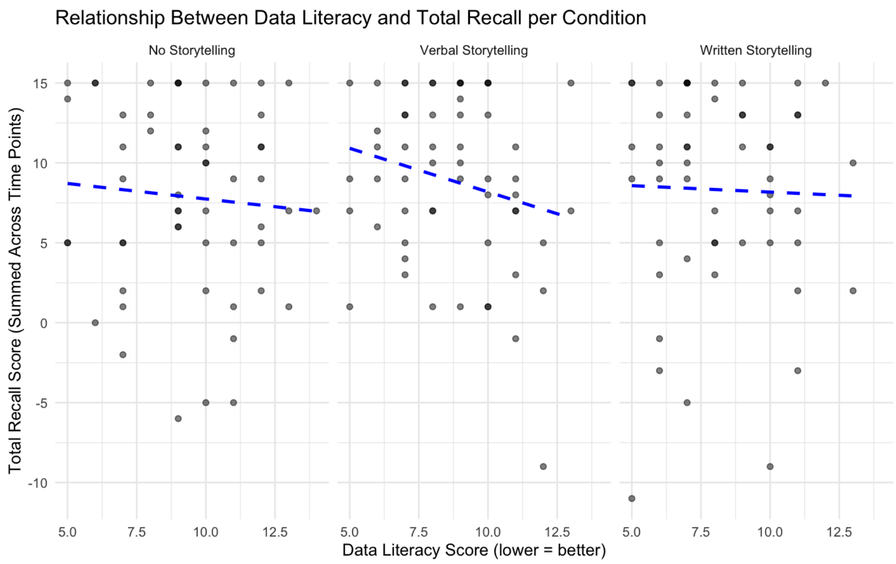


Figure 35: Sum of recall scores by level of data literacy split by conditions. The relationship between data literacy and recall is the steepest for Verbal storytelling.

The visualization shows a negative relationship throughout all conditions, suggesting that participants with higher data literacy also perform better in recalling information. Verbal Storytelling shows the steepest relationship. To test if the relationships are statistically significant, a kendall's correlation is conducted for every condition separately. All results show no statistical significance in the relationship between data literacy and recall (No storytelling:$\tau = -.060$, p = .550; Verbal storytelling: $\tau = -.129$, p = .196; Written storytelling: $\tau = -.057$, p = .583). Following, there's no clear evidence that higher data literacy leads to higher recall within any specific storytelling condition.

Overall, participants with basic data literacy have the lowest recall scores across all time points, those with moderate or high data literacy generally remember more. A small but statistically significant correlation is found between higher data literacy and better recall one month later, suggesting that data literacy may support long-term memory. This finding supports **H3**, people with higher data literacy are more likely to recall information across all storytelling mediums. However, no significant differences are found when looking at the total recall score in each condition separately, even though Verbal storytelling shows the strongest trend. This means that while there are signs that higher data literacy helps with remembering information, especially in the long run, the results are not strong enough to confirm a clear pattern within specific conditions.

4.6 Additional Data Analysis

In this section, further analysis of the data takes place to explore the additional hypotheses from Section 3.4.

4.6.1 Influences on Attitude Change

To explore which variables that were measured best predict how much participants' attitudes change over time, two Random Forest models are used, one for every statistically significant time comparison (Table 7). The outcome variable for the first model is the difference in attitude between before and one month. The model includes the variables condition (storytelling_mode), data literacy, age, political affiliation, topic, previous knowledge, initial attitude score (attitude_before_total) and short video usage. The results are shown in Figure 36. (Breiman, 2001)

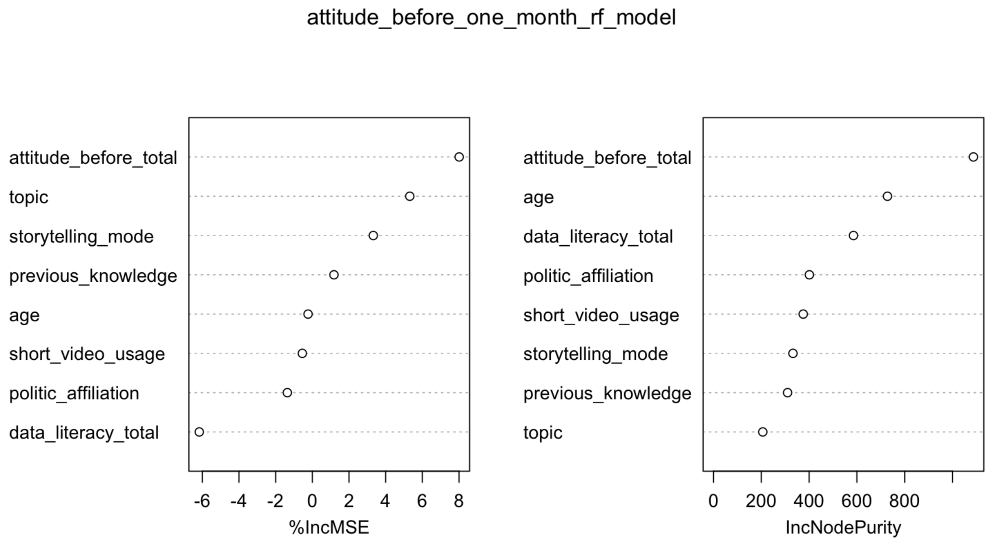


Figure 36: Random forest model with variable importance for predicting attitude change between before and one month.

The right side of the figure (IncNodePurity) shows that initial attitude, age, data literacy, and political affiliation are the most important variables for splitting the data in the decision trees. The left side of the figure (%IncMSE), which shows how much each variable contributed to the model's accuracy, highlights initial attitude, topic, storytelling mode (condition) and previous knowledge as the variables that improve predictions. Other variables have little to no positive effect on accuracy. (Figure 36)

The second Random Forest model uses the difference in attitude between immediate and one month as the outcome variable. The other variables remain the same and the results can be seen below (Figure 37)

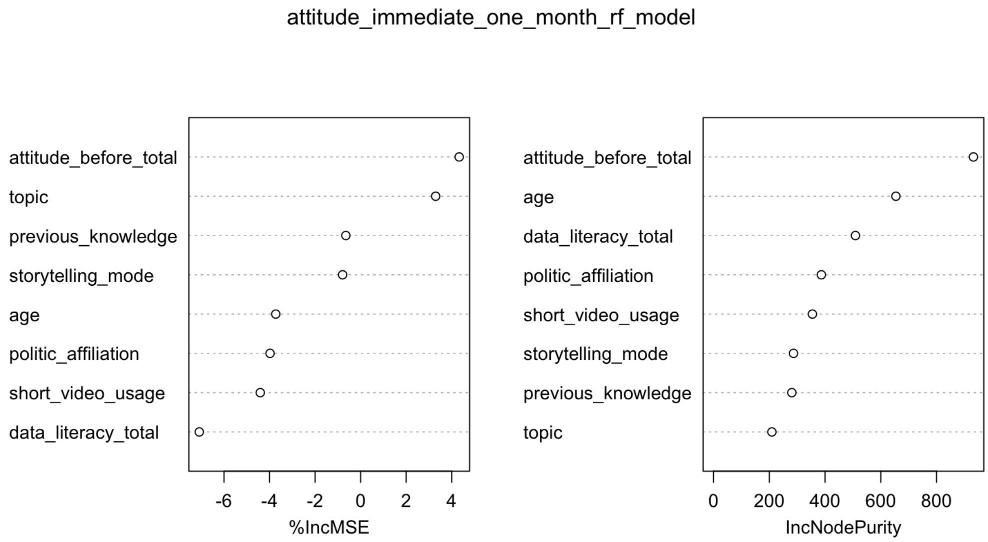


Figure 37: Random forest model with variable importance for predicting attitude change between immediate and one month.

The Random Forest model that predicts the attitude change between immediate and one month reveals that initial attitude, age, data literacy and politic affiliation are most important for splitting the data into decision trees. Initial attitude and topic are the only variables that contribute to the models accuracy. All other variables have no effect and can even lead to noise. (Figure 37)

For both time comparisons, before to one month and immediate to one month, the most important predictor is the initial attitude score. This confirms earlier findings that initial attitudes play a big role in how much participants attitude later shift (Pandey et al., 2014). Age, data literacy, and political affiliation also help separate participants into different groups within the model, but do not improve prediction accuracy. Condition and topic do help improve the model's accuracy in the before - one month model, which supports earlier results showing that attitude change differs over time and between conditions. However, for the immediate - one month model, only immediate attitude and topic improve the model's accuracy. Overall, these results suggest that while the condition has some predictive power, a participant's initial attitude, age, and background (like political views and data literacy) are more consistent indicators of how their attitude will change.

4.6.2 Influences on Recall Scores

To explore the impact of single measured variables on immediate recall, a random forest model is used (Breiman, 2001). The outcome variable is the immediate recall score and the model includes the variables storytelling mode (condition), data literacy, education, previous topic knowledge, short video usage, age and topic. The results can be seen in the following (Figure 38).

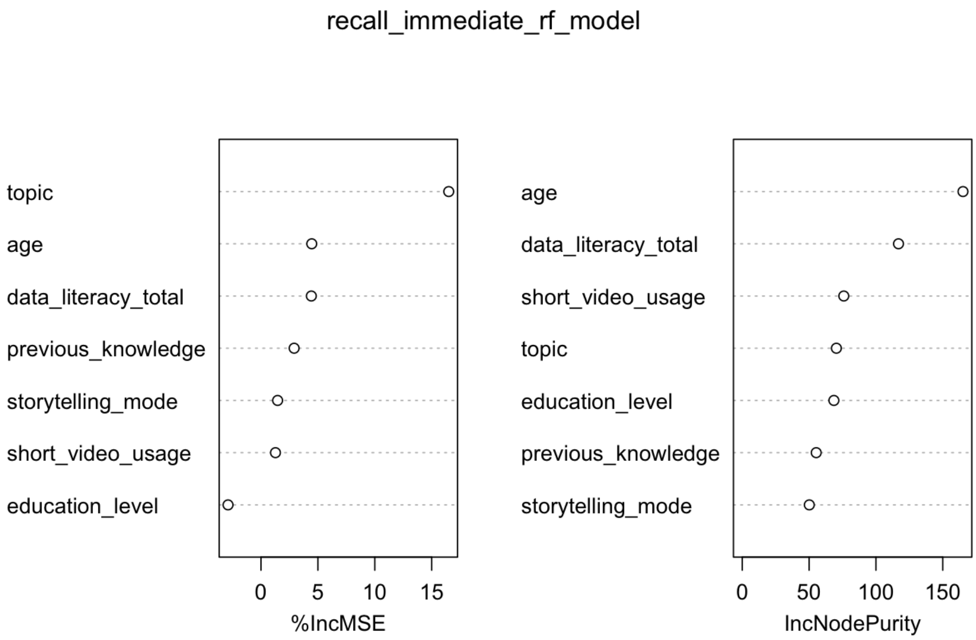


Figure 38: Random forest model with variable importance for predicting immediate recall scores.

Figure 38 shows that age, data literacy and short video usage are most important for splitting the data into decision trees. Topic, age, data literacy, previous knowledge, storytelling mode (condition) and short video usage contribute to the models accuracy. These findings align with the previous findings that recall scores differ significantly across topics. Education level does not lead to more accuracy.

To investigate further how the relationship between previous knowledge and recall and short video experience and recall looks like, a Kendall's tau correlation analysis is run with the variables and immediate recall score. There is no significant correlation between previous knowledge and immediate recall (p = .786, τ = .018) and for previous topic knowledge and immediate recall (p = .294, τ = .070).

4.6.3 Recall over Time

To see if recall scores change over time, the mean scores over time per topic and storytelling condition are visualized first (Figure 39) before a statistical analysis is conducted.

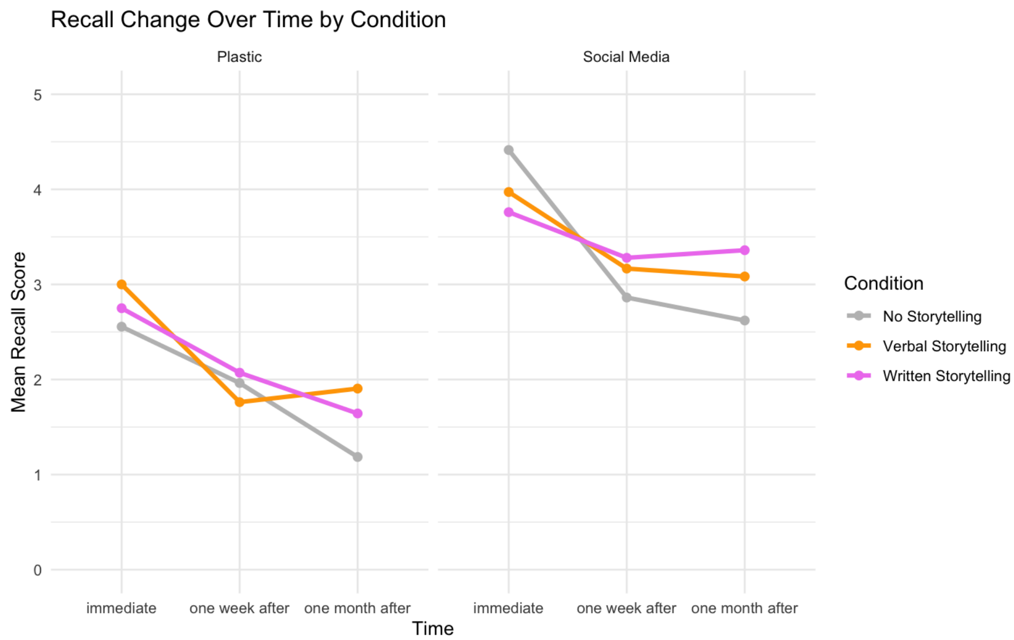


Figure 39: Recall score over time by condition. The trend goes down over time.

To evaluate if a linear mixed model regression is suitable, a baseline model without participants as a random effect is compared to a random intercept model including the participants as a random effect with a likelihood ratio test. This is done for both topics separately. The test shows a significant improvement in model fit when accounting for participant-level variance in recall scores (Social Media: $\chi^2 = 75.739$, $p < .001$, Plastic: $\chi^2 = 58.423$, $p < .001$). The condition is added to the model to see if it improves model fit. For the Social Media topic, adding an interaction between time and condition does not significantly improve prediction of recall scores over time ($\chi^2 = 8.44$, $p = .208$). This suggests that while recall changes over time, the pattern of change is similar across conditions. For Plastic, this behavior is similar ($\chi^2 = 2.49$, $p = .870$). To summarize, the models with participants as a random effect but no condition included explain 56.7% (Social Media) and 54.4% (Plastic) of the variance. (Field et al., 2012, chapter 19)

To see a pairwise comparison of the different time points, estimated marginal means are calculated based on the linear mixed model with participants as a random effect but no condition included. Bonferroni correction for multiple comparisons is used to protect from false positives (Field et al., 2012, chapter 10.5). The results are shown in the following Table 9 with significant changes highlighted ($p < .0167$).

Time Comparison	Social Media		Plastic	
	Estimate	p-value	Estimate	p-value
One Week - Immediate	**-.956**	**< .001**	**-.803**	**.010**
One Month - Immediate	**-1.044**	**< 0.001**	**-1.197**	**< 0.01**
One Month - One Week	-.089	1.000	-.395	.428

Table 9: Estimated recall score change over time by topic. Statistically significant results (Bonferroni-adjusted, $p < .0167$) are highlighted.

The results from Table 9 show that the recall score changes significantly between immediate one week and immediate and one month for both topics. No significant change in scores between one week and one month is present. The drop in scores between immediate and one month is slightly higher than between immediate and one week. This indicates that most forgetting occurs within the first week, with recall stabilizing thereafter.

4.6.4 Influence of Initial Attitude Strength on Attitude Change

The initial attitude is the attitude score measured before the participants get exposed to one of the conditions. The strength of the initial attitude is the amount the total initial attitude score differs from the neutral point. The more it differs from a neutral attitude, the higher the strength. Attitude change is the difference between the total attitude score for each time point from the initial attitude score. In the following, attitude change and attitude strength are visualized together (Figure 40).

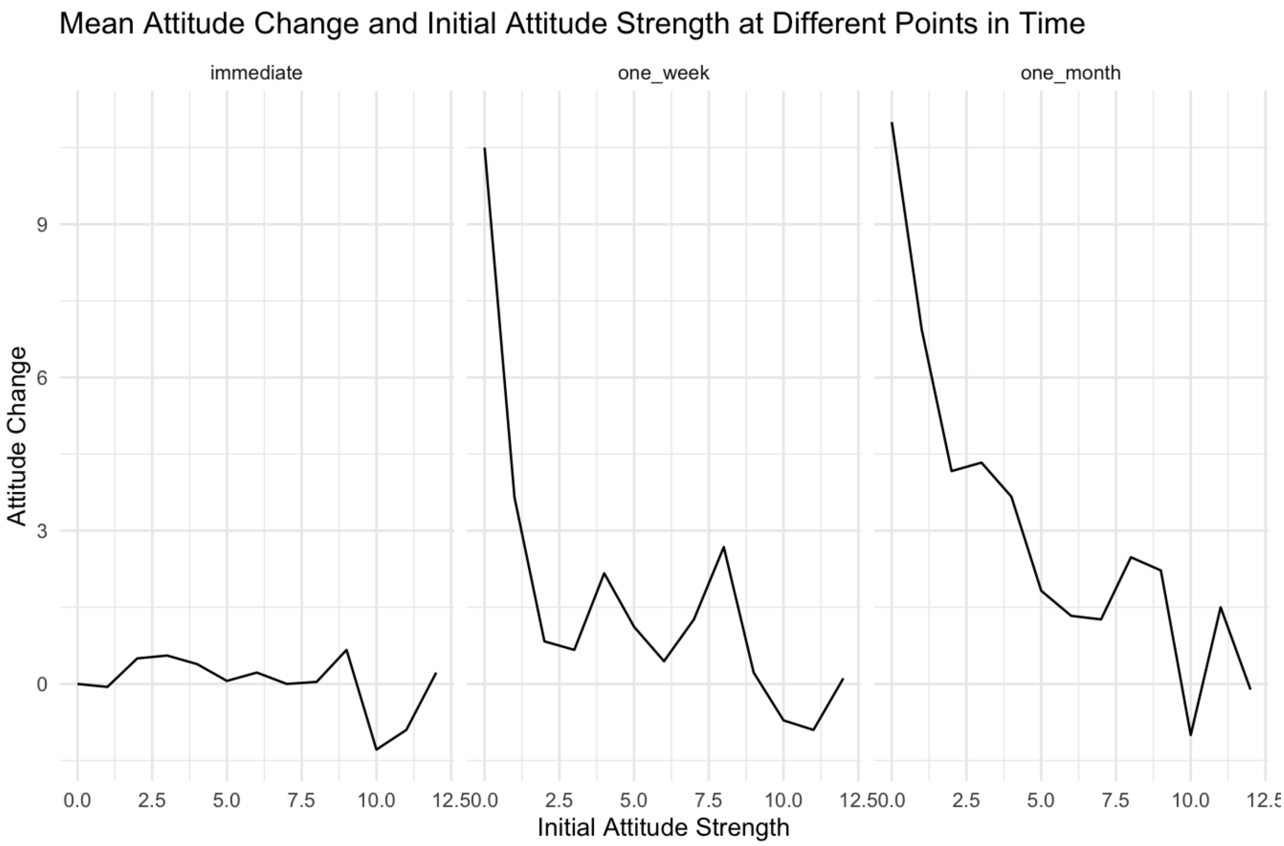


Figure 40: Attitude change and attitude strength at different time points. Attitude change from before until one week and one month is lower if the initial attitude is strong.

The line for attitude change from before until immediate, regardless of the attitude strength is hovering around zero, suggesting little immediate change. Longer term change in attitude after one week and one month shows a negative trend, suggesting the stronger the initial attitude, the less participants' attitude changes. To see if these observations are statistically significant, Kendall's correlation analysis is conducted (DATAtab Team, 2025). The results are shown in the following table (Table 10).

Time Comparison	**Kendall's τ**	**p-value**
Before → Immediate	-.046	.439
Before → One Week	**-.168**	**.003**
Before → One Month	**-.215**	**< .001**
Immediate → One Week	-.096	.096
Immediate → One Month	**-.173**	**.002**
One Week → One Month	**-.126**	**.028**

Table 10: Kendall's tau correlations between initial attitude strength and attitude change at different time comparisons. Statistically significant values ($p < .05$) are highlighted.

The results show that the attitude change from before until one month, before until one week, immediately until one month and one week until one month is influenced by the initial attitude strength. Following, if the initial attitude strength is higher, attitude change tends to decrease over time (long term).

4.6.5 Influence of First Recall on Following Recall

It should be analyzed if the first recall score influences the one-week and one-month recall scores. For that, the distribution of the three variables is checked for normality with a Shapiro-Wilk test and QQ-plot to pick the right method for correlation analysis (Field et al., 2012, chapter 5.6) (Appendix M). The test shows significant results for all variables (immediate recall score: $W = .720$, $p < .001$; one-week recall score: $W = .876$, $p < .001$; one-month recall score: $W = .871$, $p < .001$), meaning that the data is not normally distributed. Following, a kendall's tau correlation is conducted because the data has a lot of tied ranks (Field et al., 2012, chapter 6.5). The results show a statistically significant positive correlation between both, immediate and one-week recall ($p < .001$, $\tau = .457$) and immediate and one-month recall ($p < .001$, $\tau = 0.421$). The relationship is visualized on a linear regression line (Figure 41).

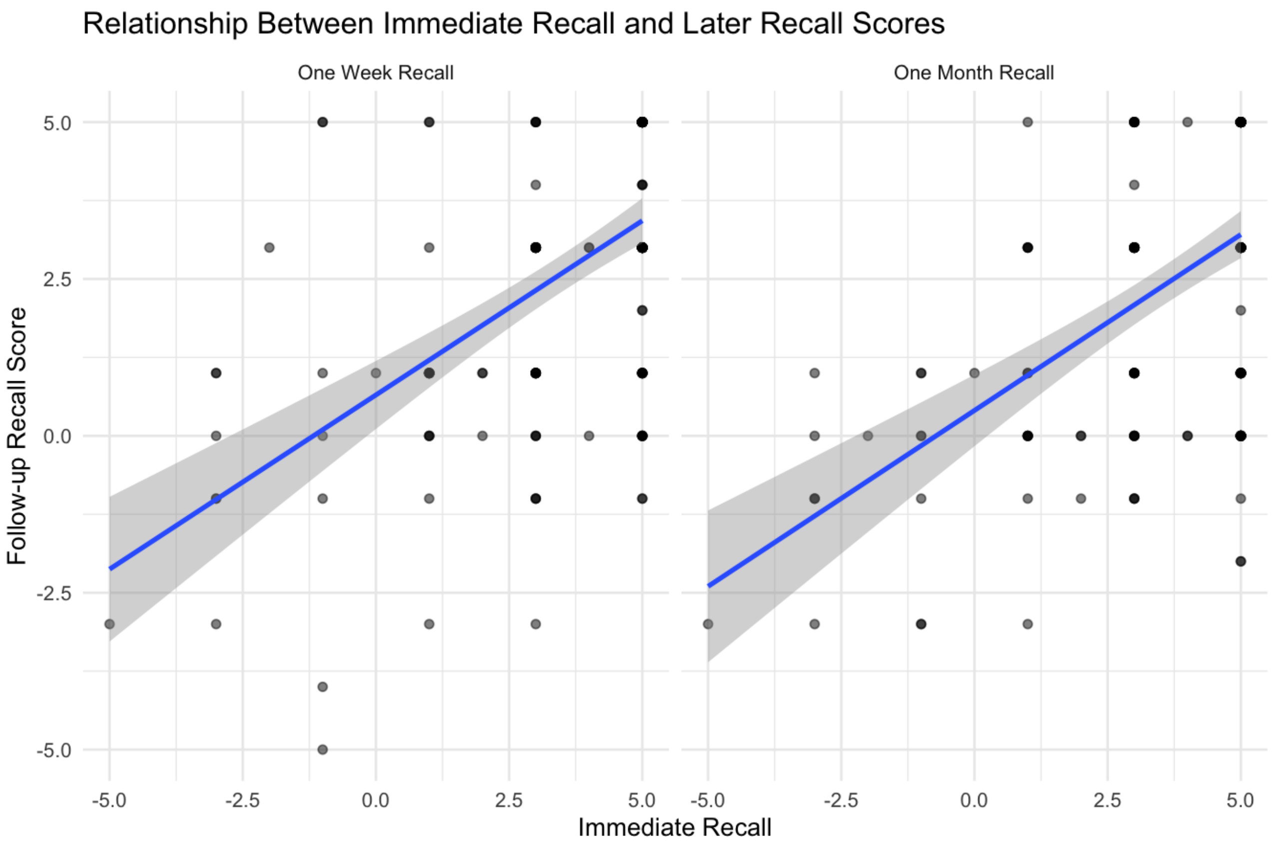


Figure 41: Relationship between immediate recall and one-week and one-month recall scores on a linear regression line.

It can be seen from the linear regression line in Figure 41 that the higher the immediate recall score, the higher the recall score after one week and one month, meaning that there is a relationship between the initial recall score and following recall scores.

4.6.6 Recall and Attitude

To see if there is a relationship between recall score and attitude, meaning if a better recall score leads to an attitude that aligns more with the information the participants are exposed to, a look at:

- immediate recall score and attitude score from before to immediately after
- one-week recall score and attitude score from immediately after to one-week
- one-month recall score and attitude score from one-week to one-month

is taken. All relations are visualized in the following with a linear regression line (Figure 42).

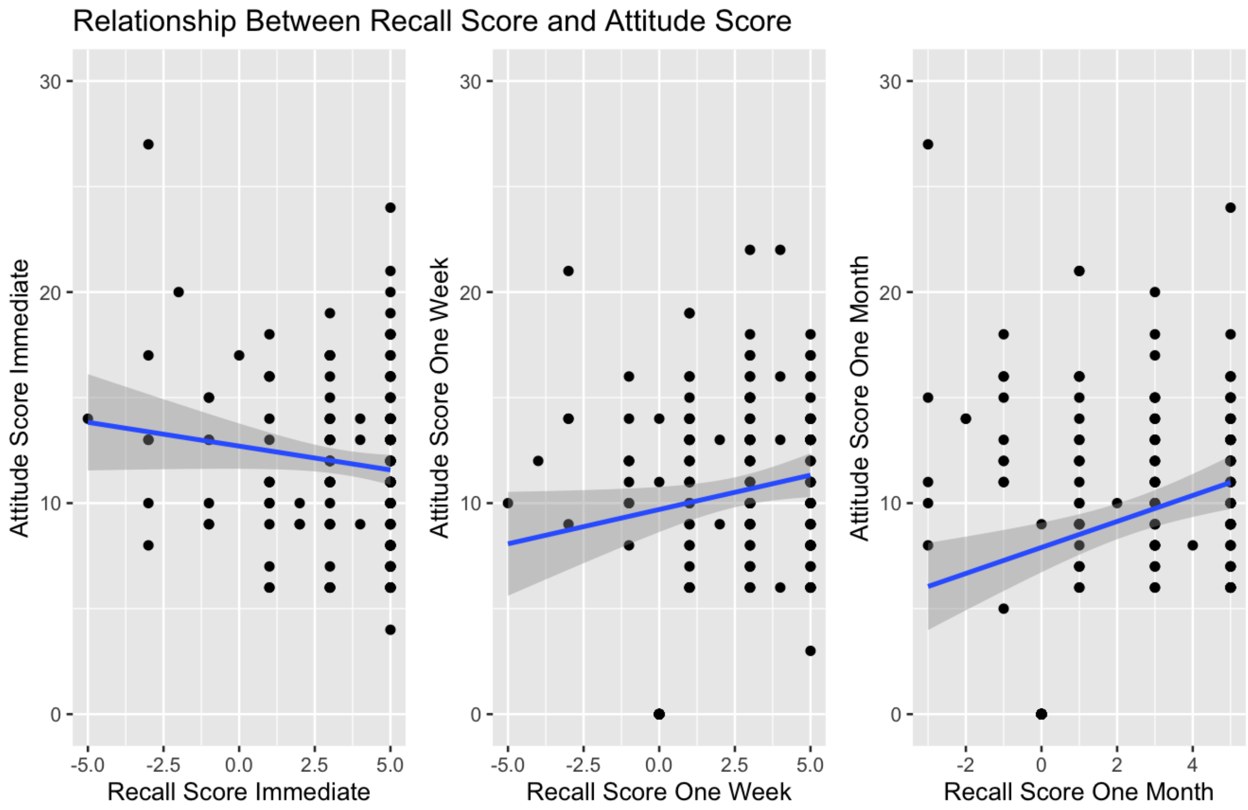


Figure 42: Relationship between recall and attitude score at different time points

Figure 42 shows that there is a negative relationship between recall score and attitude score between before and immediate, meaning that the higher the recall score the lower the attitude score. Thus, the more the attitude aligns with the data the participants were exposed to. Attitude and Recall scores after one week and one month show a positive relationship. This means that the higher the recall score the higher the attitude score, which shows that participants with higher long-term recall don't show attitudes that align with the data they were exposed to one week and one month before. To see if the visualized relationship between the recall and attitude scores is significant Kendall's tau correlation is conducted as recall scores at each time point are not distributed normally (Appendix M) (Field et al., 2012, chapter 6.5). The results of Kendall's tau show no significant correlation between immediate attitude and recall score (p = .424, τ = -.049) and between the one-week recall and attitude score (p = .158, τ = .084). The one-month recall and attitude score show a significant correlation (p = .002, τ = .187), suggesting that participants with a higher recall are not aligning their attitudes with the information they recall in the long term.

4.6.7 Recall and Time Spent on Visualization

The relationship between immediate recall score and time spent is visualized below. The exact time spent on the visualization can only be recorded for the No storytelling condition, where the participant actively decides how long to look at the visualization. Therefore, the storytelling and No storytelling conditions are visualized separately. For the storytelling conditions, instead of time spent on visualization, the number of times the video was played (based on the total time spent on the video screen) is recorded. The No storytelling condition is visualized with a linear trend line (Figure 43).

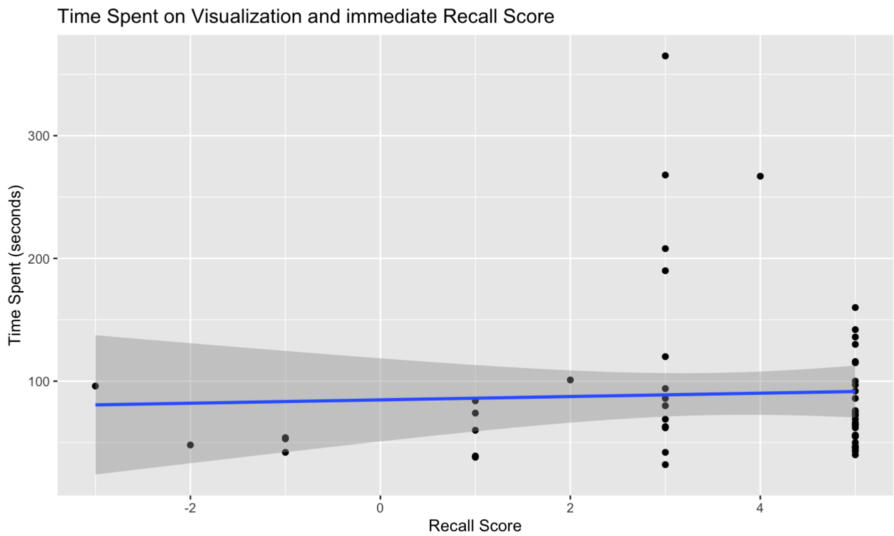


Figure 43: Relation between recall score and time spent on visualization for the no storytelling condition

The visualization shows an almost flat trend line. There are a few extreme outliers but they don't affect the line much. To test if there is a statistically significant correlation, Kendall's tau is used (Field et al., 2012, chapter 6.5). No significant correlation is found (p = .766, τ = .031). This suggests that the time a participant spends on the No storytelling visualization does not effect recall performance. In the following, the recall in the storytelling condition and the number of times the video was played are examined (Figure 44).

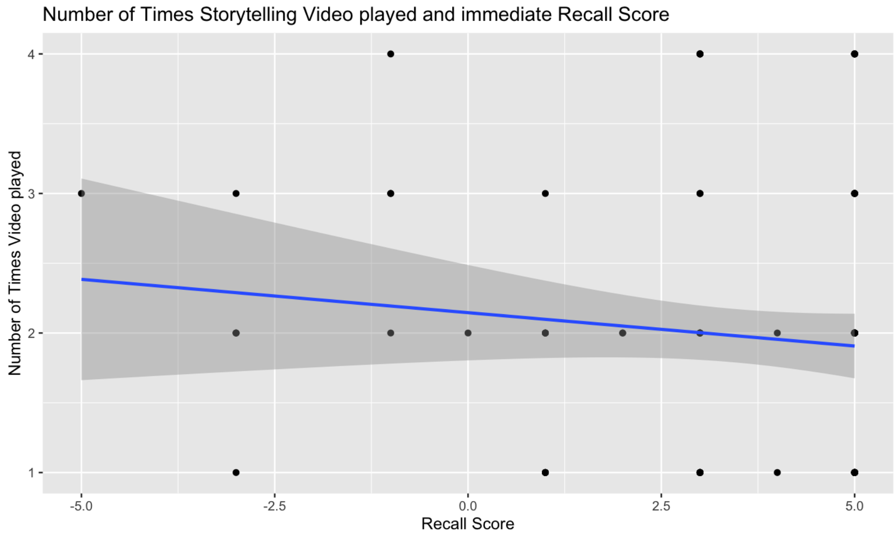


Figure 44: Relation between immediate recall score and number of times the storytelling video is played

Figure 44 shows a trend that the immediate recall score tends to decrease as the video playtime increases. Nevertheless, a correlation analysis with Kendall's tau shows no statistical significance in this trend (p = .249, τ = -.095).

Both results from the No storytelling and storytelling conditions show that there is no relationship between time spent on the visualization/video and recall score. Since there is a correlation between short and long-term recall, no further analysis for the one-week and one-month recall is conducted (Section 4.6.5).

4.6.8 Attitude Change and Time Spent on Visualization

To examine if the time spent on the visualization influences attitude change over time, the relationship is visualized for every time comparison for the No storytelling condition in Figure 45.

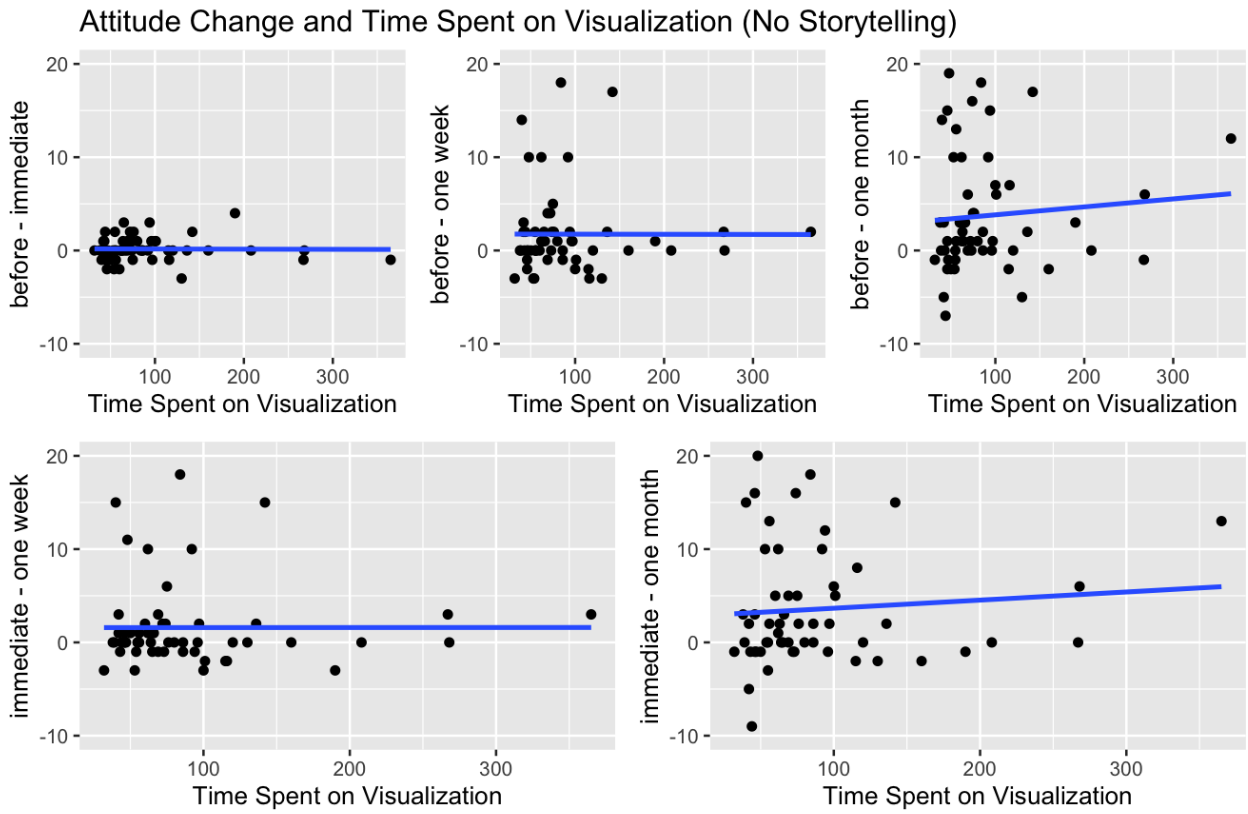


Figure 45: Relation between attitude change over time and time spent for the No storytelling condition.

Across all plots, the blue regression lines are nearly flat or slightly tilted, suggesting a weak or no clear linear relationship between time spent on the visualization and the amount of attitude change in the No storytelling condition. The relationship between attitude change from before and one month and immediate and one month and time spent on visualization reveals a slightly positive trend line. On the other hand, attitude changes from before and immediate, before and one week and immediate and one week show no relationship with time spent on visualization. To statistically test the relationships, Kendall's tau correlation analysis is used. The results can be seen in Table 11.

Time Comparison	**Kendall's τ**	**p-value**
Before $\rightarrow$ Immediate	.083	.410
Before $\rightarrow$ One Week	.041	.676
Before $\rightarrow$ One Month	.109	.250
Immediate $\rightarrow$ One Week	-.013	.891
Immediate $\rightarrow$ One Month	.068	.476

Table 11: Kendall's tau correlation between attitude change and time spent on visualization (No storytelling condition).

None of the correlations are statistically significant ($p < .05$). Following, there is no relationship between attitude change and time spent on visualization for the No storytelling condition.

Looking at attitude change and the number of times the video is played in the storytelling conditions shows that the linear trends are very small. The relationship between before and immediate and before and one week and the number of times a video gets played is slightly positive while the other relationships are slightly negative. (Figure 46)

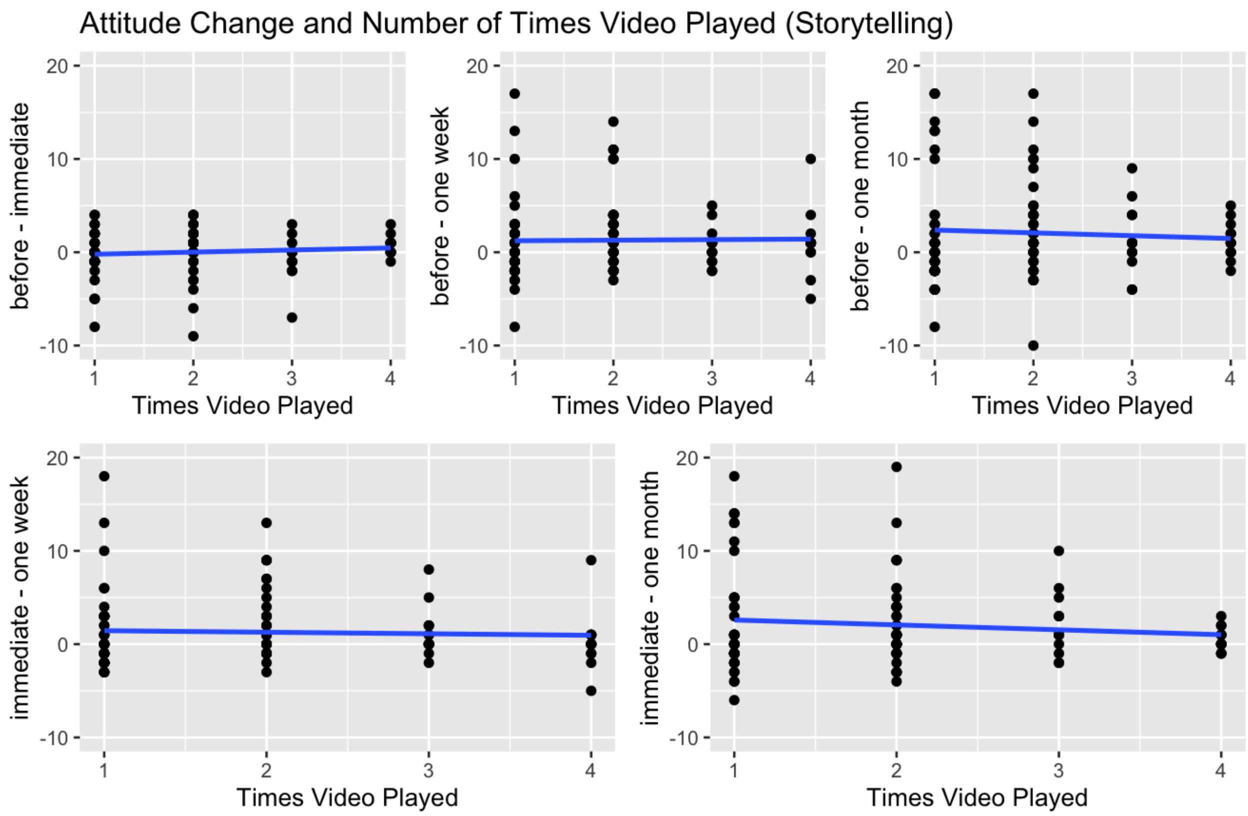


Figure 46: Relation between attitude change over time and number of times a video gets played for the storytelling conditions.

To see if the results are statistically significant Kendall's tau correlation analysis is conducted and the results are presented in Table 12. Again, no statistically significant relationships can be found. Concluding, there is no relationship between the number of times a video gets played and the attitude change of a participant.

Time Comparison	Kendall's τ	p-value
Before → Immediate	.111	.157
Before → One Week	.099	.198
Before → One Month	.069	.364
Immediate → One Week	.068	.375
Immediate → One Month	.021	.786

Table 12: Kendall's tau correlation between attitude change and number of times the video gets played (storytelling conditions only).

4.6.9 Data Literacy and Time Spent on Visualization

To see if there is a relation between data literacy and the time spent on the visualization for the No storytelling condition, a look at the distribution of data literacy across the subsample is taken first (Figure 47)

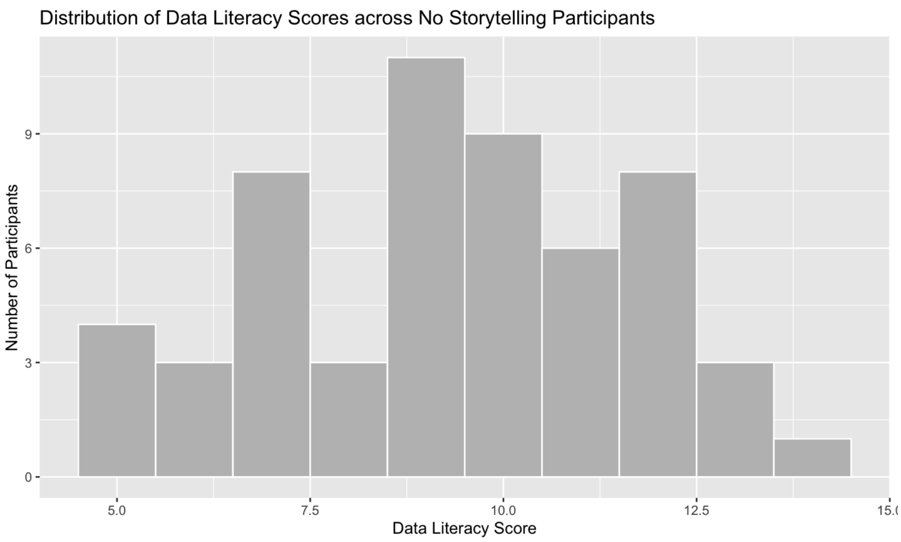


Figure 47: Distribution of data literacy across the No storytelling condition participants. lower score = higher data literacy.

It can be seen that data literacy scores are spread across the sample. Even though there are fewer participants with a very high or very low data literacy, there is a decent number of participants on each level for a correlation analysis (Figure 47). The relationship between time spent in the No storytelling condition and data literacy is visualized first (Figure 48) and after that, a correlation analysis is conducted to see if the correlation is statistically significant. (Field et al., 2012, chapter 6)

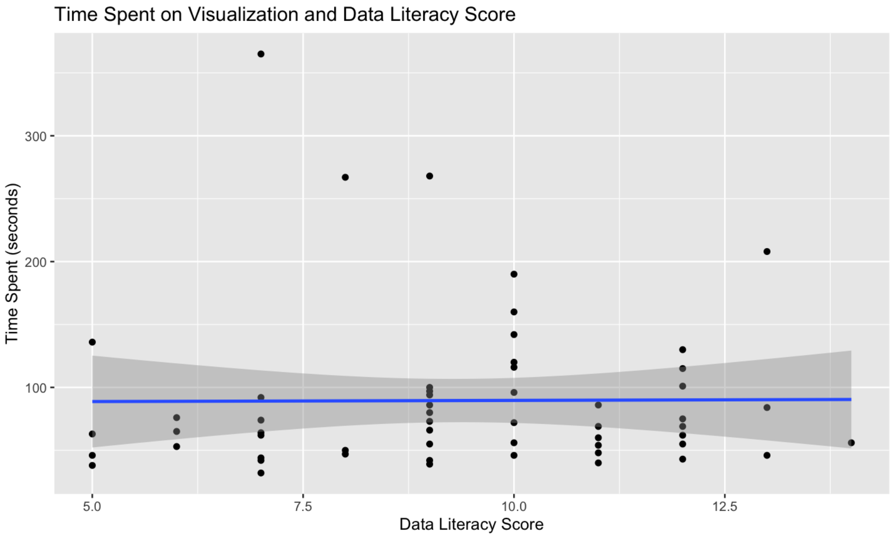


Figure 48: Relation between data literacy score and time spent on visualization (No storytelling) with a linear regression trend line.

Figure 48 shows an almost flat trend line, suggesting that there is no linear relationship between data literacy scores and time spent on visualization. The majority of participants have a time spent around 100 seconds. The visualization shows that there are some extreme outliers with a time spent around 300 seconds. Thus, a Kendall's tau correlation analysis is conducted, since the data also shows many ties (Field et al., 2012, chapter 6.5). As expected from seeing Figure 48, there is no significant correlation between data literacy score and time spent on visualization (p = .324, τ = .096). Furthermore, another look is taken at the data literacy scores of storytelling conditions and the number of times the video gets played (Figure 49).

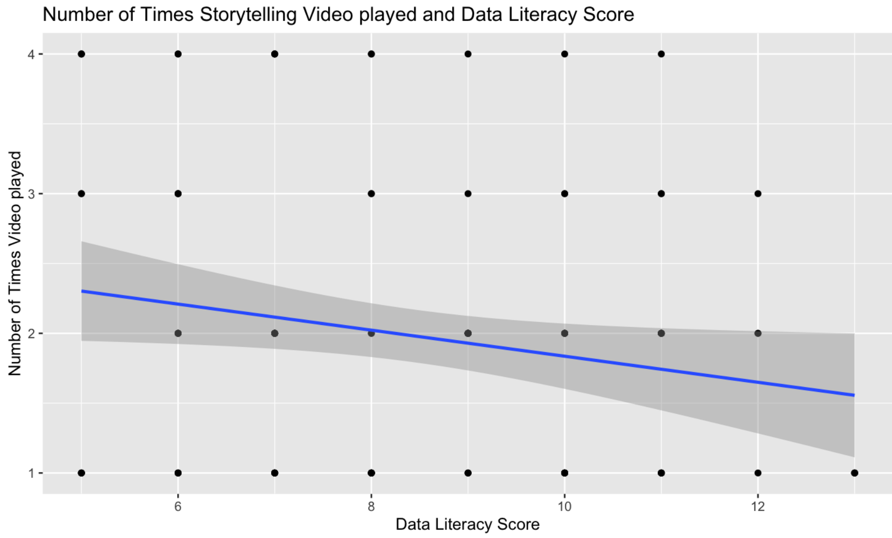


Figure 49: Relation between data literacy score and number of times the storytelling video is played

The visualization shows a negative trend, suggesting that participants with a lower data literacy score, a higher data literacy, are watching the video more often. However, this trend is not statistically significant (p = .100, τ = -.127).

Concluding, no statistically significant relationship between Data Literacy and Time Spent on Visualization can be found. The expectation that participants with a higher data literacy are spending less time or watching the video not as many times as participants with low data literacy, can be rejected in this context. The trend in Figure 49 rather shows the opposite trend that participants with a higher data literacy tend to watch the video more times than people with a lower data literacy. One possible explanation is that participants with higher data literacy engage more critically with the content, perhaps re-watching it to analyze or validate the message.

5 Discussion

The findings of Section 4 are interpreted, tied back to the Research Questions and put into a bigger context in the following. Additionally, expectations defined earlier are also touched on. Ultimately, the limitations of the research and findings are summarized to make potential areas of future research transparent.

RQ1: How might audiences' attitudes change after being presented with data storytelling communicated through different mediums?

Overall, the data shows that attitude changes in favour of the information that is presented over time. In conclusion, presenting people with data has a positive influence on their attitudes in the long term. Following, time has an influence on attitude change suggesting that attitudes develop over time. Initial attitude is an important variable that influences attitude change. Pandey et al. (2014) also suggests that the persuasiveness of visualization depends on the strength of the initial attitude. The more the initial attitude deviates from the information that is presented, the higher the attitude change over time. This aligns with the ELM by Petty and Briñol (2011), which suggests that individuals with strong initial attitudes are more likely to engage in central processing, critically evaluating the information, whereas individuals with weak attitudes might rely on peripheral cues. However, this could be also due to people with attitudes that do not align from the beginning having more room to shift their attitudes than people whose attitudes are already closely aligned with the information. The condition does not significantly influence attitude change. Breaking the data down by conditions shows that a significant change in attitude (in line with the information the participants were presented with) in the long term is only present in Verbal storytelling and No storytelling condition. This suggests that Verbal and No storytelling are more effective in changing long-term attitudes than Written storytelling. This could be due to the latter causing overload to the visual channel (modality principle, (Mayer, 2009)). Additionally, no significant changes in attitude are found immediately and one week after the exposure to the information, only one month after. This could be due to people needing time to process the information and form their new attitudes, maybe even talk about it with their peers to further process it which takes time (Behera and Dadra, 2024). It could also come from the so-called sleeper effect stating that persuasive messages have a delayed effect (Kumkale and Albarracín, 2004). There is no significant difference between the attitude change measured for the No storytelling and the

Verbal storytelling condition. Thus, it can not be said that one medium is more effective than the other and ultimately (verbal) storytelling does not perform better in changing attitudes than No storytelling in this study and Written storytelling even performs worse. However, storytelling is a complex, multifaceted intervention, and its effectiveness may depend on factors such as emotional engagement, narrative coherence, or individual differences in preference for narrative processing. This suggests that adding a narrative to data does not inherently make it more persuasive. If done incorrectly, it can perform worse than data visualizations alone. Additionally, storytelling might be more effective compared to No storytelling for more complex graphs where reading the visualization is not designed as intuitive as possible like in the ones in this experiment. Overall, these findings suggest that in contexts where long-term attitude change is the goal, ensuring that the information resonates with pre-existing attitudes may be more impactful than relying on storytelling mediums. This has implications for the design of educational or persuasive content, emphasizing the need to consider the audience's baseline beliefs in addition to data elements (Quenum et al., 2025). This highlights the importance of user-centric design for data communication.

H1: "Audiences' attitude changes after being presented with data storytelling" is only partially true. Audiences' attitude changes significantly in the long-term only if verbal data storytelling is applied (not written).

H1.1: "Audiences' attitudes are more likely to change if they are presented with verbal data storytelling than written one" is proven in the study.

RQ2: How might audiences' recall differ when being presented with data storytelling communicated through different mediums?

It is discovered that in this study initial recall scores differ significantly between the two topics (plastic and social media). Initial recall scores for the social media topic are higher than the ones for the plastic topic. This could be due to participants' being more familiar with the social media platforms and the emotions they trigger than how different countries handle plastic waste. Aligning with theories of schema-driven memory which posit that individuals recall information more easily when it aligns with existing knowledge structures (Pankin, 2013). There is no effect of the conditions on recall at any of the time points in either of the topics. The lack of effect of storytelling on recall contrasts with claims that narrative structures enhance memory retention (Ginting et al., 2024). This might result from the specific implementation of storytelling in this study or the narrative increasing cognitive load too much or led to little benefit as the data visualization on its own is easy enough to understand. Storytelling (verbal and written) could still influence recall when being applied to support the right kind of data visualization. This hypothesis needs to be examined in future research.

H2: "Audiences' recall improves after being presented with data storytelling" is not proven with statistical significance.

H2.1: "Audiences' recall is more likely to improve when presented with verbal data storytelling compared to written one" is not true.

RQ3: How might the audiences' level of data literacy influence their ability to recall information they were presented with?

Higher data literacy overall leads to better recall scores of the participants in the long-term. This may stem from participants with higher data literacy engaging with the information more critically which enhances memory retention (Dwiprabowo et al., 2024). These results underscore that no single data storytelling format fits all users, reinforcing the need for adaptable, audience-sensitive design strategies (Quenum et al., 2025). Therefore, it is important to consider the target group. However, the effect of data literacy on recall is not found within storytelling conditions which could be due to the smaller sample size or high variations of data literacy levels within the different conditions.

H3: "Audiences with a high data literacy are more likely to recall information they were presented with across all modes of data communication" is true when it comes to long-term recall.

RQ4: How might the audiences' level of data literacy influence their attitude change towards the information they were presented with?

Analyzing the impact of participants level of data literacy reveals that participants with higher data literacy show less immediate attitude change. This could be explained by them being a more critical consumer of information than people with lower data literacy, making them less open to immediate persuasion (Albarracín and Karan, 2022). Another reason could be that higher data literacy leads to more objectivity when looking at the information (focusing more on numbers), reducing the emotional impact that drives immediate attitude change. These findings highlight the importance of designing data stories that prioritize credibility and transparency, particularly for more data-literate audiences who may be less susceptible to emotional appeals. It needs to be said that the effect levels out over time. In the long term participants' attitude change, regardless of their level of data literacy is similar. When adding data literacy to a more comprehensive multilevel model that includes other variables like initial attitude and the conditions, data literacy no longer significantly predicts attitude change. This indicates that while data literacy initially moderates how information is received, its influence may be diluted when considered alongside stronger predictors such as initial attitude. Concluding, in a complex real-world scenario multiple factors interact and again it is important to design the data storytelling for the intended target group and consider their level of data literacy. Nevertheless, looking at data literacy isolated can overestimate its importance in influencing attitude change.

H4: "Audiences with high data literacy are more likely to change their attitude in the long-term compared to audiences with low data literacy" is not true. There is a difference in short-term attitude change where audiences with higher data literacy show less immediate attitude change. There is no difference between audiences with different levels of data literacy when it comes to long-term attitude change.

5.1 Summary of Additional Data Analysis

In the following Table 13, a summary of the findings for the previously defined additional hypotheses is given (Section 3.4).

Hypotheses	**Result**
More attitude change for younger participants	Age has little to no effect on attitude change
Strong political views lead to less attitude change	Political views have little to no effect on attitude change
Extreme attitudes are less likely to change	Strong initial attitude leads to less attitude change in the long-term
Previous topic knowledge leads to better recall	No significant correlation between previous topic knowledge and recall
Experience with watching short videos leads to better recall	No significant correlation between watching short videos and recall
Recall decreases over time	Most forgetting occurs within the first week, after that it stabilizes
Better recall increases attitude change	No correlation between recall and attitude scores
Higher education and data literacy are related	Significant correlation between data literacy and education
Higher time spent on visualization leads to higher recall	No correlation between time spent and recall

Table 13: Additional Hypotheses and Results

Age is not as influential in this specific sample. Additionally, different political views have little to no effect on attitude change which could be due to the topics presented not being political or polarizing enough. Participants not activating prior knowledge during the recall tasks could be the reason for it having no effects on recall. Designing data visualizations and data storytelling in a way that supports the activation of prior knowledge through cues or analogies could change that. A missing correlation between short video usage and recall shows that familiarity with video formats does not translate into deeper cognitive processing and designs must support information encoding and retrieval. The fact that there is no relationship between recall and attitude suggests that remembering information does not guarantee persuasion which is needed to influence attitudes. Finally, there is no significant correlation between time spent and recall. Thus, time spent does not equal cognitive engagement and data storytelling should focus on meaningful interaction and guiding questions.

5.2 Limitations

Some limitations in this study need to be pointed out to put the research into context and evaluate its credibility. This is particularly important since some results seem counterintuitive.

The topics selected for the experiment were of public interest and chosen with the assumption that participants would hold existing attitudes toward them, which were also measured. However, the topics were not highly polarizing and did not introduce groundbreaking new information. This may have limited their perceived relevance or emotional impact, potentially reducing the effectiveness of the storytelling interventions. Including an additional question in the survey to assess how interesting or engaging participants found the topics could have provided deeper insight into how topic perception influenced their responses.

The data visualizations are created in a TikTok-style format but are presented in a controlled experimental setup. In a real-life scenario where people scroll through their social media with different expectations and attention spans than in the controlled experiment they get reimbursed for, they might absorb the information differently and be more interested in one storytelling format than the other. In this study, all conditions were equally controlled to be able to compare them but the motivation for attention remains different.

Another limitation of this survey is the chosen chart type and narrative technique. Different chart types can potentially have different effects on audiences when presented together with storytelling. E.g. more complex graph types could be more effective when communicated in a storytelling format than basic graph types like bar charts. Moreover, in this study, the data storytelling arc is chosen as a linear, author-driven narrative structure. Other narrative techniques that are not author-driven or have a different structure can also lead to different effects on participants' attitudes and recall. Additionally, an AI-generated voice is used in this experiment even though Mayer (2005) suggests that the human voice is more effective for narration. This could have potential effects on the generalization of results in this experiment and must be further explored in the future.

Additionally, it is difficult to replicate the power of truly compelling storytelling. Although this study follows evidence-based guidelines for crafting and delivering narratives, a story's effectiveness also depends on factors beyond structure or content. Skilled storytellers, such as former U.S. President Barack Obama, possess a unique ability to connect emotionally with their audience by drawing on timing, tone, presence, and cultural resonance (Obama, 2009). Consequently, the storytelling interventions employed in this study may not fully capture the impact of professionally crafted, contextually attuned stories delivered in real-world scenarios.

The online survey that is used to capture the data also comes with some limitations as additional data can not be collected and it can not be checked if the intended respondent answers the questionnaire or somebody else. Furthermore, a survey means receiving self-reported measures that can be influenced by different biases (Section 3.10). Thus, it is not possible to measure unconscious processes or behavior changes. This has especially an impact on measuring attitude as only explicit attitudes are measured but implicit attitudes are not even though they play a significant role (Sheets et al., 2011). The research design does not include qualitative insights such as open-ended questions which limits deeper understanding of the participant's responses (e.g. emotional reactions and interpretations).

In addition to that, the majority of the participant sample of this study is well educated. It needs to be considered that for people who are less educated, data storytelling and visualizations might have a different effect. Another limitation is the sample size. Even though 166 participants are recruited, when breaking them down into the six conditions the sample size gets significantly smaller which could lead to a reduced statistical power (higher risk of type 2 errors).

6 Conclusion

This thesis investigates how different storytelling mediums influence attitude change and information recall, and how these effects vary based on the audience's level of data literacy. A quantitative analysis reveals that overall, attitudes shift in favour of the presented information, but only after one month, suggesting that attitude change is a delayed process. In addition to time, initial attitude is a key predictor of change, while the condition (Verbal, Written, No storytelling) has limited influence. Verbal storytelling and No storytelling lead to long-term attitude change, but Written storytelling does not. This suggests that with easy graphs and an author-driven narrative as used in this study, Verbal storytelling is more effective than Written storytelling. Nevertheless, Verbal storytelling is not more effective than No storytelling. Following, adding a narrative does not inherently lead to an attitude change. If designed incorrectly, it can perform worse than a data visualization alone. Additionally, neither of the conditions affect recall, contradicting claims that narratives enhance recall. However, long-term recall is higher for participants with higher data literacy, highlighting the importance of tailoring data communication to user expertise. Although higher data literacy reduces immediate attitude change, this difference evens out over time. This suggests that critical engagement may delay, but not prevent, attitude shifts. These findings emphasize the need for user-centric, data-driven communication that considers initial attitudes, literacy levels, and the time required for meaningful change.

This research contributes to the growing field of data storytelling and communication by offering empirical evidence that Verbal and Written storytelling can have different effects on attitude change, with Verbal storytelling showing more long-term impact. At the same time, it shows that data storytelling does not improve recall nor attitude change compared to sole data visualization, which is contradictory to existing research and suggests that the benefits of narrative formats may not be universal but depend on the complexity and type of visualization, topic and target audience. These findings underscore the importance of user-focused design in data visualization. Rather than relying on a one-size-fits-all approach, effective communication must consider individual differences such as data literacy, prior attitudes, and background knowledge. Storytelling is not inherently persuasive or memorable, it must be meaningfully aligned with the needs and capacities of its audience.

Effective data communication is not just about finding the perfect format, but about creating adaptive, audience-sensitive designs rooted in empathy for the people behind the numbers. The most suitable storytelling medium depends on the audience and the complexity of the data being communicated.

6.1 Future Works

The findings of this study are a valuable step toward understanding the impact of data storytelling, but they cannot be generalized broadly, as only one type of data visualization and narrative was examined, and the sample size is limited.

Future research should further investigate whether verbal data storytelling is more effective in changing attitudes than written formats. In particular, it remains to be explored how different types of data visualizations, especially more complex or abstract formats interact with verbal storytelling. Narrative techniques may become more powerful when audiences are faced with higher cognitive demands. Similarly, comparing different narrative structures beyond the linear, author-driven approach used in this study, such as interactive formats may reveal new insights into how stories can support comprehension and persuasion.

A critical direction for future work is improving the quality and authenticity of the stories themselves. While this study followed established storytelling principles, the narratives were not created by professional storytellers. A skilled communicator can influence how information is received by designing and delivering a story that tailors tone, rhythm, emotional resonance, and timing. This would make data storytelling studies more representative of real-world communication and help establish a stronger link between experimental findings and practical applications.

Additionally, future studies could examine whether AI-generated voices affect storytelling perception differently than human voices. Embedding data stories in realistic environments, such as social media feeds, would also help assess how recall and persuasion unfold when participants are exposed to competing stimuli. To deepen the understanding of recall and attitude change, future research should incorporate implicit measures in addition to self-reported ones. Finally, integrating qualitative insights, such as open-ended questions or thinkaloud protocols could offer a richer, more holistic understanding of how individuals engage with and interpret data storytelling.

References

Abbott, E. E. (1909). On the analysis of the factor of recall in the learning process. *The Psychological Review: Monograph Supplements*, 11(1):159. Publisher: The Review Publishing Company.

Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50:179–211.

Albarracin, D. and Shavitt, S. (2018). Attitudes and attitude change. *Annual review of psychology*, 69(1):299–327. Publisher: Annual Reviews.

Albarracín, D. and Karan, A. (2022). Resistance to persuasion. In *Oxford research encyclopedia of psychology*.

Aparicio, M. and Costa, C. J. (2015). Data visualization. *Communication design quarterly review*, 3(1):7–11. Publisher: ACM New York, NY, USA.

Baumeister, R. F. and Leary, M. R. (1997). Writing narrative literature reviews. *Review of general psychology*, 1(3):311–320. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Bechler, C. J., Tormala, Z. L., and Rucker, D. D. (2021). The Attitude–Behavior Relationship Revisited. *Psychological Science*, 32(8):1285–1297. _eprint: https://doi.org/10.1177/0956797621995206.

Behera, C. K. and Dadra, R. (2024). Understanding young consumers' attitude formation for new-age fintech credit products: an SOR framework perspective. *Journal of Financial Services Marketing*, 29(3):964–978. Publisher: Springer.

Berger, P. D., Maurer, R. E., and Celli, G. B. (2002). Experimental design. *CA (USA): Wadsworth Group Belmont*. Publisher: Springer.

Bonikowska, A., Sanmartin, C., and Frenette, M. (2019). Data literacy: What it is and how to measure it in the public service.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32. Publisher: Springer.

Britannica, T. E. o. E. (2024). "recall".

Briñol, P., Petty, R. E., and Guyer, J. J. (2019). A historical view on attitudes and persuasion. *Oxford Encyclopedia of the history of psychology*, pages 1–34. Publisher: Oxford University Press New York^ eNY NY.

Buckland, M. and Gey, F. (1994). The relationship between recall and precision. *Journal of the American society for information science*, 45(1):12–19. Publisher: Wiley Online Library.

BULUT, E. (2024). Social Media Usage and Emotional Well-Being.

Carpenter, S. K., Pashler, H., and Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic bulletin & review*, 13(5):826–830. Publisher: Springer.

CBS (2025). Statistics Netherlands. https://www.cbs.nl/en-gb.

Charness, G., Gneezy, U., and Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, 81(1):1–8. Publisher: Elsevier.

Choi, B. C. and Pak, A. W. (2005). Peer reviewed: a catalog of biases in questionnaires. *Preventing chronic disease*, 2(1). Publisher: Centers for Disease Control and Prevention.

Clark, T., Foster, L., Bryman, A., and Sloan, L. (2021). *Bryman's social research methods*. Oxford university press.

Cleveland, W. S. and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554. Publisher: Taylor & Francis.

Cote, C. (2021). Data Storytelling: How to effectively tell a story with data. https://online.hbs.edu/blog/post/data-storytelling.

Cronin, P., Ryan, F., and Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British journal of nursing*, 17(1):38–43. Publisher: MA Healthcare London.

Cui, Y., Chen, F., Lutsyk, A., Leighton, J. P., and Cutumisu, M. (2023). Data literacy assessments: A systematic literature review. *Assessment in Education: Principles, Policy & Practice*, 30(1):76–96. Publisher: Taylor & Francis.

DATAtab Team (2025). Spearman's rank correlation coefficient. https://datatab.net/tutorial/spearman-correlation.

Dijkstra, E. and Goedhart, M. (2012). Development and validation of the ACSI: Measuring students' science attitudes, pro-environmental behaviour, climate change attitudes and knowledge. *Environmental Education Research*, 18(6):733–749. Publisher: Taylor & Francis.

Dowarah, K., Duarah, H., and Devipriya, S. P. (2022). A preliminary survey to assess the awareness, attitudes/behaviours, and opinions pertaining to plastic and microplastic pollution among students in India. *Marine Policy*, 144:105220. Publisher: Elsevier.

Dwiprabowo, R., Supena, A., Rakhman, G. G. F., Sumantri, M. S., Islahuddin, I., Nugraheny, D. C., and Faujiah, E. (2024). Framework Brain-Based Learning: The Impact of Critical Thinking-Based Learning on Memory and Recall. *ELEMENTARY: Islamic Teacher Journal*, 12(1):75–88.

Dykes, B. (2019). *Effective data storytelling: how to drive change with data, narrative and visuals*. John Wiley & Sons.

Eckert, H.-W. (2022). *Storytelling with Data: Gaining insights, developing strategy and taking corporate communications to a new level*. Springer.

Ehret, P. J., Monroe, B. M., and Read, S. J. (2015). Modeling the dynamics of evaluation: A multilevel neural network implementation of the iterative reprocessing model. *Personality and Social Psychology Review*, 19(2):148–176. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Eldridge, S. (2025). political spectrum. https://www.britannica.com/topic/political-spectrum.

Engebretsen, M. and Kennedy, H. (2020). *Data visualization in society*. Amsterdam university press.

Ferrari, R. (2015). Writing narrative style literature reviews. *Medical writing*, 24(4):230–235. Publisher: Taylor & Francis.

Field, A., Field, Z., and Miles, J. (2012). Discovering statistics using R. Publisher: Sage.

Frank, M., Walker, J., Attard, J., and Tygel, A. (2016). Data Literacy-What is it and how can we make it happen? *The Journal of Community Informatics*, 12(3).

Geukjian, S. (2025). Fish and Overfishing. https://www.kaggle.com/datasets/sergegeukjian/fish-and-overfishing.

Ginting, D., Woods, R. M., Barella, Y., Limanta, L. S., Madkur, A., and How, H. E. (2024). The Effects of Digital Storytelling on the Retention and Transferability of Student Knowledge. *SAGE Open*, 14(3). Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Glasman, L. R. and Albarracín, D. (2006). Forming attitudes that predict future behavior: a meta-analysis of the attitude-behavior relation. *Psychological bulletin*, 132(5):778. Publisher: American Psychological Association.

Goldstein, D. G., McAfee, R. P., and Suri, S. (2011). The effects of exposure time on memory of display advertisements. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 49–58.

Gorilla (2024). Build your online experiments. https://gorilla.sc.

Gummer, E. S. and Mandinach, E. B. (2015). Building a Conceptual Framework for Data Literacy. *Teachers College Record*, 117(4):1–22. _eprint: https://doi.org/10.1177/016146811511700401.

health_blogger (2023). Unlocking the Palette of Feelings with the Color Wheel of Emotions. https://www.ascendik.com/unlocking-the-palette-of-feelings-with-the-color-wheel-of-emotions/#page-content.

Hogan, R. M. and Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10(5):562–567. Publisher: Elsevier.

Kaggle (2024). Datasets. https://www.kaggle.com/datasets.

Kennedy, H. and Engebretsen, M. (2020). Introduction: The relationships between graphs, charts, maps and meanings, feelings, engagements. In Kennedy, H. and Engebretsen, M., editors, *Data Visualization in Society*, pages 19–32. Amsterdam University Press.

Kintsch, W. (1970). Models for free recall and recognition. *Models of human memory*, pages 331–373. Publisher: Academic Press.

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Ksiazkiewicz, A., Klemmensen, R., Dawes, C. T., Christensen, K., McGue, M., Krueger, R. F., and Nørgaard, A. S. (2020). Sources of stability in social and economic ideological orientations: Cohort, context, and construct effects. *International Journal of Public Opinion Research*, 32(4):711–730. Publisher: Oxford University Press.

Kumkale, G. T. and Albarracín, D. (2004). The sleeper effect in persuasion: a meta-analytic review. *Psychological bulletin*, 130(1):143 – 172. Publisher: American Psychological Association.

Laming, D. (2021). Recall of advertisements after various lapses of time. *Psychological Research*, 85(6):2428–2443. Publisher: Springer.

Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.

Liem, J., Perin, C., and Wood, J. (2020). Structure and empathy in visual data storytelling: Evaluating their influence on attitude. In *Computer Graphics Forum*, volume 39, pages 277–289. Wiley Online Library. Issue: 3.

Lund University (2025). Block Search. https://libguides.lub.lu.se/c.php?g=677619&p=4828808.

Mat Daud, A. A. (2022). Five common misconceptions regarding flattening-the-curve of COVID-19. *History and Philosophy of the Life Sciences*, 44(3):41.

Matsumoto, A. (2025). How to ask gender survey questions on sexual orientation and identity? https://www.surveymonkey.com/curiosity/ask-survey-questions-sexual-orientation-gender-identity/.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 41(1):31–48.

Mayer, R. E. (2009). Modality Principle. In *Multimedia Learning*, pages 200–220. Cambridge University Press.

Miro (2025). The visual workspace for innovation. https://miro.com/about/.

Moreno, R. and Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of educational psychology*, 91(2):358. Publisher: American Psychological Association.

Munzner, T. (2014). *Visualization analysis and design*. CRC press.

Nakagawa, S. and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution*, 4(2):133–142. Publisher: Wiley Online Library.

NeuroLaunch editorial team (2024). Decay Psychology: Understanding Memory Loss and Forgetting. https://neurolaunch.com/decay-psychology-definition/.

Newman, N., Fletcher, R., Robertson, C. T., Arguedas, A. R., and Nielsen, R. K. (2024). Digital News Report 2024. Technical report, Reuters Institute for the Study of Journalism.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Oba, D. and Berger, J. (2024). How communication mediums shape the message. *Journal of Consumer Psychology*, 34(3):406–424. Publisher: Wiley Online Library.

Obama, B. (2009). President Obama Makes Historic Speech to America's Students - English subtitles. https://youtu.be/5a5VSkkdpB4?si=oezAYXdjWuVKbUQA.

Ojo, A. and Heravi, B. (2018). Patterns in award winning data storytelling: Story types, enabling tools and competences. *Digital journalism*, 6(6):693–718. Publisher: Taylor & Francis.

Olson, K. (2021). What are data? Issue: 9 Pages: 1567–1569 Publication Title: Qualitative Health Research Volume: 31.

Pandey, A. V., Manivannan, A., Nov, O., Satterthwaite, M., and Bertini, E. (2014). The persuasive power of data visualization. *IEEE transactions on visualization and computer graphics*, 20(12):2211–2220. Publisher: IEEE.

Pankin, J. (2013). Schema Theory.

Petty, R. E. and Briñol, P. (2011). The elaboration likelihood model. *Handbook of theories of social psychology*, 1:224–245. Publisher: Sage Publications Ltd Los Angeles.

Pickard, A. J. (2013). *Research methods in information*. Facet publishing.

Prolific (2025). Prolific. https://www.prolific.com.

Proton AG (2025). Complete guide to GDPR compliance. https://gdpr.eu/?cn-reloaded=1.

QlikTech International AB (2025). How Data Literate Are You? https://thedataliteracyproject.org/assessment/.

Quenum, V., Gagnon, E., and Mosconi, E. (2025). Enhancing Data Storytelling: A Conceptual Framework.

Qurban, B. (2024). Global Health Indicators Dataset. https://www.kaggle.com/datasets/bushraqurban/world-health-indicators-dataset/data.

Rdocumentation (2025). emmeans: Estimated marginal means (Least-squares means). https://www.rdocumentation.org/packages/emmeans/versions/1.10.5/topics/emmeans.

Ritchie, H., Samborska, V., and Roser, M. (2023). Plastic Pollution.

Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, 55(1):68. Publisher: American Psychological Association.

Sadiku, M., Shadare, A. E., Musa, S. M., Akujuobi, C. M., and Perry, R. (2016). Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 2(12):11–16.

Schöner, G. (1989). Learning and recall in a dynamic theory of coordination patterns. *Biological Cybernetics*, 62(1):39–54. Publisher: Springer.

Segel, E. and Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics*, 16(6):1139–1148. Publisher: IEEE.

Shao, H., Martinez-Maldonado, R., Echeverria, V., Yan, L., and Gasevic, D. (2024). Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension? In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–21.

Sheets, P., Domke, D. S., and Greenwald, A. G. (2011). God and country: The partisan psychology of the presidency, religion, and nation. *Political Psychology*, 32(3):459–484. Publisher: Wiley Online Library.

Silverman, J. (2023). Great Pacific Garbage Patch: The World's Biggest Landfill in the Pacific Ocean. https://science.howstuffworks.com/environmental/earth/oceanography/great-pacific-garbage-patch.htm.

Song, H., Healey, J., Siu, A. F., Wigington, C., and Stasko, J. (2023). Experts prefer text but videos help novices: an analysis of the utility of multi-media content. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–9.

SurveyMonkey (2025). Demographic survey question guide: Tips, examples, & best practices. https://www.surveymonkey.com/mp/gathering-demographic-information-from-surveys/.

The Danish National Center for Ethics (2024). The Four Scientific Ethical Principles. https://researchethics.dk/research-ethics-themes/the-four-scientific-ethical-principles.

The R Foundation (2025). The R Project for Statistical Computing. https://www.r-project.org.

TikTok (2025). TikTok. https://www.tiktok.com/about.

Vehovar, V., Toepoel, V., and Steinmetz, S. (2016). *Non-probability sampling*, volume 1. The Sage handbook of survey methods.

Weber, W., Engebretsen, M., and Kennedy, H. (2018). Data stories: Rethinking journalistic storytelling in the context of data journalism. *Studies in communication sciences*, 2018(1):191–206. Publisher: Seimso.

Wickham, H. and Grolemund, G. (2016). R for Data Science.

Wilke, C. O. (2019). *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media.

WRAP (2025). Understanding recycling symbols. https://www.recyclenow.com/how-to-recycle/recycling-symbols.

Yasser, M. (2022). Titanic Dataset. https://www.kaggle.com/datasets/yasserh/titanic-dataset.

Zdanovic, D., Lembcke, T. J., and Bogers, T. (2022). The influence of data storytelling on the ability to recall information. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, pages 67–77.

Zhang, Y., Reynolds, M., Lugmayr, A., Damjanov, K., and Hassan, G. M. (2022). A Visual Data Storytelling Framework. *Informatics*, 9(4).

Zmigrod, L. (2020). The role of cognitive rigidity in political ideologies: Theory, evidence, and future directions. *Current Opinion in Behavioral Sciences*, 34:34–39. Publisher: Elsevier.

7 Appendix

Appendices can be found directly in this section, in the attached file under the file names stated below, or through their hyperlink.

A Reduction of Dataset for Plastic Waste Data

Plastic Waste Dataset:

Entity	Year	Share of waste recycled from total regional waste	Share of waste incinerated from total regional waste	Share of littered and mismanaged from total regional waste	Share of waste landfilled from total regional waste
Americas (excl. USA)	2002	xx	xx	xx	xx
Americas (excl. USA)	2003	xx	xx	xx	xx
United States	2017	xx	xx	xx	xx
Europe	2017	xx	xx	xx	xx

Reduce/ Filter:

Entity	Year	Share of waste recycled from total regional waste	Share of waste incinerated from total regional waste	Share of littered and mismanaged from total regional waste	Share of waste landfilled from total regional waste
Americas (excl. USA)	2002	xx	xx	xx	xx
Americas (excl. USA)	2003	xx	xx	xx	xx
United States	2017	xx	xx	xx	xx
Europe	2017	xx	xx	xx	xx

Reduce/ Aggregate:

Entity	sum of share of waste recycled from total regional waste	sum of share of waste incinerated from total regional waste	sum of share of littered and mismanaged from total regional waste	sum of share of waste landfilled from total regional waste
Americas (excl. USA)	xx	xx	xx	xx
United States	xx	xx	xx	xx
Europe	xx	xx	xx	xx

Figure 50: Reduction of the plastic waste dataset to prepare it for the visualization

B R Code for Data Visualizations

The R code for the data visualizations can be found in the following folder of the attached file:
Appendix/Data_visualization_R

C Data Storytelling Arc for Plastic Waste Data

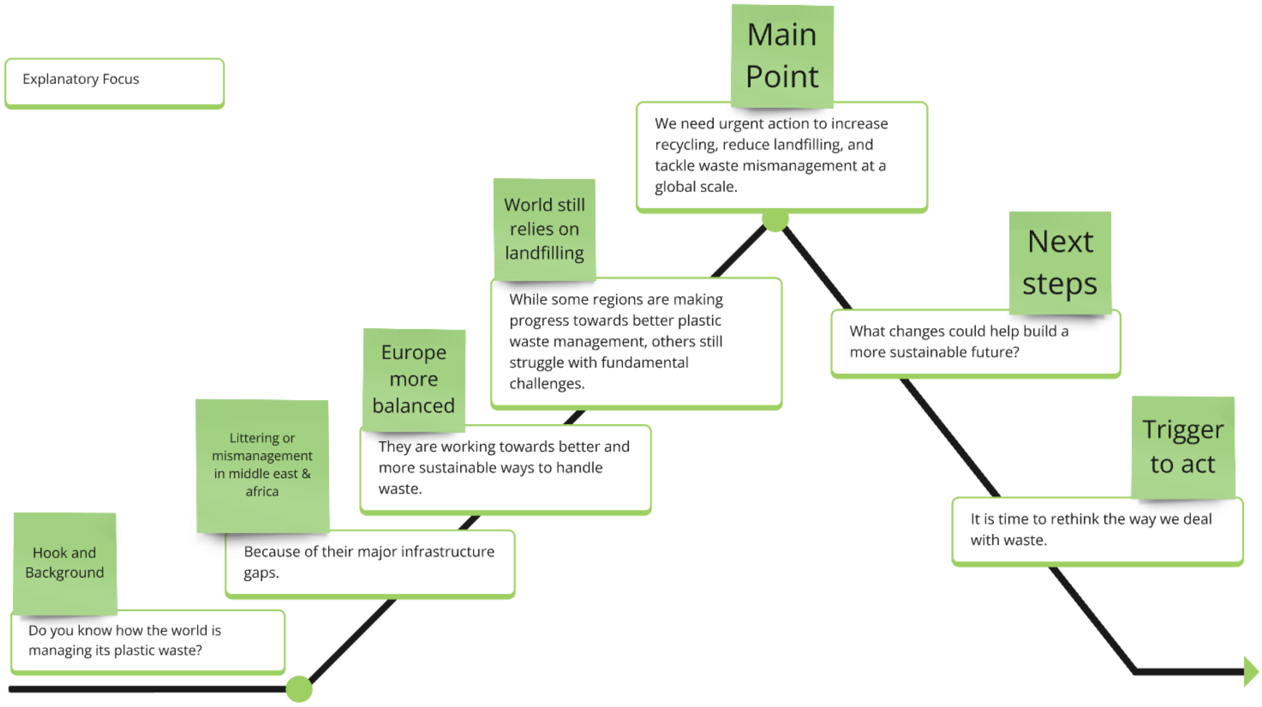


Figure 51: Data Storytelling Arc for the plastic waste data (Dykes, 2019, chapter 6)

D Experimental Conditions

Four videos used in the experiment can be found in the following folder of the attached file: *Appendix/Experiment_conditions*

E Survey Consent Form

Welcome to the study

Please read the information about the study below carefully.

What is this study about?
You are being asked to participate in a longitudinal study about different methods of communicating insights about data on a topic of general importance and the effects of these different methods. This study is part of a master's thesis in the Information Studies degree at Aalborg University in Copenhagen, Denmark.

What will participation involve?
You will be shown one data visualization and given the time to study it carefully. After that, you will be asked a set of questions about the visualisation and your attitudes towards the topic in question. We wish to ask you about your attitudes at different points in time. This means we will contact you again 1 week after completing the initial study and 1 month after to answer additional questions. You will be contacted through Prolific each time. Completing the third set of questions after 1 month will complete your participation in the study.
All three study parts combined will take approximately 20 minutes in total. None of the questions are of a sensitive or personal nature. You will be compensated for your participation through Prolific after you have completed all three questionnaires. Your reward of £3 will be deposited in your Prolific account.

Your contribution
Your participation in this study is a critical step towards understanding how best to present raw and complex data by visualizing it in such a way that makes it easier to understand and remember.

Privacy
We do not collect any of your personal information, in compliance with the EU’s General Data Protection Regulation (GDPR). Apart from sending you reminders after 1 week and 1 month to complete your participation, we will not contact you about this study. If you have any questions about the study, please contact us at tnagel23@student.aau.dk.

- ○ I consent to participating in the study described above
- ○ I do not consent to participating in the study described above

Figure 52: Text of Survey Consent Form

F Questions about Attitude and Recall

Questions about Attitude: Social Media and Mental Health

I believe social media platforms should take responsibility for promoting content that supports mental well-being.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- I don't know

I am willing to spread awareness about the influence social media can have on mental health

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- I don't know

I am willing to avoid social media platforms that make me feel anxious, stressed, or insecure.

- Strongly Agree
- Agree
- Neutral
- Disagreee
- Strongly Disagree
- I don't know

I think I rarely use social media in ways that negatively impact my mood or mental health

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- I don't know

I am willing to actively engage with content or communities on social media that focus on mental health and self-care.

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- I don't know

I am aware of how social media can affect my mental health positively and negatively and take steps to minimize negative effects

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- I don't know

Figure 53: Questions about Attitude towards social media and mental health

Questions about Attitude: Plastic Waste Management

I am willing to pay more for biodegradable plastic alternatives

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- Don't know

I take reusable shopping bags with me to the market/shop

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- Don't know

After using a plastic product, I reuse it or make sure it goes in the recycling bin

- Strongly Agree
- Agree
- Neutral
- Disagreee
- Strongly Disagree
- Don't know

I use single use plastics such as plastic cups, bags, plates, bottles etc. rarely

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- Don't know

I am willing to spread awareness about plastic pollution to my friends and family

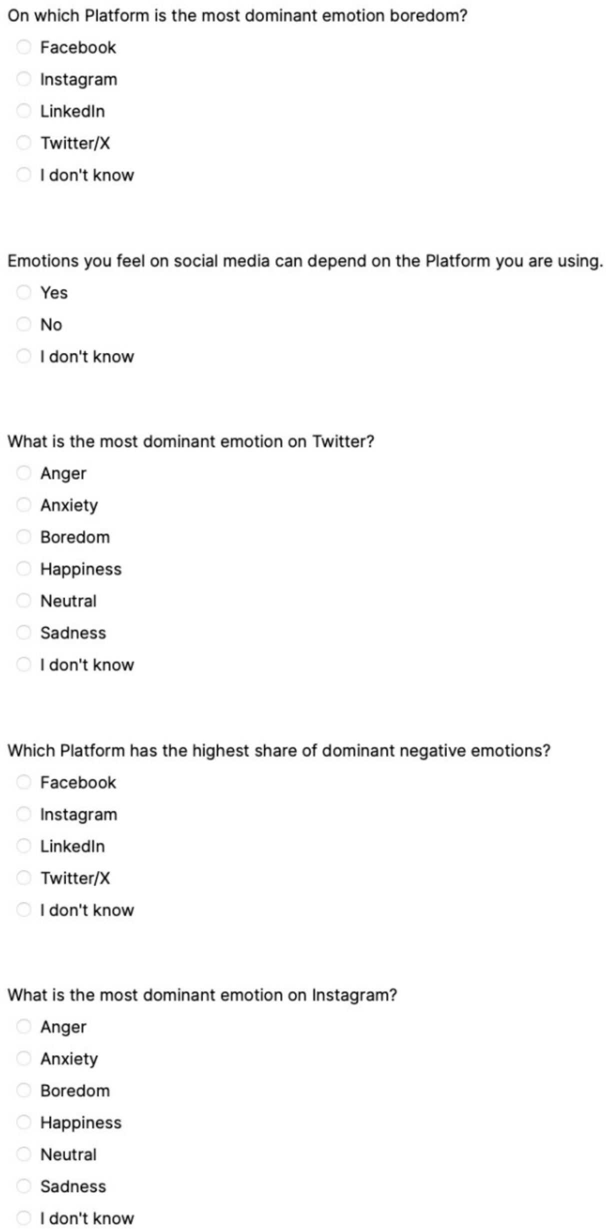
- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- Don't know

I do not prefer plastic bags to other packaging made from natural materials like paper, jute or cloth

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
- Don't know

Figure 54: Questions about Attitude towards plastic waste management

Questions about Recall: Social Media and Mental Health



On which Platform is the most dominant emotion boredom?

- Facebook
- Instagram
- LinkedIn
- Twitter/X
- I don't know

Emotions you feel on social media can depend on the Platform you are using.

- Yes
- No
- I don't know

What is the most dominant emotion on Twitter?

- Anger
- Anxiety
- Boredom
- Happiness
- Neutral
- Sadness
- I don't know

Which Platform has the highest share of dominant negative emotions?

- Facebook
- Instagram
- LinkedIn
- Twitter/X
- I don't know

What is the most dominant emotion on Instagram?

- Anger
- Anxiety
- Boredom
- Happiness
- Neutral
- Sadness
- I don't know

Figure 55: Questions to measure recall of social media and mental health insights

Questions about Recall: Plastic Waste Management

What region is best at adopting more modern and sustainable waste management approaches?

- ○ Europe
- ○ America
- ○ Middle East & Africa
- ○ Asia
- ○ Oceania
- ○ I don't know

How is the majority of plastic waste in the world managed?

- ○ Landfilled
- ○ Littered and mismanaged
- ○ Incinerated
- ○ Recycled
- ○ I don't know

The majority of plastic waste is managed in a sustainable way.

- ○ Yes
- ○ No
- ○ I don't know

How is the majority of waste managed in the Middle East & Africa?

- ○ Landfilled
- ○ Littered and mismanaged
- ○ Incinerated
- ○ Recycled
- ○ I don't know

Which region faces the biggest challenges in managing waste responsibly?

- ○ Europe
- ○ America
- ○ Middle East & Africa
- ○ Asia
- ○ Oceania
- ○ I don't know

Figure 56: Questions to measure recall of plastic waste management insights

G Raw Survey Data

The full raw survey data can be found in the following folder of the attached file:
Appendix/2104_final_data_export

H R Code for Analysis

The R code for the data analysis can be found in the following folder of the attached file. A guide about how to run the R code can be found in the readme.Rmd file:
Appendix/Experiment_analysis

I QQ-Plot Data Literacy Score

Figure 57: QQ-Plot of data literacy score. The data is not normally distributed

J QQ-Plot Attitude Before by Topic

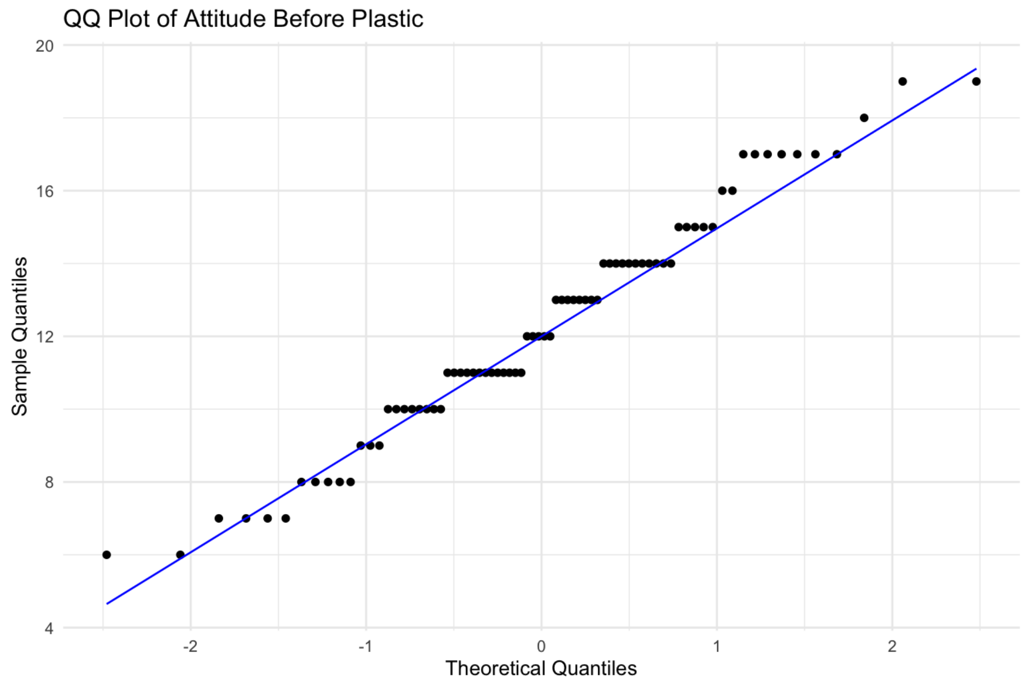


Figure 58: QQ-Plot of attitude before for the topic plastic. The data is normally distributed

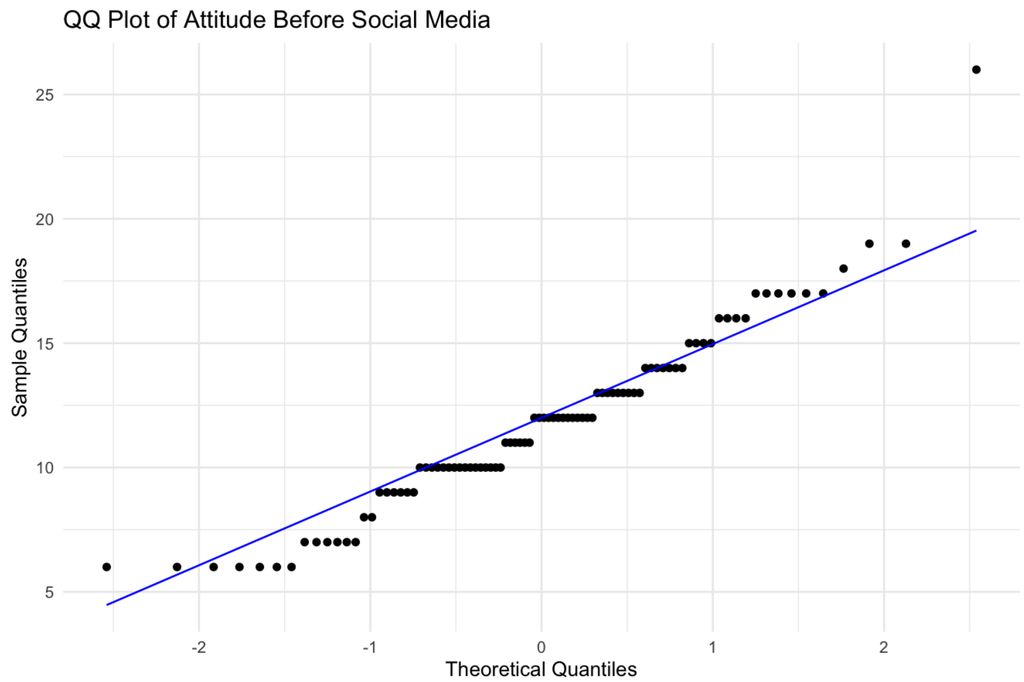


Figure 59: QQ-Plot of attitude before for the topic social media. The data is not normally distributed

K QQ-Plot Attitude Change by Storytelling Mode

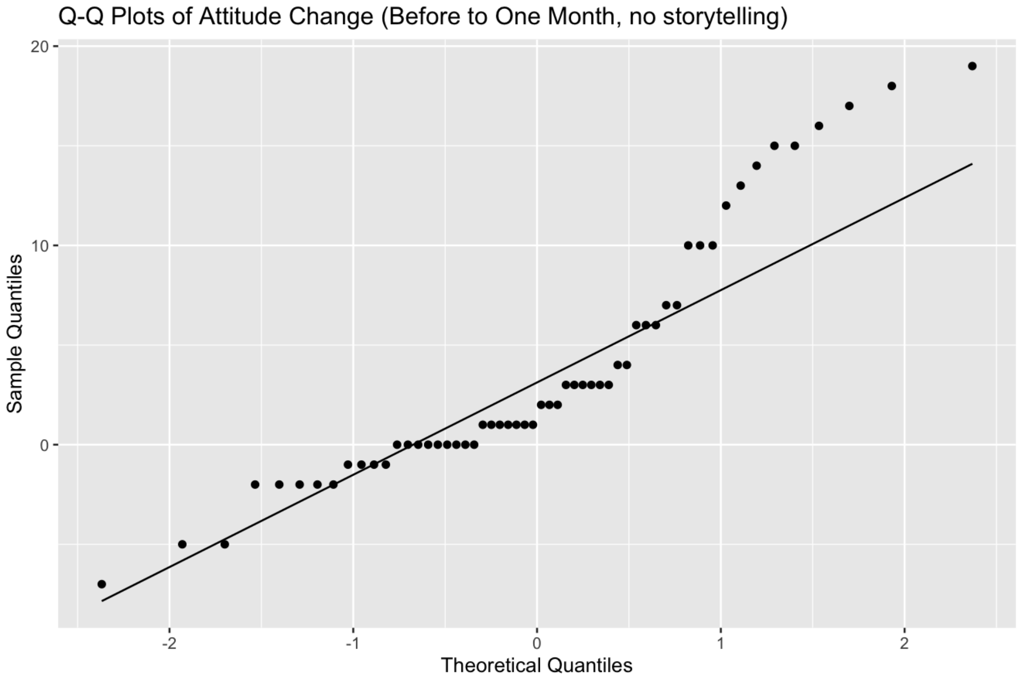


Figure 60: QQ-Plot of attitude change before - one month for the no storytelling condition. The data is not normally distributed

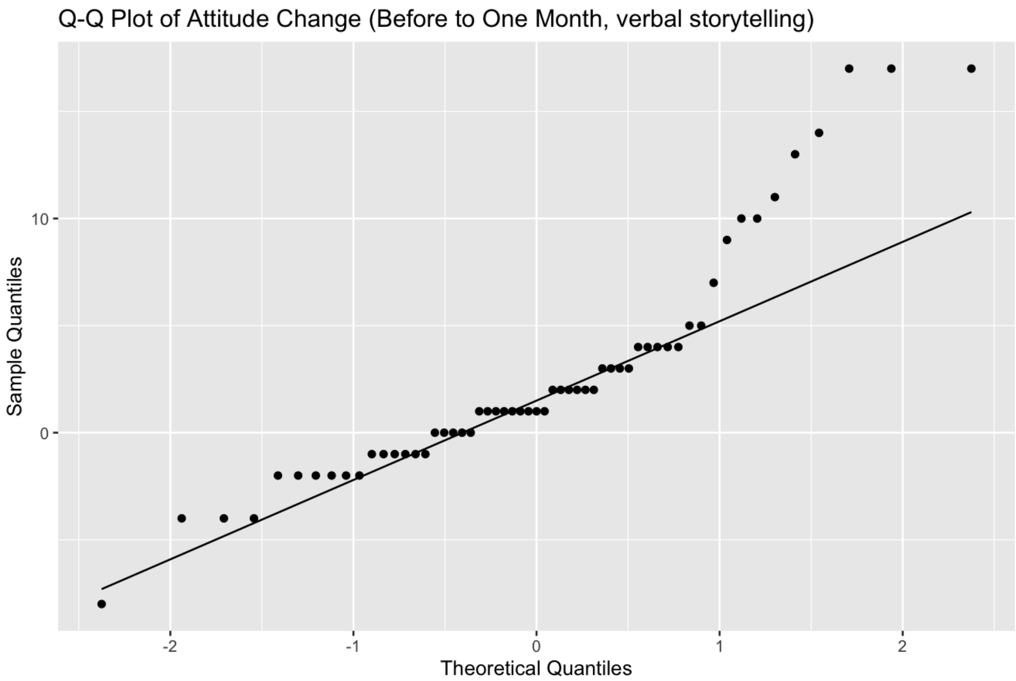


Figure 61: QQ-Plot of attitude change before - one month for the verbal storytelling condition. The data is not normally distributed

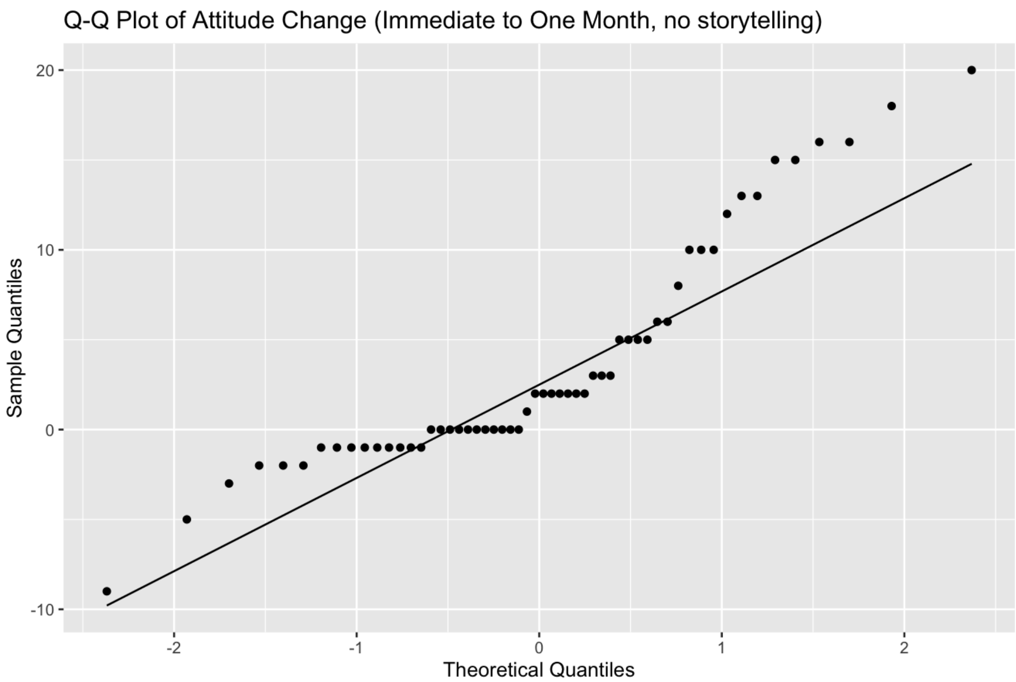


Figure 62: QQ-Plot of attitude change immediate - one month for the no storytelling condition. The data is not normally distributed

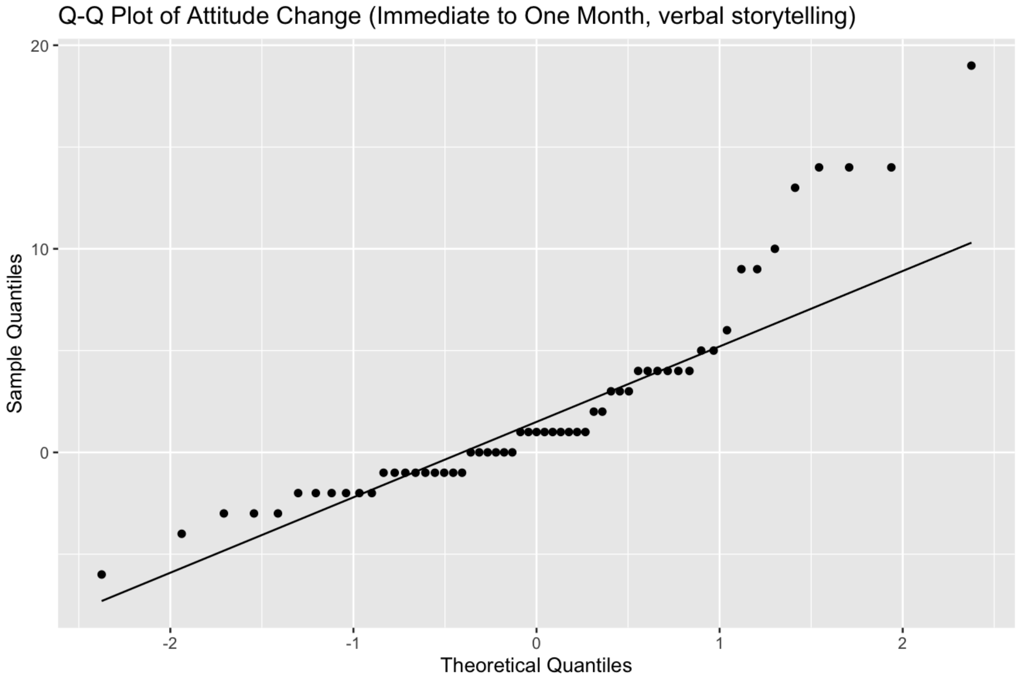


Figure 63: QQ-Plot of attitude change immediate - one month for the verbal storytelling condition. The data is not normally distributed

L QQ-Plot Immediate Recall by Topic

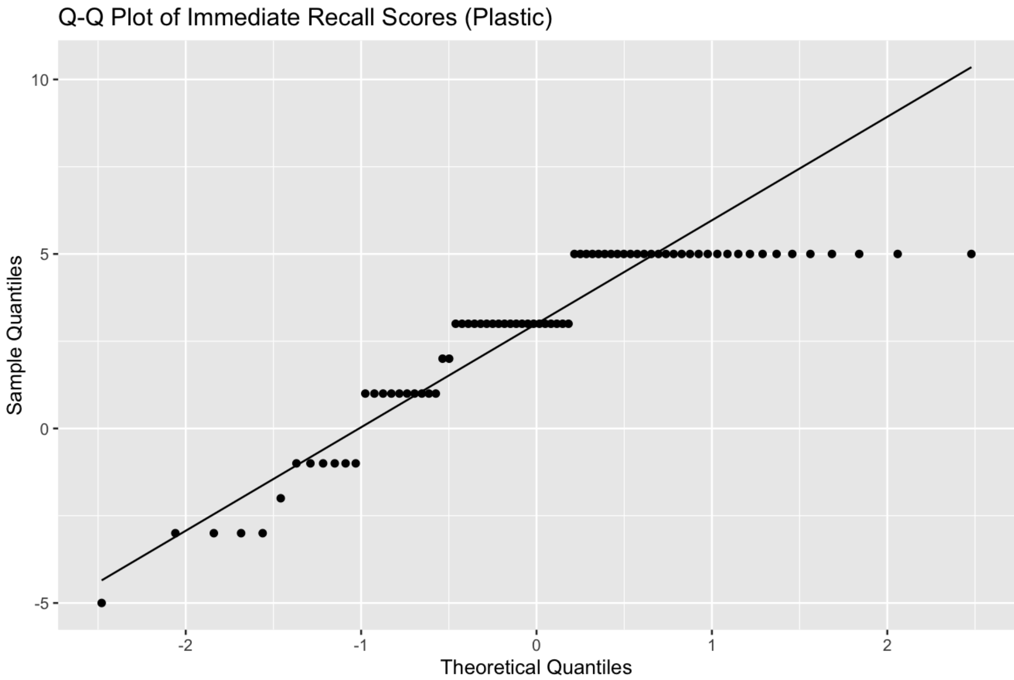


Figure 64: QQ-Plot of immediate recall for the topic plastic. The data is not normally distributed

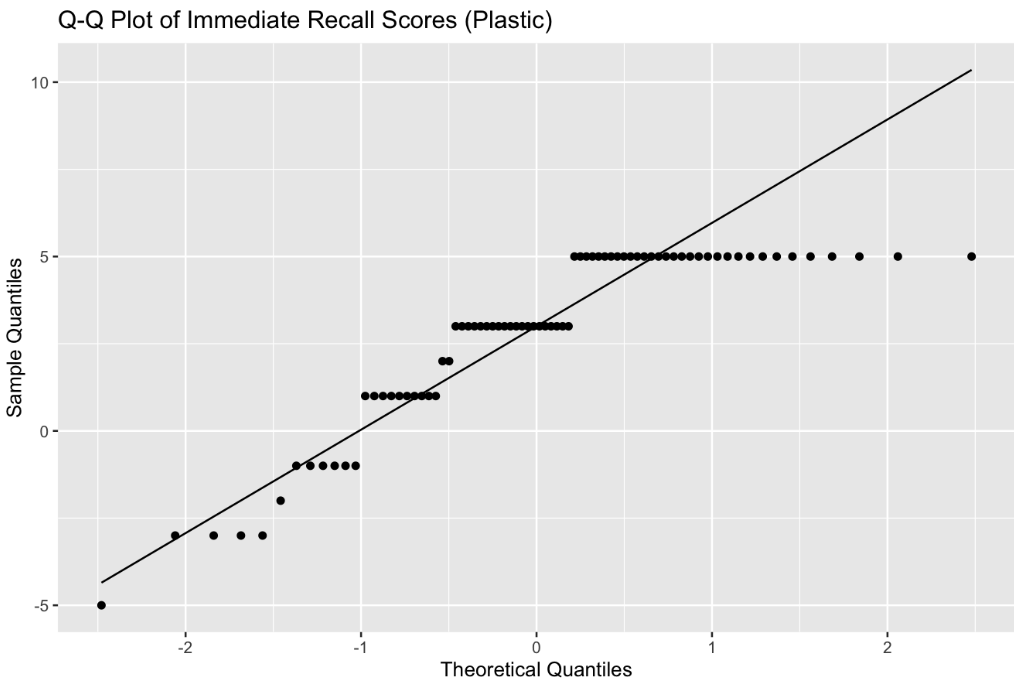


Figure 65: QQ-Plot of immediate recall for the topic social media. The data is not normally distributed

M QQ-Plot Recall Scores at Different Times

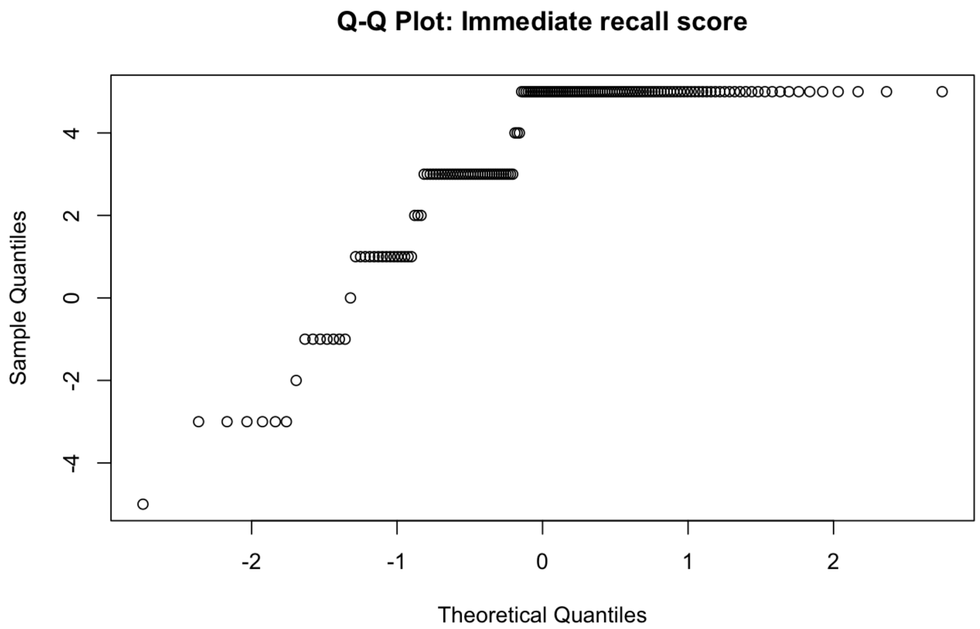


Figure 66: QQ-Plot of immediate recall. The data is not normally distributed

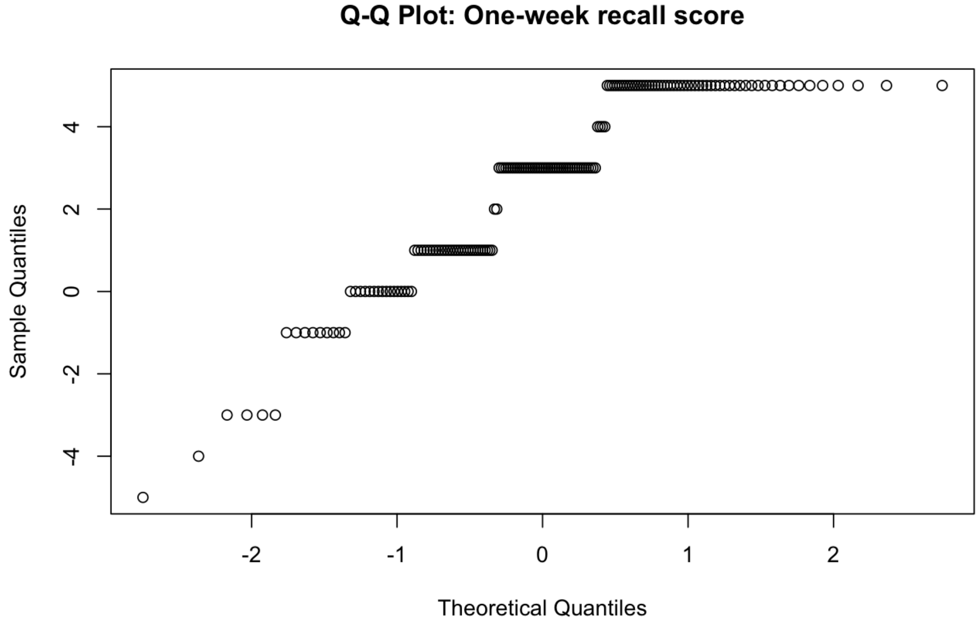


Figure 67: QQ-Plot of one-week recall. The data is not normally distributed

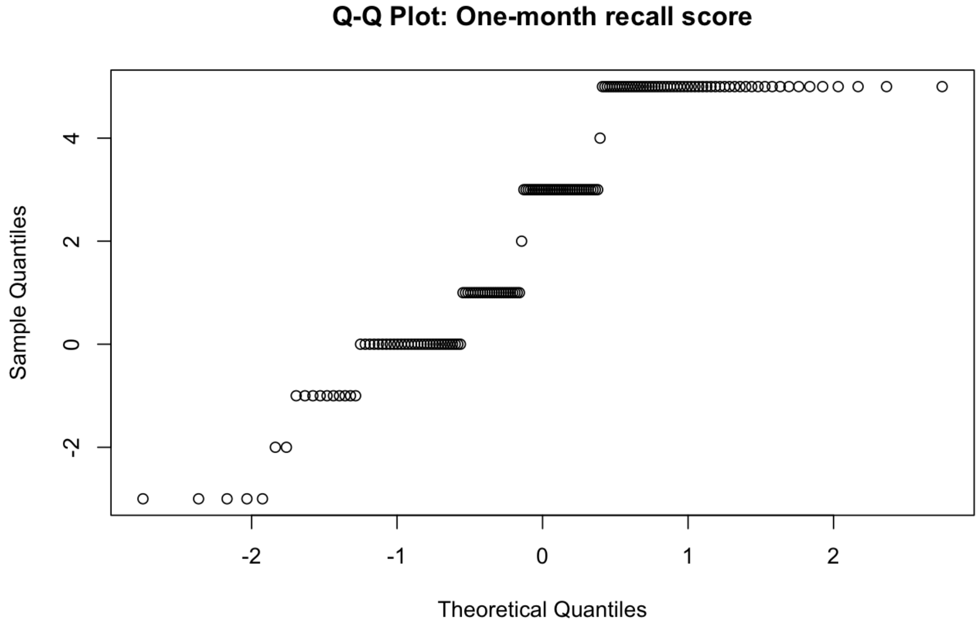


Figure 68: QQ-Plot of one-month recall. The data is not normally distributed

N Literature Approval

Toine Bogers<tobo@itu.dk>

An: Tamara Nagel

Di, 20.05.2025 10:18

Hi Tamara,

I hereby approve your literature list. Please save this reply as a PDF and include it as the last appendix in your thesis :)

Best regards,
Toine

———
Department of Computer Science, IT University of Copenhagen
Rued Langgaards Vej 7 | 2300 Copenhagen | Denmark
Phone: +45 31 44 42 84 | Room 3F11
URL: http://www.toinebogers.com | Skype: toine.bogers

IT UNIVERSITY OF COPENHAGEN

Figure 69: Literature Approval