MASARYKOVA
UNIVERZITA

FILOZOFICKÁ FAKULTA

Comprehensibility of Data Visualizations for the General Public

Identification of Procedures and Recommendations for Creating and Presenting Comprehensible Visualizations

Magisterská diplomová práce

BcA. Kateřina Mahdalová

Vedoucí práce: doc. RNDr. Michal Černý, Ph.D.
Konzultant: Mgr. Tomáš Marek, Ph.D.

Katedra informačních studií a knihovnictví
Program Informační studia a knihovnictví

Brno 2025

MUNI
ARTS

Bibliografický záznam

Autorka:	Kateřina Mahdalová Filozofická fakulta Masarykova univerzita Katedra informačních studií a knihovnictví
Název práce:	Srozumitelnost datových vizualizací pro nejširší veřejnost: Identifikace postupů a doporučení pro tvorbu a prezentaci srozumitelných vizualizací
Studijní program:	Informační studia a knihovnictví, magisterský studijní program
Vedoucí práce:	doc. RNDr. Michal Černý, Ph.D.
Konzultant:	Mgr. Tomáš Marek, Ph.D.
Rok:	2025
Počet stran:	128
Klíčová slova:	data journalism, data visualization, reader interaction, attention, user behavior, online news, visualization exposure

Bibliographic record

Author: Kateřina Mahdalová

Faculty of Arts

Masaryk University

Department of Information and Library Studies

Title of Thesis: Comprehensibility of Data Visualizations for the General Public: Identification of Procedures and Recommendations for Creating and Presenting Comprehensible Visualizations

Degree Programme: Information and Library Studies

Supervisor: doc. RNDr. Michal Černý, Ph.D.

Consultant: Mgr. Tomáš Marek, Ph.D.

Year: 2025

Number of Pages: 128

Keywords: data journalism, data visualization, reader interaction, attention, user behavior, online news, visualization exposure

Anotace

Diplomová práce je zaměřena na výzkum interakcí čtenářů s datovými vizualizacemi v online zpravodajství. Empirické zkoumání čtenářského chování je propojeno s teoretickými východisky z oblasti vizualizačního designu, kognitivní psychologie a mediálních studií. Kvantitativní analýza délky zobrazení vizualizací v reálném publikačním prostředí nabízí nový vhled do toho, jaké vizualizace čtenáři skutečně registrují a jak s nimi interagují.

Abstract

The diploma thesis focuses on the research of reader interactions with data visualizations in online news. The empirical analysis of reader behavior is integrated with theoretical frameworks from the fields of visualization design, cognitive psychology, and media studies. A quantitative analysis of the display duration of visualizations in a real publishing environment offers new insight into which visualizations readers actually register and how they interact with them.

Declaration

I hereby declare that the thesis titled Comprehensibility of Data Visualizations for the General Public: Identification of Procedures and Recommendations for Creating and Presenting Comprehensible Visualizations that I have submitted for assessment is entirely my original work, and that no part of it has been taken from the work of others unless explicitly cited and acknowledged within the text of my thesis.

I certify that I have used AI tools in accordance with the principles of academic integrity and that I have appropriately referenced the use of these tools in my work.

....

Brno, April 30, 2025

BcA. Kateřina Mahdalová

Acknowledgments

I would like to thank my family for their patience and support throughout my studies and while writing this thesis. I often think of my mother, who sadly passed away in the spring of 2024, and who would have been proud to see me complete my studies and pursue topics that I find both important and engaging.

From the bottom of my heart, I thank my wonderful partner for his calmness, patience, and unwavering belief that I would succeed. I'm also grateful to my children who, in their own unique ways, supported me throughout this journey – offering encouragement, understanding, and occasional welcome distractions when I needed them most :-)

I am also grateful to my colleagues who stood by me during a professionally challenging period in early autumn 2024. Their support meant more to me than words can express. That time also reminded me of the importance of adhering to professional principles – even when it is difficult. Fairness and integrity are core values for me in journalism, and I deeply respect everyone who shares this perspective.

Special thanks to my colleague Mgr. Tereza Karlová for walking this academic path alongside me – for the consultations, shared frustrations and joys, and simply for ensuring I didn't have to go through it all alone. Her support made this journey considerably easier.

Finally, I wish to thank everyone who supported, inspired, or helped me along the way – especially my thesis supervisor, doc. RNDr. Michal Černý, Ph.D., and my consultant, Mgr. Tomáš Marek, Ph.D. I am also grateful to be part of the KISK community, where I found inspiration, perspective, and a place that truly nurtures learning.

Table of Contents

1. Introduction

Data journalism has established itself as a significant phenomenon in contemporary digital news reporting, combining analytical approaches, visualization competencies, and storytelling techniques to interpret and make accessible (not only) complex statistical data and their interrelationships (Heravi & Lorenz, 2020; Fink & Anderson, 2015). In the context of growing information overload (Bawden & Robinson, 2009), data visualizations represent an essential tool for media communication, enabling the transformation of abstract numerical values into visual representations that can be cognitively processed and interpreted more efficiently (Ware, 2004). Typical data journalism products integrate textual and graphical elements to help readers understand causal, correlational, or temporal relationships in data (Cairo, 2016; Weber & Rall, 2012).

Despite significant advances in visualization techniques and interactive possibilities over the last decade (Kostkova, 2021; Boy et al., 2014), empirical evidence about how visualizations are perceived and used in real media environments remains notably limited (Bach et al., 2018). Existing research primarily focuses on isolated evaluation of visualizations in controlled laboratory conditions (Heer & Bostock, 2010; Franconeri et al., 2021), which does not adequately assess their actual effectiveness in the natural context of news reading. This methodological shortcoming is particularly evident in the Czech media environment; there has been only a limited number of empirical studies quantifying basic interaction parameters, such as exposure to visualizations, temporal aspects of interaction, or whether readers notice all visualizations in an article. Based on available literature and a review of Czech academic journals in media studies, it can be stated that systematic research in this specific area represents a research gap that this work attempts to partially fill.

This thesis addresses the identified research gap through empirical investigation of how readers actually interact with data visualizations in the real environment of online news media. The research is conducted on the specialized portal datovazurnalistika.cz, which features typical formats of Czech data journalism. The main goal is to measure how much time readers spend on individual visualizations, how they progressively scroll through articles, and what proportion of them reach the end of the article. Special attention is paid to whether readers actually see all visualizations published within a single article. The main contribution of this study is the observation of user behavior in a natural context, rather than in laboratory conditions.

Theoretically, the research is based on three interconnected concepts that directly relate to how readers approach visualizations in online news.

The first is the theory of information overload (Bawden & Robinson, 2009; Eppler & Mengis, 2004), which explains why readers in digital environments often overlook or only superficially skim content. This theory is crucial for our research because it explains why some visualizations may remain completely unnoticed - in a flood of information, readers selectively pay attention to only certain elements and skip many others.

The second concept is readers' visualization literacy (Boy et al., 2014; Kennedy et al., 2016), meaning their ability to understand and work with graphically presented information. This concept is relevant because it helps explain why certain types of visualizations may attract more attention than others - readers naturally focus on types of graphs they better understand and can interpret more quickly.

The third related concept is news efficacy (Park et al., 2019), which helps explain why some readers choose a strategy of completely avoiding certain types of content when facing information overload. For our research, this means that readers with low confidence in their ability to understand data visualizations may deliberately skip these elements, which affects the measured patterns of scrolling and time spent on different parts of the article.

From a methodological perspective, the work uses digital analytics tools to precisely measure several key indicators: time spent on individual visualizations, scrolling patterns, and progression through the article, i.e., how far users go when reading or interacting with visualizations. While existing research often focuses on laboratory studies of how people interpret visualizations (e.g., Hullman & Gelman, 2021), this study concentrates on measurable parameters of actual user behavior in the natural environment of readers. It is based on the assumption that for readers to understand a visualization, they must first notice and pay attention to it - without this basic interaction, it is impossible to talk about any influence of the visualization.

The work is naturally divided into four main parts. The theoretical section deals with cognitive aspects of data visualizations in journalism, including the concept of distributed cognition as a framework for understanding the interaction between reader and visual content, as well as the characteristics of the contemporary digital media environment. The methodological section describes the research design, the method of measuring the observed variables, and the analytical procedures used. The empirical section presents the results of user behavior analysis and examines the patterns of interaction with visual elements in journalistic content. Finally, the discussion section interprets the findings in relation to previous research, reflects on their implications for data journalism practice, and outlines the main limitations of the study along with suggestions for future research directions.

The research was designed with consideration for ethical aspects of working with user behavior data. All data was collected anonymously, without the possibility of identifying specific users. Measurements were conducted only at the level of technical interactions (scrolling, visibility of elements on the screen), not at the level of personal data. This ensured the protection of readers' privacy while maintaining research integrity and obtaining relevant insights about interaction with visualizations. The ethical aspects of the research are discussed in more detail in the methodological section of the thesis.

The insights gained contribute not only to the academic debate on the effectiveness of visual representations in the media but also to the practical improvement of editorial procedures and visualization strategies. Subsequent research focusing on identifying specific patterns of user behavior and factors influencing reader attention can help create empirically based recommendations for visualization design, their placement within articles, and the overall structure of content combining text and visual elements.

At a time when data journalism is gaining increasing importance as a means of communicating complex information, understanding how audiences actually interact with visualizations in a natural environment is a crucial step toward more effective use of this communication tool.

It is important to clarify that while this thesis is titled "Comprehensibility of Data Visualizations," it specifically examines the prerequisites for comprehension—namely, exposure and visibility—rather than comprehension itself. This methodological choice reflects our pragmatic-analytical framework, which prioritizes measuring observable behavioral traces in naturalistic settings. We consider exposure a necessary (though not sufficient) condition for comprehension to occur. By focusing on whether and how long visualizations are actually seen by readers, we establish an empirical foundation for understanding the initial conditions under which comprehension becomes possible.

2. Theoretical Framework

2.1 Development and Context of Visualizations in Data Journalism

Data visualization represents a key tool for extending human cognition, going beyond mere aesthetic representation of numbers. It is a communication medium that transforms abstract values into visual elements that the human brain can process and interpret more efficiently (Ware, 2004). Although historical roots reach back to pioneers of the 18th and 19th centuries, contemporary data visualization primarily serves as a means of reducing cognitive load when processing complex information and supporting decision-making (Franconeri et al., 2021).

In professional terminology, we distinguish several key concepts. Data visualization refers to the precise graphical representation of quantitative or qualitative data with an emphasis on effective communication of relationships and patterns. Infographics represent a broader format combining graphical elements with explanatory text and illustrations, often emphasizing narrative. Interactive visualization enables user input, filtering, and exploration of data for personalized information discovery (Cairo, 2016). This terminological differentiation is not an academic nuance but reflects different communication intentions and approaches to working with data.

With the development of digital technologies, there has been a fundamental shift in the visualization paradigm – from static illustrations to tools for active argumentation and decision support. This transition is not just technological but also conceptual: visualizations no longer serve only to display known facts but to reveal hidden patterns, communicate uncertainty, and mediate narrative (Hullman & Gelman, 2021). The principle of cognitive fit – the alignment between visualization format and cognitive task – thus becomes a key factor in designing effective visualizations (Vessey & Galletta, 1991).

An important aspect of visualizations, which is often overlooked, is their function as tools for building mental models (visualization as model-building). Visualizations are not just passive "reflections of data" but actively shape hypotheses and enable simulation of various scenarios (Card et al., 1999). This concept is particularly relevant for understanding the difference between exploratory and explanatory functions of visualizations. While exploratory visualizations serve to discover previously unknown relationships in data (e.g., in scientific research), explanatory visualizations communicate already discovered relationships to a specific audience (e.g., in media). Each of these functions requires different approaches to visualization design, further underlining the complexity of this communication tool.

Contemporary data journalism reflects these theoretical insights in editorial practice. Leading global media such as The New York Times, The Washington Post, or The Economist have built specialized data teams that combine principles of journalism, data analysis, and visual design. Emerging visualization genres such as scrollytelling, explainer graphics, or interactive maps represent hybrid formats responding to changing patterns of how people access and engage with information (Segel & Heer, 2010). We observe similar developments in the Czech environment, where data-focused newsrooms such as iRozhlas, Deník N, or formerly Seznam Zprávy (until 2024) integrate visualizations as constitutive elements of news reporting.

In an era of information overload and the rise of generative artificial intelligence, data visualization takes on new urgency. During the COVID-19 pandemic, data visualizations became a critical tool for crisis communication, mediating complex epidemiological data to the general public. This phenomenon revealed both the strengths and risks of visualizations: well-designed visualization can make complex data comprehensible, but inappropriately chosen formats can lead to misinterpretations or distrust (Cairo, 2019). The ongoing integration of artificial intelligence tools into visualization creation brings new questions about automation, interpretation, and validity of visualized data.

It is precisely in this context – when visualization serves as a key tool for mediating complex information but simultaneously faces challenges in the form of limited attention and varying levels of visualization literacy among users – that this thesis emerges. It focuses on empirical examination of the interaction between reader and visualization in a natural online environment, aiming to identify factors that determine whether a visualization will be noticed, correctly interpreted, and effectively utilized.

2.2 Cognitive Ecology and Attention Allocation in Digital News Environments

The digital news environment presents readers with a unique challenge: navigating through dense information with limited cognitive capacity. This section establishes a theoretical framework for understanding how readers engage with data visualizations in this context. We adopt a pragmatic-analytical perspective, which privileges observable behavioral traces over inferred mental states, aiming to identify patterns of real-world media interaction without recourse to unobservable cognitive processes. This term refers to a methodological position that prioritizes empirical tractability and practical insight over theoretical abstraction. This approach aligns with recent methodological trends in media studies that prioritize ecological validity over laboratory precision (Laban et al., 2017; Grinberg, 2018).

Before proceeding, we must clarify our central concept of "exposure". In this study, exposure refers specifically to the measurable presence of a visualization in the user's viewport (visible area of the screen) for a quantifiable duration. This operationalization does not imply that users cognitively registered the content – only that the technical precondition for such registration was met. This distinction is critical for interpreting our findings within appropriate epistemological boundaries.

2.2.1 Information Saturation and Selective Attention

Digital media environments flood users with content without clear relevance hierarchies. This condition, termed "information overload" (Toffler, 1970; Castells, 2010), creates a context where visualizations must compete for readers' limited attention. Although Levy (2008) rightly questions whether an "optimal" information level exists – a criticism we acknowledge – the concept remains valuable for understanding why readers prioritize certain content elements over others.

Recent empirical research has moved beyond abstract capacity models to document actual reading behaviors. Grinberg (2018), analyzing over 7.7 million page views, identified six distinct reading patterns ranging from brief scanning to extensive reading. These patterns show how readers prioritize content under cognitive pressure, making rapid decisions about what deserves attention. This tendency to prioritize is reinforced by editorial strategies themselves. The NEWSROOM dataset (Grusky et al., 2018), comprising 1.3 million articles and their human-written summaries, shows how newsrooms adopt diverse summarization strategies – ranging from purely extractive to highly abstractive – to tailor information presentation to audience needs and platform constraints. These summaries act as framing tools that shape what readers attend to and how deeply they engage. Building on Grinberg's work, our research classifies readers into typological groups based on their observed interaction patterns (e.g., non-scrollers, partial readers, complete readers). These typologies form the basis for Hypothesis 5 (H5), which explores whether different reader types exhibit distinct patterns of visualization exposure.

This selective attention directly relates to our research on visualization exposure. Laban et al. (2017) demonstrated that these attention patterns follow predictable trajectories, with a strong position effect – content placed higher in articles generally receives more attention than lower-placed content. Their viewport-based methodology, similar to our approach, captured real-world reading behaviors without the artificiality of laboratory settings.

The position effect has been consistently observed in digital reading studies (Holmqvist & Wartenberg, 2005), though it varies with factors like visual salience and topic interest. Rather than representing a fixed "importance heuristic," it likely reflects both reading conventions and practical filtering strategies in time-limited scenarios. This position effect directly informs our first hypothesis (H1) about the systematic decrease in visualization exposure with increasing article depth.

2.2.2 Attentional Dynamics During News Reading

Reading digital news rarely follows a linear path. Readers frequently shift between content elements and process multiple information streams simultaneously. Understanding these dynamics helps explain patterns of visualization engagement.

When readers switch between tasks or content elements, cognitive processing from previous elements may interfere with processing new information – a phenomenon Leroy (2009) calls "attention residue." For visualizations, this means that reader engagement may be affected by preceding content elements.

Dabbish et al. (2011) found that digital media users regularly practice media multitasking, rapidly alternating between different content streams. This multitasking context means visualizations typically receive fragmented rather than sustained attention – a challenge particularly relevant for complex visual formats.

Device type significantly shapes these attention dynamics. Mobile devices, with their smaller viewports and touch-based interactions, create different constraints and affordances compared to desktop environments. Lee et al. (2021) found that mobile reading typically involves more frequent but shorter attention bursts, with greater viewport limitations affecting how much content appears simultaneously on screen. These device-specific differences directly inform our hypothesis H3, which predicts different exposure patterns between mobile and desktop users.

While both studies employ viewport-based metrics, Laban et al. (2017) focus on modeling in-page attention trajectories, whereas Lagun and Lalmas (2016) develop user engagement typologies predictive of future behavior. Lagun and Lalmas showed that traditional engagement measures like time-on-page fail to account for how attention distributes across page elements. Their work supports our methodological choice to track not just if visualizations appear on screen but for how long – providing a more nuanced view of exposure patterns and informing our hypothesis H2 regarding the relationship between visualization complexity and exposure duration.

2.2.3 Contextual Factors and Critical Perspectives

While our research focuses primarily on behavioral aspects of visualization exposure, we recognize that these patterns exist within broader contexts that influence what readers attend to and why. Critical perspectives offer important correctives to purely behavioral approaches.

While Gitelman (2013) does not use the term directly, her critique implies what we might call a "transparency illusion“ – visualizations appearing to offer direct access to reality while obscuring subjective choices in data collection and representation. This illusion can lead readers to accept visualized data without critical engagement. For example, COVID-19 dashboards were often presented as objective pandemic representations, potentially obscuring methodological limitations in their creation.

Kennedy et al. (2016) further emphasize that visualization literacy includes not just cognitive skills but also critical awareness of how visualizations can reflect power structures and cultural assumptions. This perspective suggests that interaction patterns might be influenced by readers' trust or ideological stance toward presented data – factors that extend beyond our behavioral measurements but provide important interpretive context.

Boy et al. (2014) experimentally demonstrated that up to 65% of users completely ignore interactive elements or use them very superficially. Bach et al. (2018) documented "visualization blindness" – readers completely overlooking visuals even in prominent positions. These findings challenge assumptions about visualizations' inherent effectiveness.

While these factors are not directly incorporated into our testable hypotheses, they provide a necessary interpretive lens through which our behavioral findings should be understood. They remind us that patterns of visualization exposure reflect not just cognitive constraints but also social and cultural factors that shape reading practices.

2.2.4 An Integrated Framework for Research Design

The theoretical perspectives outlined above directly inform our research design and variables. Table 1 shows how these theoretical concepts map our research variables, operational measurements, and specific hypotheses.

Theoretical Concept	**Measured Variable**	**Operationalization**	**Related Hypothesis**
Selective Attention	Position of Visualization	Scroll depth in pixels; relative vertical position in article	H1
Visibility Threshold	Exposure Probability	Binary: visualization appeared in viewport (yes/no)	H1
Cognitive Load & Design Complexity	Visualization Type	Categorized as: simple, interactive, multi-slide	H2
Device Affordances	Device Type	Mobile vs. Desktop	H3
Attention Allocation	Exposure Duration	Time (in seconds) visualization remained in viewport	H2, H3, H4
Reader Behavior Typology	User Engagement Patterns	Cluster analysis based on scroll behavior and exposure time	H5

Table 1: Theoretical framework linking core concepts, measured variables, and hypotheses in the study of visualization exposure

This pragmatic-analytical framework allows us to examine observable proxies of reader attention without making claims about internal cognitive processes. We acknowledge the limitations of this approach – exposure doesn't guarantee comprehension or engagement – but it provides valuable insights into the baseline conditions for visualization effectiveness. Our research primarily examines the behavioral dimensions of reader-visualization interaction, using viewport-based measurements to capture patterns that emerge in naturalistic reading environments. This approach complements laboratory-based studies of visualization comprehension by revealing how visualizations function "in the wild" of everyday news use. The five research hypotheses (H1-H5) that emerge from this framework will be formally presented and operationalized in Chapter 3.

These theoretical considerations lay the foundation for our methodological design, which aims to empirically trace the conditions under which readers are likely to encounter, perceive, and potentially engage with data visualizations in the dynamic environment of digital news. The following chapter details how we operationalize these concepts within our research methodology.

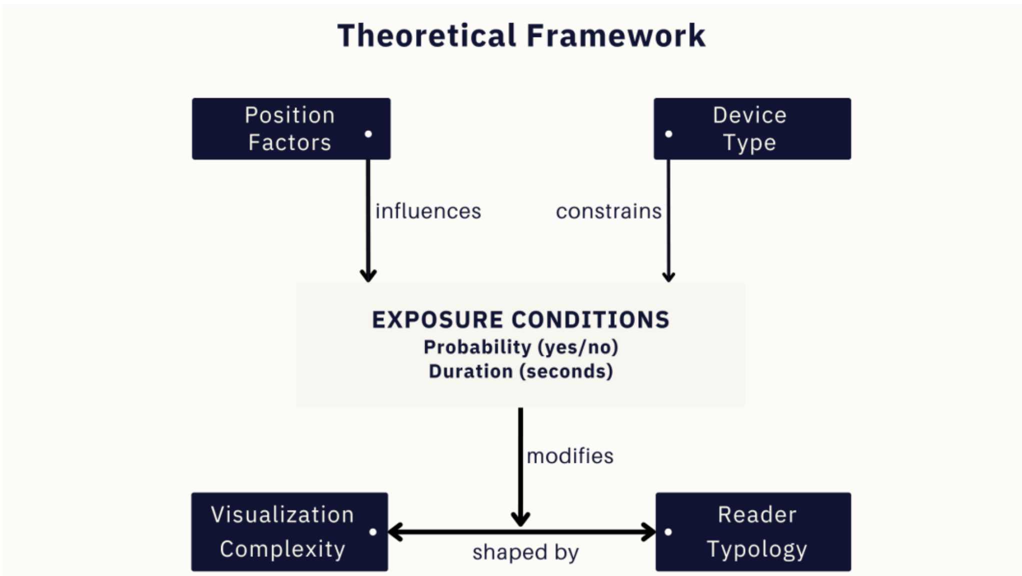


Figure 1: Theoretical Framework for Visualization Exposure

Figure 1 illustrates how structural features (position, device) and content properties (visualization type, complexity) are linked to measurable user behavior. While this framework cannot capture internal cognitive mechanisms such as comprehension or reflection, it enables empirically grounded observations of exposure patterns. As such, it offers a pragmatic model for identifying factors that influence whether visualizations are even noticed – necessary precondition for any further engagement.

2.3 Cognitive Ecology of Visualizations: An Integrative Perspective

Building on our previous discussion of attention dynamics, we now turn to the role of visualizations within this cognitive-ecological system – not just as stimuli, but as cognitive tools shaped by perceptual, interactive, and sociocultural dynamics. The concept of cognitive ecology, as applied in this work, represents an interdisciplinary framework that connects cognitive psychology, an ecological approach to perception, and the theory of distributed cognition. Building primarily on Hutchins' conception of distributed cognition and extended by recent ecological perspectives in visualization research (e.g., Liu et al., 2018), this framework recognizes that visualization processing is not merely an internal mental operation but emerges from the interaction between the reader's cognitive capabilities, the visualization's design features, and the socio-technological context in which the interaction occurs.

Unlike classical cognitive psychology, which primarily focuses on internal mental processes, cognitive ecology emphasizes that cognitive processes are embedded in complex systems including material artifacts, social interactions, and cultural practices (Hutchins, 2010). This approach allows us to examine not only how individuals cognitively process visual information but also how these processes are shaped by the technological, social, and cultural environment – a perspective particularly relevant for understanding visualization engagement in the dynamic context of digital news use.

2.3.1 Perceptual and Cognitive Foundations of Visualization Effectiveness

At the core of visualization effectiveness lies the human visual system's ability to detect patterns, relationships, and anomalies in graphically represented data. This ability forms the foundation for all data visualization work.

Cleveland and McGill's (1984) seminal research on graphical perception established an empirically based framework for understanding which visual encodings are most effective for communicating quantitative information. Their experiments showed that position-based encodings (as used in bar charts) allow for more accurate value extraction than length encodings (as in bar charts with a non-zero baseline), which in turn outperform area or angle encodings (as in pie charts).

The influence of this work on visualization design has been profound, informing principles that prioritize certain visual encodings over others. For our research, this perceptual hierarchy has direct implications for how visualization types might affect exposure patterns. Visualizations using perceptually more efficient encodings (like position) may require less cognitive effort to process, potentially affecting how long readers engage with them before feeling they have extracted the necessary information.

However, this classical perceptual research has methodological limitations that we must acknowledge. Studies like Cleveland and McGill's typically involved controlled experiments with simple stimuli and small samples, limiting their ecological validity. Heer and Bostock (2010) attempted to address these limitations by replicating and extending the experiments with larger, more diverse samples. While they confirmed the general ranking of visual encodings, they found smaller differences between them and significant individual variation in performance – suggesting that perceptual efficiency is not universal but interacts with individual factors and prior experience.

Moreover, classical perceptual studies primarily measured accuracy of individual value extraction, neglecting tasks more common in real-world visualization use, such as trend identification, pattern recognition, or comparison of data groups. Franconeri et al. (2021) highlight that different visualization types may be optimal for different analytical tasks, regardless of their ranking in simple perceptual hierarchies. For instance, while bar charts excel at precise value comparison, line charts better reveal trends over time, even though both use position encoding.

The concept of cognitive fit (Vessey, 1991) offers a complementary perspective that emphasizes alignment between visualization format and cognitive task. According to this theory, visualization effectiveness depends not only on inherent perceptual properties but also on how well the representation matches the specific cognitive operation the user needs to perform. For example, a heat map might be excellent for identifying spatial patterns but poor for precise value comparisons between specific regions – a task better served by a simple table or bar chart. This perspective explicitly connects perceptual complexity with task appropriateness, making it central to our investigation of whether more cognitively demanding visualizations correspond with longer exposure durations (H2).

This concept of cognitive fit guided our selection of visualization formats for the research articles used in this study. For instance, Sankey diagrams were chosen for visualizing voter shifts because they excel at showing flow relationships between categorical data sets. These diagrams allow readers to trace voter movement between parties while maintaining a holistic view of the overall electoral landscape – a task that would be difficult to accomplish with simpler chart types like bar or pie charts. Similarly, race charts were employed for temporal comparisons of polling data, as they effectively communicate changing rankings over time, making them particularly suitable for visualizing political party competition dynamics.

Recent advancements in visualization research have moved beyond static effectiveness rankings to consider the role of interaction, animation, and narrative techniques in supporting cognition. Interactive techniques can offload cognitive operations to the interface, potentially reducing mental effort and enhancing engagement. These affordances function as cognitive offloading mechanisms (Risko & Gilbert, 2016), allowing users to shift demanding operations – such as value comparison or pattern extraction – onto the interface, thus reducing internal cognitive load. However, as we discussed in the previous section, the actual use of interactive features in naturalistic settings is often limited, creating a gap between theoretical potential and practical reality.

These perceptual challenges illustrate the limits of 'objective' visualization – even technically accurate representations can lead to systematically biased interpretations. For our research, this means that patterns of visualization exposure may reflect not only interest in the data or effectiveness of design but also perceptual challenges that influence how readers interpret and engage with different visualization types.

The concept of 'visualization equilibrium' proposed by Kayongo et al. (2022) further extends this discussion by examining how visualizations affect strategic decision-making in group contexts, where users not only interpret visualizations but also anticipate how others might interpret and act upon the same visual information. This socio-cognitive dimension adds another layer of complexity to understanding how visualizations function in public communication contexts such as journalism.

2.3.2 Visualization Literacy as a Sociocultural Phenomenon

Visualization literacy represents a multidimensional concept encompassing the skills needed to interpret and critically evaluate visual representations of data. This literacy extends beyond mere perceptual processing to include cultural, contextual, and critical dimensions that influence how readers engage with visualizations.

Boy et al. (2014) define visualization literacy as "the ability to confidently understand visually represented data, extract information, and draw conclusions from data visualizations." This definition emphasizes the cognitive and practical aspects of interpretation. In contrast, Kennedy et al. (2016) propose a broader conceptualization that includes critical awareness of how visualizations reflect particular worldviews, power structures, and cultural assumptions. This expanded view suggests that visualization literacy is not just a technical skill but a form of critical media literacy essential for meaningful engagement with data in public discourse.

This broader conception is particularly appropriate in a journalistic context, where visualizations are not neutral but rhetorical objects embedded in public discourse. It is important to note that visualization literacy here refers to the receptive ability to read and interpret visualizations – the skills that readers need – rather than the productive capacity to create visualizations, which would be the domain of data journalists and visualization designers. This distinction is crucial for our research, as we are investigating how readers interact with visualizations, not how they are created.

These differing definitions have direct implications for research practice. If we were to operationalize visualization literacy according to Boy et al., we would focus primarily on measuring the accuracy of information extraction. Conversely, Kennedy et al.'s definition would require examining whether and how users critically reflect on the presented data, including questioning sources, methodologies, or framing choices. In this work, we align more with Kennedy et al.'s broader conception, which better reflects the complex reality of interaction with visualizations in a journalistic context. Empirical research shows that components of visualization literacy vary significantly across individuals and contexts.

Boy et al. (2014) found that visualization literacy levels differ not only between individuals but also across cultural contexts. These differences cannot be reduced to a simple binary division of "literate" versus "illiterate" readers but represent a continuum of competencies that vary across visualization types and contexts of use.

This sociocultural perspective on visualization literacy has implications for our research on exposure patterns. Readers with different literacy levels may adopt different strategies when encountering visualizations – some may engage deeply, while others might skip complex visualizations entirely. These varying engagement patterns inform our hypothesis H5 about reader typology, as we expect to identify distinct groups of readers based on their interaction patterns with visualizations.

Current approaches to fostering visualization literacy include explicit onboarding techniques that guide readers through unfamiliar visualization types. Stoiber et al. (2022) evaluated different onboarding methods, finding that techniques like scrollytelling tutorials and contextual guidance can significantly improve readers' ability to interpret complex visualizations. These findings suggest that visualization design should not assume universal literacy but should incorporate supportive elements that help readers develop the necessary skills for interpretation.

2.3.3 Perceptual Challenges and Cognitive Constraints

Consistent with our pragmatic-analytical perspective, we interpret perceptual constraints not as flaws in user performance, but as integral elements of the cognitive system that visualization design must accommodate. Despite advances in visualization design, several perceptual and cognitive challenges remain that can affect how readers engage with visualizations. These challenges represent not just limitations of individual cognition but structural properties of human perception that visualization designers must account for.

Color perception presents particular challenges for visualization design. While color is visually salient and often intuitive, its interpretation can be highly subjective and context-dependent. The phenomenon of simultaneous contrast – where perception of a color is influenced by surrounding colors – can significantly affect how color-encoded data is interpreted (Albers, 2006). This effect means that the same color values may be perceived differently depending on their visual context, potentially distorting data interpretation.

Beyond color, other perceptual effects can influence visualization interpretation. When visualizing continuous data, readers often mentally categorize values around reference points – a phenomenon Xiong et al. (2020) documented in their studies of graphical perception.

They found that values near 25%, 50%, or 75% in pie charts or stacked bar charts are often remembered as exactly matching these reference points, even when they differ significantly. This perceptual "snapping" to reference points represents not just an individual error but a systematic property of human perception. These perceptual challenges illustrate the limits of "objective" visualization – even technically accurate representations can lead to systematically biased interpretations. For our research, this means that patterns of visualization exposure may reflect not only interest in the data or effectiveness of design but also perceptual challenges that influence how readers interpret and engage with different visualization types.

In our researched articles, we addressed these perceptual challenges in several ways. For example, when using Sankey diagrams for voter shift analysis, we implemented a color-coding system that maintained perceptual consistency across related data elements, added interactive hover states to clarify precise values, and included contextual annotations to guide interpretation. For multi-slide visualizations presenting complex survey data, we broke down information into progressive disclosure sequences rather than presenting all data at once, reducing cognitive load while maintaining narrative coherence. These design choices aimed to mitigate the potential cognitive barriers associated with more complex visualization formats.

From an applied perspective, research on scrollytelling and narrative visualization offers promising approaches to addressing these challenges. Oesch et al. (2022) identified five standard techniques in scrollytelling: graphic sequences, animated transitions, panning and zooming, scrolling through movies, and showing/autoplaying animated content. These techniques can guide readers through complex data, potentially mitigating perceptual challenges by providing structured narrative paths through the information.

Similarly, Mörth et al. (2022) demonstrated how scrollytelling can support scientific communication by allowing readers to explore dynamic narratives at their own pace. Their approach combines the familiar gesture of scrolling with progressive disclosure of information, providing readers with a sense of control while maintaining narrative coherence. These narrative techniques may be particularly valuable for addressing perceptual challenges and supporting readers with varying levels of visualization literacy.

From a cognitive-ecological perspective, these perceptual challenges and constraints should not be viewed simply as limitations to overcome but as features of the cognitive system that visualization design must work with rather than against. Effective visualizations align with human perceptual tendencies rather than fighting against them – a principle that guides our analysis of exposure patterns in the empirical portion of this study.

The following table synthesizes core design implications distilled from the cognitive-ecological approach discussed above.

Principle	**Implication for Visualization Design**
Perceptual salience is context-dependent	Use redundancy and clear labeling
Cognitive fit affects exposure time	Align format with task type
Visualization literacy varies	Include onboarding or contextual cues
Readers multitask	Favor scrollytelling over static blocks
Position influences attention	Place critical information higher in layout
Device constraints affect viewport	Design responsive, adaptive visualizations

Table 2: Design Considerations from a Cognitive-Ecological Perspective

This table summarizes key principles derived from our cognitive-ecological framework and their practical implications for visualization design in journalistic contexts. These considerations not only inform our research design but also provide a foundation for interpreting our empirical findings and developing evidence-based recommendations for practice. Together, these cognitive, perceptual, and sociocultural factors provide the conceptual foundation for our investigation of exposure patterns (H2) and reader typologies (H5), as developed in the following empirical chapters.

2.4. Interaction with Visualizations in the Digital Journalistic Environment

Having established the cognitive-ecological framework for understanding visualization processing, we now focus specifically on how readers interact with visualizations in the context of digital journalism. This section bridges theoretical principles with the practical realities of online news use, examining how the digital environment shapes visualization engagement. This understanding is crucial for contextualizing our empirical findings on visualization exposure and for developing effective recommendations for journalistic practice.

2.4.1 Patterns of User Engagement in Online News Contexts

Digital news use differs fundamentally from traditional print reading in its non-linear, fragmented, and selective nature. Online readers navigate content actively, making moment-by-moment decisions about what deserves attention and what can be skipped. These behavioral patterns create distinct challenges and opportunities for visualization designers seeking to communicate effectively in digital journalistic contexts.

Eye-tracking studies have provided valuable insights into these patterns. Early research by Pernice (2017) identified the now-famous "F-pattern" of web reading – users scan horizontally across the top, then vertically down the left side, followed by another, shorter horizontal scan. This pattern suggested that content positioned in the upper-left quadrant would receive disproportionate attention, while lower-positioned elements might be overlooked entirely.

However, more recent research has revealed greater complexity in digital reading behaviors. Laban et al. (2017), analyzing 1.2 million news reading sessions through viewport data, found that attention patterns vary significantly based on content type, user motivation, and device characteristics. Rather than following a universal F-pattern, readers adopt different scanning strategies depending on their specific goals and the structural cues provided by the page design. This finding aligns with our approach to reader typology (H5), suggesting that different reading patterns may be associated with distinct user segments and visualization engagement strategies.

These viewport-based studies offer particularly relevant methodological precedents for our research. While eye-tracking reveals precise gaze locations, viewport measurement captures which content elements enter the user's visible screen area – providing ecologically valid data at scale. Lagun and Lalmas (2016) further refined this approach by developing engagement metrics based on scrolling patterns and viewport time, demonstrating how these measurements can predict user satisfaction and return visits. Their distinction between shallow, moderate, and deep engagement parallels our interest in identifying different reader types through behavioral observation.

Grinberg (2018) extended this work by identifying six distinct reading patterns across 7.7 million page views, ranging from brief scanning to deep, methodical reading. These patterns were not randomly distributed but correlated with content characteristics and user demographics, suggesting systematic differences in how readers approach online content. For visualization exposure, this implies that certain formats may better serve different reading styles – a consideration that informs our analysis of visualization complexity (H2) and its relationship to exposure duration.

In the context of data visualizations specifically, readers often employ a pattern that Hegarty et al. (2011) termed "just-in-time processing" – selectively engaging with visual elements as needed to answer specific questions or verify claims, rather than systematically processing the entire visualization. This selective attention strategy represents an adaptation to information-rich environments where comprehensive processing of all visual elements would be cognitively overwhelming. When designing our research articles, we recognized this behavior by incorporating progressive disclosure in multi-slide visualizations and by ensuring that even partial engagement could yield meaningful insights. "Progressive disclosure" refers to the design principle of presenting information gradually, revealing additional details only when needed or requested, rather than overwhelming users with all information at once.

A particularly relevant phenomenon for our research is what Bach et al. (2018) termed "visualization blindness" – cases where users completely overlook charts or graphs even when they occupy substantial screen space. This effect seems most pronounced for visualizations that:

- Appear below the initial viewport ("below the fold")
- Lack clear visual differentiation from surrounding content
- Fail to signal their interactive capabilities
- Do not visually communicate their informational value

This phenomenon directly relates to our hypothesis about the influence of position on visualization exposure (H1) and informs our approach to measuring when and how visualizations enter the user's viewport.

2.4.2 The Impact of Device Type on Visualization Interaction

A crucial factor influencing visualization interaction is the type of device used for accessing news content. According to data from the Reuters Institute Digital News Report (Newman et al., 2023), mobile access now accounts for 60–80% of total traffic to news websites. This shift toward mobile use has profound implications for how visualizations are designed, displayed, and interacted with.

Mobile devices impose several constraints on visualization interaction. The most obvious is reduced screen size, which limits the amount of information that can be displayed simultaneously and often necessitates simplified or segmented presentations. Touch-based interaction replaces mouse hovering, eliminating common interactive techniques that rely on pointer movement without clicking. The "fat finger problem" makes precise selection of small elements difficult, while one-handed usage creates additional ergonomic limitations that can affect scrolling patterns and interaction choices.

Empirical studies examining these differences have yielded mixed results. Lee et al. (2021) found that mobile users spent 42% less time with visualizations than desktop users, controlling for total article engagement. Their interactions were also less exploratory and more goal-directed, suggesting qualitatively different engagement strategies. However, Kim and Heer (2018) documented cases where mobile users engaged more deeply with certain visualization types, particularly those optimized for touch interaction. These contradictory findings suggest that the relationship between device type and visualization engagement is complex and likely moderated by design choices and content characteristics.

Beyond simple duration metrics, mobile and desktop users exhibit different interaction patterns. Mobile reading typically involves more frequent but shorter attention bursts, with greater reliance on semantic markers and structural cues to guide navigation. The predominantly vertical orientation of mobile screens also influences how content is parsed, potentially enhancing the position effect whereby elements placed higher receive disproportionate attention.

These device-specific characteristics directly inform our hypothesis H3, which predicts differences in visualization exposure between mobile and desktop users. Our methodological approach, which distinguishes between device types while measuring the same exposure variables, allows us to empirically examine whether these differences manifest in natural reading environments.

When designing visualizations for our research articles, we employed responsive approaches that adapted to different screen sizes. For instance, complex Sankey diagrams for voter flows were redesigned for mobile viewing with simplified layouts and larger touch targets, while maintaining the same core information. Similarly, multi-slide visualizations used touch-friendly navigation controls on mobile devices while offering different interaction affordances on desktop. These adaptations recognize that effective visualization design must consider not just perceptual principles but also the specific interaction constraints of different devices.

2.4.3 Interactivity: Promise and Reality

Interactive elements represent one of the most distinctive features of digital visualizations compared to their static predecessors. In principle, interactivity offers multiple advantages: it can facilitate exploration of complex datasets, personalize information to individual interests, reveal multiple layers of detail, and engage readers more actively in the meaning-making process. The theoretical promise of interactivity has led to its widespread adoption in data journalism, with interactive charts, maps, and dashboards becoming standard features in many digital news publications.

This enthusiasm for interactivity is grounded in influential frameworks such as Shneiderman's (1996) "overview first, zoom and filter, then details-on-demand" mantra, which posits that effective information seeking follows a progressive disclosure pattern. Yi et al. (2007) further developed this approach by taxonomizing interactive techniques into categories including selection, exploration, reconfiguration, encoding, abstraction/elaboration, filtering, and connection. These frameworks suggest that well-designed interactivity can enhance both comprehension and engagement with visualized information.

However, empirical studies of actual user behavior reveal a substantial gap between the theoretical potential of interactivity and its practical utilization. Boy et al. (2014), based on a combination of behavioral observation and self-reported data, found that only 35% of users interacted with the visualizations more than once, while the remaining 65% either did not engage at all or interacted only superficially (just once). Similarly, Ziemkiewicz and Kosara (2010) reported that 68% of users did not interact at all, and another 22% made fewer than three interactions. These findings suggest what might be termed an interaction gap – a consistent disparity between designed affordances and actual user behavior.

This discrepancy is not limited to interactivity alone. Bach et al. (2018) introduced the concept of visualization blindness – a phenomenon in which users overlook visualizations entirely, even when they are placed prominently within a digital article. This occurs particularly when visualizations are not visually differentiated from surrounding text, or when their communicative value is not immediately apparent.

Together, these findings challenge assumptions about the inherent effectiveness of visual and interactive elements. Readers may fail to notice, use, or value interactive features, especially in fast-paced reading environments. As discussed earlier (2.3.1), this undermines the cognitive offloading potential of interactivity, which – when successfully used – can help reduce mental effort by delegating tasks like filtering, sorting, or comparing data to the interface (Risko & Gilbert, 2016). When interactions fail to occur, these cognitive benefits remain unrealized.

Several factors contribute to this interaction gap. First, many readers approach digital news with passive content expectations, focusing on efficient information extraction rather than active exploration. Second, cognitive load and effort involved in discovering or learning to use novel interaction models may deter users. Third, poor signaling of interactive affordances—such as the absence of animations, buttons, or hover states—can render interactive features effectively invisible, particularly for users scanning quickly.

These challenges have led to the development of what some researchers term parsimonious interactivity – a design strategy that prioritizes a limited set of well-signaled, high-utility features. Instead of offering a broad set of complex tools, this approach ensures that the visualization communicates its core message in its default state, with interaction serving only as a layer of optional enhancement.

In our research articles, we applied this principle deliberately. Sankey diagrams illustrating voter shifts were designed to show the main trends at a glance, while hover interactions revealed precise numeric values for more engaged users. Similarly, multi-slide visualizations offered progressive disclosure, enabling readers to explore data at their own pace but without requiring interaction to understand the main message.

These observations directly inform our Hypothesis H2, which predicts that more complex visualizations – such as multi-slide formats – will correlate with longer exposure duration, under the assumption that readers invest more time when processing more demanding content. Likewise, Hypothesis H3 posits that device type may moderate interaction patterns, with mobile users expected to engage differently than desktop users due to varying affordances.

In summary, interactivity holds significant potential for enhancing engagement and comprehension—but only under specific conditions. Its actual use depends not only on interface design, but also on reader expectations, motivation, and situational constraints. Understanding these dynamics is essential for interpreting patterns of visualization exposure and for designing interactive visualizations that are both usable and impactful.

2.4.4 Methodological Approaches to Measuring Visualization Engagement

Studying how readers interact with visualizations in natural settings presents significant methodological challenges. Unlike laboratory studies, which can employ techniques like eye-tracking, think-aloud protocols, or comprehension tests, research on visualization engagement "in the wild" must balance ecological validity with measurement precision.

Several methodological approaches have been developed to address this challenge:

Laboratory studies with eye-tracking provide highly detailed data on visual attention, fixation patterns, and gaze sequences. While offering precise measurements, these studies typically involve small samples in artificial reading contexts, limiting their generalizability to natural behaviors. Eye-tracking can reveal which elements of visualizations receive visual attention but cannot directly measure comprehension or engagement.

Self-reported measures through surveys or interviews offer insights into subjective experiences and interpretations but suffer from recall biases and social desirability effects. Readers often cannot accurately report their own interaction patterns, particularly for behaviors that occur automatically or without conscious awareness.

Server-side analytics capture aggregate user behaviors across large samples but typically lack granularity below the page level. Standard metrics like time-on-page or bounce rate provide general engagement indicators but cannot isolate interactions with specific visualization elements.

Viewport-based measurement, the approach adopted in this study, tracks which elements appear in the user's visible screen area and for how long. This method, employed by Laban et al. (2017) and Lagun and Lalmas (2016), offers several advantages: it works at scale, captures behavior in natural reading contexts, and provides element-level precision without disrupting the reading experience. However, viewport visibility represents only a necessary condition for engagement, not a sufficient one – elements may be technically visible without receiving active attention.

Interaction logging records specific user actions like clicks, hovers, or scrolls. While providing concrete behavioral data, this approach captures only explicit interactions, missing passive engagement or visual processing without interaction. It is also typically limited to measuring interactions within a visualization rather than the relationship between the visualization and surrounding content.

Each of these methods involves tradeoffs between scale, precision, and ecological validity. Our research employs viewport-based measurement as a primary method, supplemented by interaction logging for interactive elements. This combination allows us to capture both passive exposure (visualization appearing in the viewport) and active engagement (deliberate interaction with visualization features) across a naturalistic sample of readers.

This methodological approach aligns with our pragmatic-analytical framework by focusing on observable behavioral indicators rather than attempting to directly measure internal cognitive processes. By tracking when visualizations enter and exit the viewport, we can determine not just if they were technically visible but for how long they remained so – a proxy measure for the opportunity to engage with the content.

The viewport-based approach also addresses several practical challenges of visualization research in journalistic contexts. It works across different device types, allowing comparison between mobile and desktop experiences. It functions without requiring user consent beyond standard site terms, avoiding selection biases that arise when recruiting participants for more invasive studies. And it captures behavior at scale, providing robust sample sizes that can reveal patterns across different reader segments.

Factor	Impact	Implication for Design or Measurement
Device type	Alters interaction style and viewport size	Use responsive design; measure exposure across devices
Scroll position	Significantly affects visibility probability	Place key visuals higher; measure position effect
Interactivity	Often underutilized in naturalistic settings	Make interaction obvious; default state should be informative
Reader typology	Drives engagement depth and reading patterns	Identify different reader segments; analyze exposure by type
Visualization complexity	Influences processing time and effort	Match complexity to task; measure exposure duration
Progressive disclosure	Controls information flow and cognitive load	Structure complex visualizations in digestible sequences

Table 3: Design Factors Influencing Visualization Exposure in Digital News

These findings clarify the behavioral landscape in which visualizations compete for user attention. They also shape our methodological priorities: to capture not only whether users had the opportunity to see a visualization (exposure), but also how this opportunity varies across formats, devices, and reader types. The following chapter details how we operationalize these concepts within our research methodology, including the specific variables measured, the technical implementation, and the analytical strategies employed to interpret the resulting data.

2.5 Research Problem, Objectives, and Theoretical Foundations

Having established the theoretical framework for understanding visualization processing, user behavior, and digital journalism contexts, we now turn to the specific research problem and objectives of this study. This section synthesizes the theoretical perspectives presented in previous chapters and articulates how they inform our research questions, hypotheses, and methodological approach.

2.5.1 Articulating the Research Problem

Data journalism has established itself as a significant approach to contemporary news reporting, combining analytical skills, visualization competencies, and narrative techniques to interpret complex statistical data and make them accessible to general audiences (Heravi & Lorenz, 2020; Fink & Anderson, 2015). At the heart of data journalism lies the visualization – the graphical representation of quantitative information designed to facilitate understanding, reveal patterns, and support evidence-based claims.

Despite significant advances in visualization techniques and interactive possibilities over the past decade (Kostkova, 2021; Boy et al., 2014), empirical evidence about how visualizations are actually perceived and used in real media environments remains notably limited (Bach et al., 2018). While considerable research has been conducted on visualization effectiveness in controlled laboratory settings (Heer & Bostock, 2010; Franconeri et al., 2021), there is a substantial gap in our understanding of how these findings translate to naturalistic reading environments—particularly in online news contexts where visualizations compete with other content elements for limited attention resources.

This gap is particularly evident in the Czech media environment, where there has been only a limited number of empirical studies quantifying basic interaction parameters, such as exposure to visualizations, temporal aspects of interaction, or whether readers notice all visualizations in an article. Based on available literature and a review of Czech academic journals in media studies, we can identify a research gap that this work attempts to partially fill.

The central research problem can be formulated as: How do readers in the natural environment of news websites interact with data visualizations, and what factors determine whether a visualization will be noticed, exposed to reader attention, and potentially engaged with?

This problem is significant for both theoretical and practical reasons. Theoretically, it addresses the ecological validity of visualization research by examining how readers engage with visualizations in authentic contexts rather than laboratory settings. Practically, it seeks to provide evidence-based recommendations for data journalists and visual designers, helping them create more effective visualizations that better serve information needs in natural reading environments.

2.5.2 Research Objectives and Questions

The main objective of this study is to empirically investigate how readers actually interact with data visualizations in the real environment of online news media. Specifically, we aim to measure:

- **Whether readers see visualizations at all (visibility)**
- **How much time they spend with visualizations (exposure duration)**
- **How these metrics vary across different visualization types, positions, and devices**

Rather than focusing on comprehension or interpretation, we deliberately constrain our study to these observable behavioral metrics, which constitute the necessary preconditions for any deeper engagement with visualization content. This methodological choice aligns with our pragmatic-analytical framework outlined in section 2.2.

These objectives translate into the following specific research questions:

- **What influence does the position of a visualization within an article have on the probability that a reader will interact with it?**
    - Is the assumption confirmed that visualizations placed in the upper third of an article receive significantly more attention?
    - Are there specific scrolling patterns around visualizations?
- **How does device type (mobile phone vs. desktop) influence interaction with visualizations?**
    - Do mobile users really spend less time with visualizations than desktop users?
    - Do preferred types of visualizations differ between mobile and desktop users?
- **What is the relationship between visualization complexity and duration of exposure?**
    - Do more complex visualizations (e.g., multi-slide formats) maintain reader attention longer than simple single-image graphs?
    - How does the type of visualization influence exposure patterns?
- **Can we identify distinct reader typologies based on visualization engagement patterns?**
    - What proportion of readers engage deeply with visualizations versus those who merely glimpse them?
    - Are there systematic patterns in how different reader types approach various visualization formats?

These research questions directly stem from the theoretical frameworks discussed in previous sections, particularly the cognitive-ecological perspective on visualization (section 2.3) and the patterns of interaction in digital journalistic environments (section 2.4). They also directly inform our five testable hypotheses, which we will present in the methodology chapter.

2.5.3 Theoretical Foundations and Conceptual Framework

Our research is grounded in several interconnected theoretical frameworks that together create a comprehensive foundation for analyzing visualization effectiveness in natural reading environments.

These frameworks have been discussed in detail in previous sections, but here we synthesize their key elements and articulate how they specifically inform our research design.

Cognitive Ecology and Distributed Cognition

The theory of distributed cognition (Hutchins, 1995) provides the conceptual foundation for understanding how cognitive processes are not confined to the individual mind but distributed across people, artifacts, and environments. In the context of data visualizations, this implies that well-designed visuals can serve as cognitive tools that extend users' mental capacities—but only if users actually encounter and engage with them.

Building on this, the concept of cognitive ecology (Hutchins, 2010), as discussed in Section 2.3, emphasizes that visualization processing emerges from the interaction between a reader's cognitive capabilities, the design features of the visualization, and the broader socio-technological context. This perspective supports our choice to study visualizations in their natural media environment rather than in isolation. This theoretical framing justifies our focus on measuring exposure and viewport visibility as the basic conditions necessary for any deeper engagement to occur.

Selective Attention and Information Processing

As discussed in section 2.2, theories of selective attention help explain why readers in digital environments often overlook or only superficially process content. The concept of information saturation (Toffler, 1970; Castells, 2010) and the empirical findings on reading patterns (Laban et al., 2017; Grinberg, 2018) provide a framework for understanding how readers allocate limited attention resources when navigating content-rich environments.

This theoretical perspective informs our hypothesis about the influence of position on visualization exposure (H1), suggesting that content position serves as an implicit prioritization cue that influences attention allocation. It also underpins our interest in reader typologies (H5), recognizing that different readers may employ different strategies for managing attention in information-rich environments.

Perceptual and Cognitive Foundations of Visualization

The perceptual theories discussed in section 2.3.1, particularly cognitive fit theory (Vessey, 1991), inform our hypothesis about visualization complexity (H2). These theories suggest that the alignment between visualization format and cognitive task influences processing efficiency, which may in turn affect how long readers engage with different visualization types. The concept of cognitive offloading

(Risko & Gilbert, 2016) further enriches this perspective by highlighting how interactive visualizations can externalize cognitive operations to reduce mental effort. This theoretical insight informs our approach to measuring exposure patterns across different visualization types, particularly those with varying levels of complexity and interactivity.

Device Affordances and Constraints

Theories of device affordances and constraints, as discussed in section 2.4.2, provide a foundation for our hypothesis about device type (H3). The empirical findings on differences between mobile and desktop reading behaviors (Lee et al., 2021; Kim & Heer, 2018) suggest that device characteristics may influence not only how visualizations are displayed but also how readers interact with them.

This theoretical perspective informs our decision to compare exposure patterns across device types, recognizing that mobile and desktop environments may create different conditions for visualization engagement due to differences in screen size, interaction mechanisms, and typical usage contexts.

Methodological Approaches to Measuring Engagement

Finally, the methodological perspectives discussed in section 2.4.4 inform our choice of viewport-based measurement as the primary approach for this study. The work of Laban et al. (2017) and Lagun and Lalmas (2016) provides empirical precedents for using viewport tracking to measure content exposure in naturalistic reading environments. This methodological framework aligns with our pragmatic-analytical approach by focusing on observable behavioral indicators rather than attempting to directly measure internal cognitive processes. It allows us to collect empirically robust data on visualization exposure while acknowledging the limitations of this approach for assessing deeper aspects of comprehension or engagement.

2.5.4 Connecting Theory to Empirical Research

The theoretical frameworks outlined above directly inform the operational design of our research, shaping both what we measure and how we interpret the results. Figure 2 illustrates the conceptual model that guides our empirical investigation.

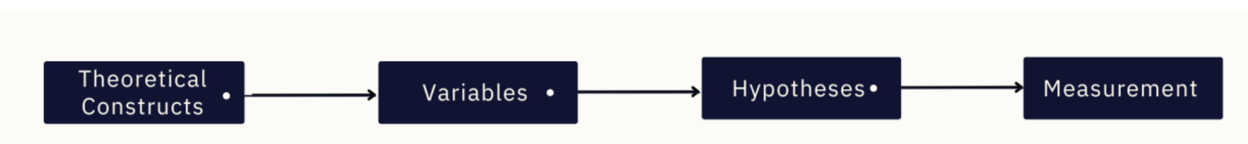


Figure 2: Conceptual model linking theoretical frameworks to research variables and hypotheses

This model shows how position and device type create the foundational conditions for visualization exposure, while complexity and reader typology moderate these relationships. The central dependent variables—visibility probability and exposure duration—represent the observable manifestations of these relationships in reader behavior.

Our research hypotheses, which will be formally presented in Chapter 3, map directly onto these relationships:

- **H1** addresses the relationship between position and exposure probability
- **H2** examines how complexity influences exposure duration
- **H3** investigates the impact of device type on both exposure metrics
- **H4** explores how the position of visualizations affects exposure duration
- **H5** explores how readers can be segmented based on their engagement patterns

The methodological approach, detailed in Chapter 3, operationalizes these theoretical concepts into measurable variables. For example:

- Position is measured as both ordinal position in the article and pixel distance from the top
- Complexity is categorized as simple static visualizations, interactive single-frame visualizations, or multi-slide visualizations
- Exposure is tracked as both binary (visualization appeared in viewport: yes/no) and continuous (seconds in viewport) variables
- Device type is determined through user-agent detection
- Reader typology is derived from clustering analysis of scrolling and exposure patterns

This explicit connection between theoretical concepts, research variables, and methodological approach enhances the validity of our research design and provides a clear framework for interpreting the results. It also acknowledges the limitations of our approach, recognizing that exposure represents a necessary but not sufficient condition for deeper forms of engagement.

The following chapter details the specific methodological procedures employed to implement this research design, including the selection of articles and visualizations, the technical solution for measurement, and the analytical strategies used to test our hypotheses.

Although this study employs a purely quantitative approach focused on behavioral indicators (e.g., exposure time, viewport visibility, interaction logging), this design choice should not be interpreted as a disregard for the qualitative dimensions of data visualization engagement. On the contrary, the theoretical framework and the inclusion of reader typology were informed by a strong awareness of the diverse, situated, and context-dependent nature of how readers interact with visualizations. Rather than attempting to blend methodological paradigms, this study deliberately focuses on scalable, ecologically valid behavioral data, while acknowledging that future research could complement these findings through mixed-method approaches that explore subjective meaning-making and interpretive depth.

3. Methodology and Research Design

3.1 Research Design and Approach

This study focuses on examining how readers interact with data visualizations in online news articles. It is important to precisely define what we mean by "interaction" in the context of this study - we are primarily measuring basic exposure parameters: whether visualizations were displayed on the reader's screen, how long they remained in the field of view, and how far the reader scrolled through the article. This deliberate narrowing of the research scope to measurable exposure indicators (rather than cognitive processing or understanding of content) represents a methodological decision that allows us to obtain empirically substantiated data in the natural context of engaging with news.

The chosen methodology reflects the need to overcome the limitations of laboratory experiments, which often derive from artificial situations with low ecological validity (Heer & Bostock, 2010). In contrast, this research observes reader exposure to visualizations in real time and real context - specifically on the pages of the journalistic portal datovazurnalistika.cz, which publishes typical formats of contemporary Czech data journalism.

It is important to note that the current research design represents a modification of a more ambitious original plan. Initially, this study was conceived as a collaborative project with a major media house that would provide access to a large audience. In addition to pilot measurements, control measurements were planned to run simultaneously on my private website datovazurnalistika.cz using the same methodology as a validation mechanism. In both cases, the same script was to be implemented - the technical execution was identical, and measurements ran in parallel, with the difference being that the audience of the large media house was significantly larger in scale.

Due to unexpected changes in the policy and approach of the media house, it was necessary to reconsider the research strategy. In order to maintain ethical integrity and protect all stakeholders, the decision was made to work exclusively with data from the datovazurnalistika.cz website. Despite these modifications, the research maintains its validity thanks to several strengths: the measurements captured more than 1,000 unique interactions per article during periods of high traffic; the measuring tool demonstrated consistent reliability with data capture rates exceeding 95%; and the research questions were carefully calibrated to align with the original research goals, remaining unchanged throughout the adaptation process. This approach reflects the reality of conducting applied research in a dynamic media environment while maintaining scientific integrity.The use of an authentic environment allows for the consideration of complex factors, such as article length, page layout, device type, or the current reading context, which significantly influence whether visualizations are displayed at all (Bach et al., 2018).

It should be emphasized that this study deliberately does not examine the qualitative aspects of interaction - that is, whether the reader understood the visualizations, how they interpreted them, or whether they actively used them to construct meaning. This limitation does not represent a methodological weakness, but a conscious choice of research design that corresponds to the established goal: to determine if and under what conditions readers encounter visualizations at all. As Franconeri et al. (2021) point out, exposure itself represents a necessary, albeit not sufficient condition for any cognitive processing of visual content. Without basic exposure to visualization, comprehension or utilization cannot occur.

The research design is based on three key concepts, elaborated in the theoretical part of the paper. The first is information overload (Bawden & Robinson, 2009; Eppler & Mengis, 2004), which explains readers' selective attention and tendency to skip visual content in an oversaturated information environment. The second is the concept of visualization literacy (Boy et al., 2014; Kennedy et al., 2016), which reflects differences in readers' ability to interpret different types of visualizations and their willingness to pay attention to them. The third is the theory of distributed cognition (Hutchins, 1995), which understands visualizations as cognitive tools - not just illustrations, but active components of the comprehension process.

This theoretical framework led to the decision to track not the declared comprehensibility of graphs, but the actual exposure level - whether they were displayed at all, how long they remained visible on the screen, and whether the reader scrolled to them. More complex aspects of the cognitive processing of visualizations would require different methodological approaches (e.g., eye-tracking, think-aloud protocols, or in-depth interviews), which would, however, significantly disrupt the natural context of news use.

The chosen approach has its positives, but also limitations, which this paper explicitly reflects:

- The advantage is non-invasiveness and ecological validity of measurement - data was collected on a regular media website without interfering with the user experience, using a script recording the presence of visualization in the field of view.

- The disadvantage is the absence of data on cognitive processing - that is, whether the reader actually paid attention to the visualization, how they interpreted it, or whether they understood it.

These limitations are not overlooked but form an integral part of the interpretative framework of the results, which is addressed in a separate section in the discussion chapter. At the same time, they create space for follow-up research that could analyze the cognitive and interpretative level of interaction with visualizations in greater depth.

While the research problem outlined in Chapter 2 includes a broader set of exploratory questions, the empirical component of this study focuses on the following core questions and five testable hypotheses:

- Which visualizations do readers actually register?
- How much time do they devote to individual visualizations and what are the patterns of their exposure within the article?
- What influence does the placement, type, and complexity of visualization have on the level of exposure?
- Are there differences between devices (desktop vs. mobile)?

Five specific tested hypotheses correspond to these questions, the evaluation of which is the subject of Chapter 4:

H1: The probability that a reader will see a visualization (i.e., it will be displayed on their screen) systematically and rapidly decreases with its position downward within the article.

H2: Visualizations with a higher level of complexity (e.g., multi-slide formats maintain reader exposure for a longer duration compared to single-image graphs).

H3: Users on mobile devices exhibit a lower level of exposure to visualizations than users of desktop devices.

H4: Visualizations placed higher in the article (i.e., earlier during scrolling) receive, on average, more exposure time from readers than visualizations placed lower.

H5: Among readers, different groups can be identified according to patterns of interaction with visualizations, including a subgroup with a high level of engagement (highly engaged users).

To verify these hypotheses, metrics for visualization visibility on the screen and the duration of this visualization in the field of view were established, as detailed in the following subchapters. The results of this approach allow for the formulation of empirically substantiated claims about the basic conditions of exposure to visualizations in online journalism, although they do not provide insight into deeper aspects of cognitive processing.

3.1.1 Comparative Dimension: Bayerischer Rundfunk Article

To verify the validity of our findings beyond the Czech context, we also included an article published by the German public broadcaster Bayerischer Rundfunk (CO_2-Rechner, Bayer, 2024) in the analysis. Although it was not possible to implement measurements using the same technique for this case (access to server infrastructure was lacking), thanks to the author's (Constanze Bayer) willingness, I had detailed aggregated data on user behavior at my disposal. This article serves as a comparative case, and its results will be used in Chapter 4 to contextualize the main trends.

3.2 Selection of Articles and Visualization Sample

The empirical part of the research was conducted on four articles published on the website datovazurnalistika.cz during the period of June–July 2024. These articles were selected based on their thematic diversity, representation of various visualization formats, and their structural length, which allowed for the analysis of reader behavior in different contexts.

The selection was purposive, not random – it reflected the need to create a representative sample of contemporary Czech data journalism. All articles contain more than two data visualizations, some of which were interactive, some static, and some multi-panel (multi-slide).

3.2.1 Selection Criteria

Articles were selected according to the following parameters:

- thematic diversity (defense, elections, preferences, voter motivations),
- representation of different visualization strategies (interactive graphs, race chart, time trends),
- presence of multiple visualizations in different positions within the article,
- technical feasibility of measurement (own infrastructure, possibility of scripting),
- editorial availability for placement of the measuring code.

It was also considered whether the article reflected the genre elements of data journalism – especially non-linear structure, combination of textual and graphical content, and work with a narrative framework (partial scrollytelling).

3.2.2 Structure of Articles and Visualizations

The following table summarizes the key characteristics of individual articles, including their topic, number of visualizations, and predominant type of graphic processing:

Article	Topic	Number and types of visualizations
A 1	**Support for Mandatory Military Service in the Czech Republic**: A survey shows skepticism about the country's independent defense capabilities, with half of respondents supporting the reintroduction of mandatory military service, particularly for men, alongside growing but insufficient interest in voluntary military reserves.	3 Multi-slides
A 2	**Motivations Behind Czech Voters in European Elections**: Record-high turnout in Czech European elections was driven by protest votes against the government and EU membership, with younger and progressive voters leaning toward pro-European parties and older voters favoring traditional options.	3 Single chart Moving chart Single Chart with more interactions
A 3	**Evolution of Czech Political Preferences**: Tracking voter trends since the 2021 elections highlights coalition dynamics for 2025, showing shifts influenced by undecided voters, with coalitions like Spolu and Přísaha-Motoristé playing key roles in reshaping political competition.	5 Multi-slides Single charts
A 4	**Voter Shifts and Radicalization in European Elections**: Analysis of voter movement reveals significant abstentions and radicalization among ANO supporters, alongside struggles for most parties to retain their 2021 voters, while protest sentiment benefits fringe and opposition parties.	5 Single Charts

Table 4: Overview of Articles

3.2.3 Characteristics of Visualizations

A total of 16 visualizations were analyzed, which can be divided into three basic categories:

- Simple static visualizations – for example, single-column graphs with a caption.
- Interactive graphs in single-slide mode – responding to movement or clicks, but without a narrative structure.
- Multi-slide visualizations (horizontal scrollytelling) – a series of slides showing different aspects of one data set, typically accompanied by text.

Visualizations were placed in different parts of the article – some directly below the headline, others in the final part of the text. This vertical diversity allowed for monitoring the influence of placement on the level of exposure and length of interaction (see hypotheses H1 and H4).

3.2.4 Comparative Case: Bayerischer Rundfunk Article

The interactive article CO_2-Rechner published by the German public broadcaster Bayerischer Rundfunk (Bayer, 2024) serves as a supplementary case. This case is not part of the main data set, as it was not possible to deploy our own measuring infrastructure, however, the editorial team provided aggregated data on reader behavior. The article is used for comparative comparison, especially in terms of engagement with complex scrollytelling visualizations in a foreign environment (see also Chapter 4.3).

3.3 Technical Measurement Solution

For the purposes of the research, a custom measuring tool was developed to enable monitoring of reader behavior in the real environment of web articles. Measurement was carried out in a passive form – without any interference with the user experience – and was based on a combination of JavaScript, HTML elements, and Firebase cloud.

The basic limitation of the measurement was that it was not possible to measure activities within the visualization itself (e.g., interactions, mouse movements, etc.). All visualizations were created on the Flourish platform and inserted into the body of the article using the HTML iframe element. This is a so-called sandboxed element, where the surroundings (in this case, the article) have no access to the embedded content for security reasons, in this case, to the embedded visualization.

3.3.1 Basic Principle of Measurement

The aim was to obtain three key variables:

1. visibility of visualization (whether it was displayed on the screen),
2. display time (number of seconds for which the visualization was visible),
3. scrolling behavior (i.e., depth of reader's progress in pixels).

Visualization data collection: Given that all visualizations were created using Flourish and inserted into articles using HTML iframe elements, it was not possible to measure direct reader interactions within visualizations (e.g., clicks, mouse movements). Instead, the time during which the visualization was in the visible part of the screen (the so-called viewport) was monitored.

Each visualization was wrapped with three elements:

- Upper hidden element (HTML <div>),
- Iframe with visualization (HTML <iframe>),
- Lower hidden element (HTML <span>).

Two events were monitored for these three elements:

- "Entered visible" – the element entered the visible area of the screen,
- "Exited visible" – the element left the visible area of the screen.

Event recording process: Each interaction event (the moment when the visualization appeared in or disappeared from the reader's field of view) triggered two records in the database: one from the iframe and one from the corresponding upper or lower hidden element.

For example:

- When a reader scrolled to a visualization, the following were recorded:
    - "Entered visible" for the upper element,
    - "Entered visible" for the iframe.
- When the reader continued scrolling and the visualization disappeared from the screen:
    - "Exited visible" for the iframe,
    - "Exited visible" for the lower element.

The recorded data in this case also contained:

- Article ID (according to URL),
- Visualization ID (based on its position in the article).

The following information was also recorded for each reader for further analysis:

- Randomly generated unique ID,
- Device type: mobile or desktop.

The first timestamp was created when the upper part of the visualization entered the screen, the second when the lower part appeared. The difference between these two time points gives the time during which the reader actually perceived the visualization. The same principle applied when moving from bottom to top.

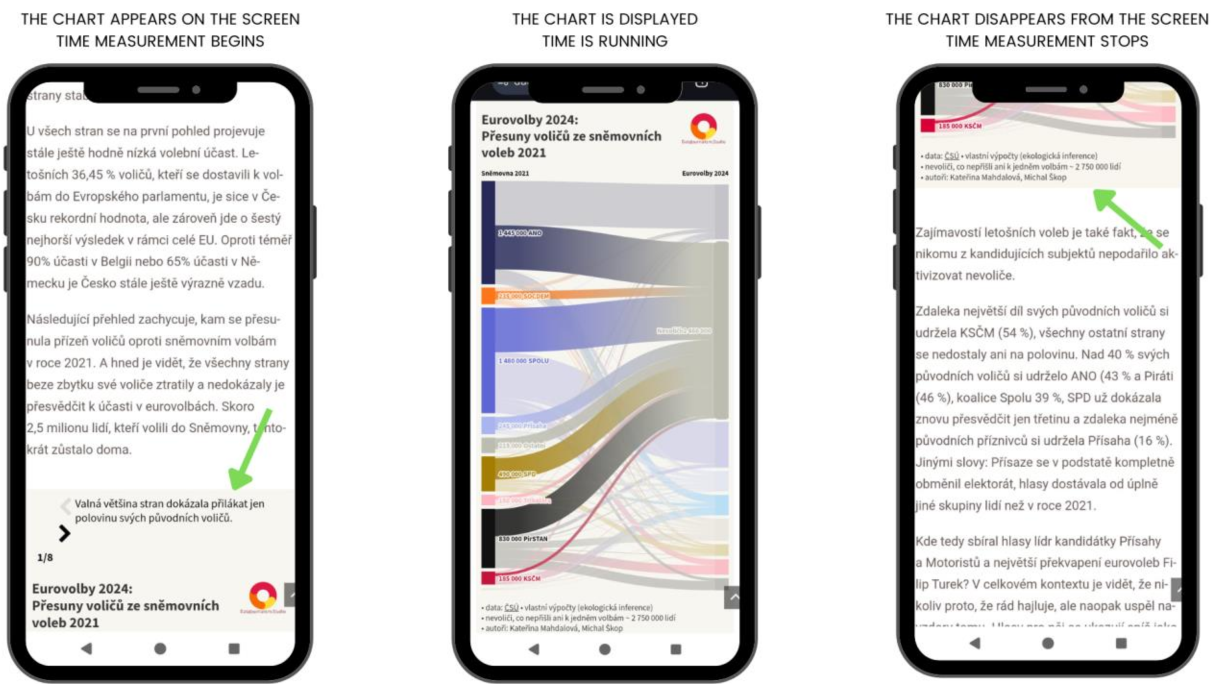


Figure 3: Event recording process

The recorded data was stored in the Firebase Realtime Database (Google, 2024) using a JavaScript script. Due to the technical limitations of the editorial publishing system and templates used, it was not possible to record detailed interaction within the visualization itself.

3.3.2 Custom Script vs. Ready-Made Analytics

Instead of commonly used web analytics tools (e.g., Google Analytics, Plausible, Matomo, and others), a custom measuring script was chosen. This step was guided by the following reasons:

- Higher data granularity – it was necessary to track specific visualizations within the article, not just the page as a whole.

- Full control over measurement – allowed defining custom collection logic, timing, and data structure.
- Ethical and legal clarity – standard analytical tools often process personal data, while this approach was completely anonymous and under the full control of the researcher.

On the other hand, analytics tools like Google Analytics also estimate additional information, e.g., demographic data about age and gender or reader location. This information was not available within the chosen approach with a custom measuring script.

3.3.3 Reliability and Operational Aspects

The script had minimal impact on page performance, was easy to implement, and its operational costs (e.g., Firebase hosting) were in the order of several thousand Czech crowns. Data was stored in the Firebase database (Google, 2024), which was chosen for its simple implementation and reliability. However, several limitations emerged:

- Exporting data for analysis was relatively complex,
- Operational costs could increase with more extensive analyses.

Further key observations:

- Time records of the lower elements showed high precision (in 95% of cases, the difference was less than 0.02 s).
- Redundant monitoring through three sensors (top, iframe, bottom) created a robust system for data validation and backup in case of failure of one of the elements.
- Due to asynchronous HTML rendering, it was shown that data from the upper <div> tags were often less accurate and were therefore excluded from the analyses.

The approach used proved to be very effective for passive attention measurement in web articles. It had low technical requirements, was ethically undemanding, and easily transferable between different publishing environments. The technical solution proved to be very stable – more than 99% of events were correctly recorded. The use of three measuring points made it possible to detect and resolve cases of failure of one of the elements.

For the first article, a random sample was deployed for preventive reasons (approximately 5% of readers), but due to the trouble-free operation, the entire audience was measured in subsequent articles.

An example code with the detailed description of the script is provided in Annex 1.

3.3.4 Operationalization of Observed Variables

For the purposes of empirical analysis, it was necessary to transform theoretical concepts into measurable variables, which enabled systematic testing of the established hypotheses. This subchapter defines how individual key variables were operationalized to ensure methodological clarity and research reproducibility.

Visualization complexity was operationalized based on the formal characteristics of visualizations, not their content complexity. We established three distinct levels using the following criteria:

- Simple static visualizations: Single-frame graphs without interactive elements, requiring only passive viewing.
- Interactive single-frame visualizations: Graphs that respond to user actions (hovering, clicking) but maintain a single information view.
- Multi-slide visualizations (horizontal scrollytelling): Sequential presentations showing different aspects of one dataset across multiple slides, requiring active navigation.

This classification reflects increasing demands on user engagement and cognitive processing, considering both the volume of information presented and the interaction mechanisms required to access it. In quantitative analyses, complexity was coded as an ordinal variable (1 = simple, 2 = interactive, 3 = multi-slide).

A visualization was considered "displayed" if any part of it appeared in the visible area on the user's screen (viewport). This binary variable (0 = not displayed, 1 = displayed) was determined by tracking "entered visible" and "exited visible" events for HTML elements surrounding the iframe with the visualization. This approach allowed identification of whether the reader had the opportunity to perceive the visualization at all.

Exposure duration was defined as the total time in seconds during which the visualization was visible on the user's screen. It was measured as the sum of all intervals between "entered visible" and "exited visible" events. While this variable quantifies potential attention devoted to the visualization, it does not guarantee active cognitive processing.

Visualization position was operationalized in two complementary ways: (1) as an ordinal position (first, second, third, etc.) reflecting narrative sequence, and (2) as a metric distance (in pixels) from the top of the page to the upper boundary of the visualization. This dual approach allowed for nuanced analysis of how both structural and spatial positioning influence user behavior.

Device type was determined using the user-agent string and screen resolution, distinguishing between desktop (computers and laptops) and mobile devices (smartphones and tablets). This binary classification enabled testing for differences in visualization interaction across device categories, given their differing screen sizes and usage contexts.

Reader engagement segments were identified based on scrolling behavior and exposure time. Three primary groups emerged:

- Inactive readers (“non-scrollers”) who performed minimal or no scrolling.
- Partial readers who progressed partway but did not reach the final visualization.
- Complete readers who reached the last visualization.

Within the last group, highly engaged readers were defined as those in the top quartile of exposure duration for a given visualization. These users showed exceptional interest, spending significantly more time on visual content than the average reader.

This multidimensional operationalization enabled the testing of all hypotheses while ensuring ecological validity. By precisely defining these variables, the research established a robust methodological foundation for analyzing real-world interactions with data visualizations in digital journalism.

3.4 Limitations and Ethical Aspects of Research

Every research design brings, alongside its benefits, certain limitations and risks – both in terms of the informative value of the data obtained, and in terms of the ethical principles of collecting and interpreting information. This chapter addresses precisely these aspects which were taken into account during the design and implementation of the research.

3.4.1 Ethical Principles

The research was conceived with an emphasis on fulfilling the basic ethical principles of working with digital data, especially in the context of monitoring user behavior on the web. The key starting point was to ensure that the collection and processing of data took place in full compliance with privacy protection requirements, legal regulations including GDPR, and the principles of research transparency.

Data was collected exclusively in the form of technical records of user behavior on the page – specifically scrolling, visibility of selected elements, and the timestamp of these events. No personally identifiable data (PII) was collected, such as IP addresses, location data, demographic information, or persistent cookies. Identification of an individual was not possible at any point, which corresponds to the principles of data minimization according to GDPR.

Moreover, the technical solution was fully under the control of the researcher. Data was stored in the secure Firebase environment (Google, 2024), without the involvement of external analytics service providers. Independence from commercial analytics platforms was motivated both by the desire for a higher level of granular measurement and by the ethical preference for full control over what data is collected, where it is stored, and who has access to it.

Methodological transparency was also an important principle. All research choices – from the selection of measured articles through the technical parameters of the script to the decision about which events to record – are described in detail and justified in this paper. This approach allows for replication of the research or critical reconsideration of the chosen procedures.

3.4.2 Methodological Limitations

The chosen approach to measuring reader attention offers several advantages – primarily the ability to observe user behavior in their natural context – but also carries important limitations that affect the interpretation of results. These limitations are not only technical in nature but also relate to what exactly can be captured by the chosen method.

The basic limitation lies in the fact that the research measures only whether the visualization was displayed on the page, not whether the reader paid attention to it, let alone understood it. The fact that a graph was in the visible part of the screen for several seconds does not necessarily mean that it was actively perceived or interpreted. Such an event represents only exposure, not engagement or understanding (Kennedy et al., 2016).

Another specificity concerns the very manner of reading in the digital environment. Some visualizations may be technically displayed (e.g., during fast scrolling), but the reader may not actually register them. This phenomenon corresponds to the so-called visualization blindness, i.e., the tendency to completely ignore some visual stimuli, whether consciously or unconsciously (Bach et al., 2018).

It is also important to distinguish that tracking movement and time on the page does not reveal attitudes, values, or cognitive strategies of readers. The obtained data do not allow for the reconstruction of interpretative processes – we do not know whether the reader identified with the graph, whether they trusted it, or considered it comprehensible.

These methodological limitations were nevertheless deliberately accepted in exchange for higher ecological validity of the research – that is, the ability to capture real behavior of readers in the real environment of a news website, without researcher intervention. This approach corresponds to the growing emphasis on studying visualizations "in the wild," outside laboratory conditions, as formulated, for example, by Hullman and Gelman (2021), who point out the discrepancy between the results of controlled studies and the reality of interaction with visualizations in the common media environment.

For these reasons, the research deliberately limited itself to basic interaction indicators – time, scrolling, visibility – which form an empirically graspable framework for comparing different types of visualizations.

3.4.3 Epistemological Limitations

In addition to methodological limitations related to the measurement technique, it is also necessary to reflect on the epistemological boundaries of the entire research approach – that is, the question of what type of knowledge can be obtained from the given data, and what, on the contrary, remains beyond their reach.

The fundamental premise is that all measured variables – scrolling, visibility of elements, or the duration of visualization display – capture only the external manifestations of reader activity, not the understanding itself, semantic interpretation, or cognitive processes that take place "inside the reader's head." Such data may suggest which parts of the article were attractive to users or which visualizations might have engaged more than others, but they say nothing about what the reader understood, what meaning they derived from the visualization, or how they placed it in a broader interpretative framework.

In this sense, the research should be understood as a study of attention, not understanding – as an empirical insight into what the reader "could" perceive, not into what they perceived or understood. These limitations are in line with the theoretical framework of distributed cognition (Hutchins, 1995), which was presented in Chapter 2. Visualizations here are not understood as neutral carriers of information, but as part of a broader cognitive system, in which not only the graph itself plays a role, but also the context, user, medium, and specific reading situation. The meaning of visualization is thus not encoded only in its form but arises in the interaction between the reader and the environment. From this perspective, it is not possible to guarantee that a technically recorded interaction automatically means the emergence of understanding – just as its absence does not mean its non-existence.

Moreover, visualizations often have an implicit effect – the reader may use them to orient themselves, structure the article, or perceive them as a "backdrop" of credibility without actively interpreting them. This aspect of interaction, which is often emotional, aesthetic, or socially mediated, is not captured by the research and cannot be captured.

The purpose of this work is not to question the role of data visualizations in journalism, nor their ability to convey information. However, the research tries to rigorously define what can be said about reader behavior based on the measured data – and what remains beyond their reach. Precisely this distinction between what we know and what we merely presume is essential for the interpretation of results, which is part of the following Chapter 4.

3.4.4 Critical Reflection: Why This Approach?

The chosen approach to measuring reader attention stems from a conscious decision to focus on such aspects of interaction that can be reliably and non-invasively captured in the real environment of a news website. It is not a "technological maximum," but a thoughtful research compromise that takes into account both the practical limitations of measurement and the epistemological boundaries of data interpretation.

The reason we focused specifically on the visibility of visualizations, time spent on individual graphs, and the process of scrolling is the fact that these variables form the basic condition for any further interaction. A visualization that is not displayed or is immediately skipped cannot fulfill any informative, orientational, or persuasive function. Exposure thus represents a minimal prerequisite for understanding – although in itself it does not guarantee understanding.

At the same time, it was clear that alternative methods of data collection – such as questionnaires, comprehension tests, or qualitative interviews – would significantly disrupt the natural reading environment, and thus the validity of the behavior itself. That is why the research decided to monitor real interactions, as they spontaneously occur in regular media operations, and not under laboratory or controlled conditions.

This approach allows us to describe certain general patterns – for example, when readers typically leave articles, which types of visualizations tend to be overlooked, or whether the position of a graph in the text affects the degree of its registration. On the other hand, it tells us very little about the motivations of readers, their understanding, or evaluation of visualizations. The research thus does not deal with what people think, but with what they do on the page – and it is precisely this level that it considers insufficiently described and methodically graspable.

It is noteworthy that even though these metrics are very basic – whether an element was on the screen and for how long – data of this type from real journalistic operations usually remain unavailable. Commercial analytics tools may measure the performance of entire articles or pages, but they typically do not reach the level of individual visualizations. As a result, the actual interaction of the audience with visualization content is often merely speculated upon – without empirical support.

The chosen method thus does not capture the entire semantic potential of data visualizations – but focuses on their minimum conditions of functioning in the context of reader behavior. By doing so, it offers a framework that can be further developed in combination with other types of research (e.g., ethnographic, psychological, or experimental).

4. Analysis and Results

This chapter presents the results of measuring reader exposure to visualizations in selected news articles. The aim is not only to describe basic statistics about how many users saw a given part of the article and how long they stayed with it, but primarily to compare different types of visualizations, their position in the text, differences between devices, and behavior patterns across the audience. The results are structured to address the formulated research questions and to test the five research hypotheses outlined in Chapter 3.

The analysis is based on data collected on the website datovazurnalistika.cz during June and July 2024. The behavior of visitors who interacted with four articles containing a total of 16 data visualizations was analyzed (see Chapter 3.2). The technical solution for measurement was described in Chapter 3.3, and the basic methodological and epistemological frameworks in Chapter 3.4.

4.1 General Patterns of Reader Behavior

This subchapter focuses on how readers navigate through articles as a whole, regardless of the content of individual visualizations. We track their progressive scrolling of the page, the rate of readership retention, and typical exit points. The aim of this section is to answer the research question: *How do readers move through the article?* while simultaneously testing hypothesis H1, which assumes that the probability of a visualization being displayed to a reader systematically decreases with its position moving downward within the article.

4.1.1 Funnel Visualization of Readership Retention

A basic view of reader behavior is offered by so-called funnel graphs, which show the proportion of readers who reached individual visualizations in a given article. For each of the four articles published on the datovazurnalistika.cz portal, these proportions were calculated based on data from the recording system (see Chapter 3.3), with each visualization block tagged at the beginning and end. Based on the capture of these tags, it was possible to determine how many readers actually "saw" the visualization, i.e., it had the opportunity to be displayed on their screen.

The results suggest a consistent trend: as reading progresses through the article, the proportion of users who reach subsequent visualizations decreases. For example, in the case of Article 1, approximately 80% of all visitors reached the first visualization, while only about 44% displayed the last visualization. Similar trends can be observed in other articles – the most significant decline was recorded in the longer Article 3, where the proportion of readers dropped from 94% for the first visualization to 32% for the last.

Graph 1 Funnel Charts of Reader Retention by Visualizations

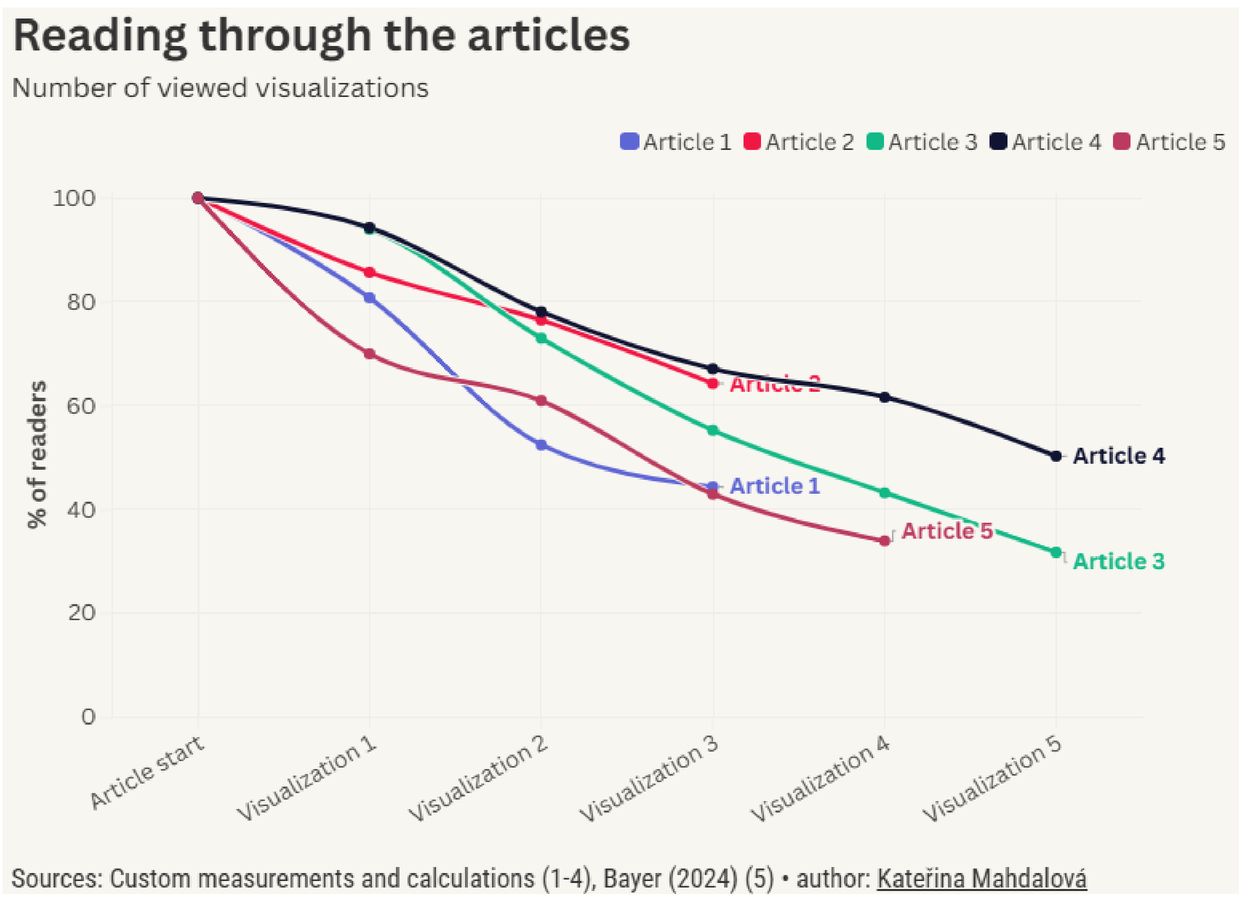
This graph shows the percentage of readers who progressively moved through individual visualizations in the four analyzed articles. Each funnel represents one article with a trend of decreasing readership as readers progress to further visualizations.

4.1.2 Aggregated Patterns of Readership Retention

To understand reader behavior across individual articles and to compare the development of readership retention depending on content position, two sets of aggregated graphs were created. The first tracks readership according to the order of visualizations in the article (i.e., first, second, third, etc.), while the second tracks readership depending on the vertical position of the visualization on the page (in pixels). These two measurement methods provide a complementary view: the first reflects the narrative structure of the article, the second the physical length and layout of the content.

A fifth article published by the German public broadcaster Bayerischer Rundfunk (CO_2-Rechner, Bayer, 2024) was also included in this part of the analysis. This article was included as a comparative case outside the Czech context (see Chapter 3.1), although it was not possible to implement the same measurement technique for it. The author of the article provided aggregated readership data, which allowed it to be included in the summary visualization. It is labeled as Article 5 in the graphs.

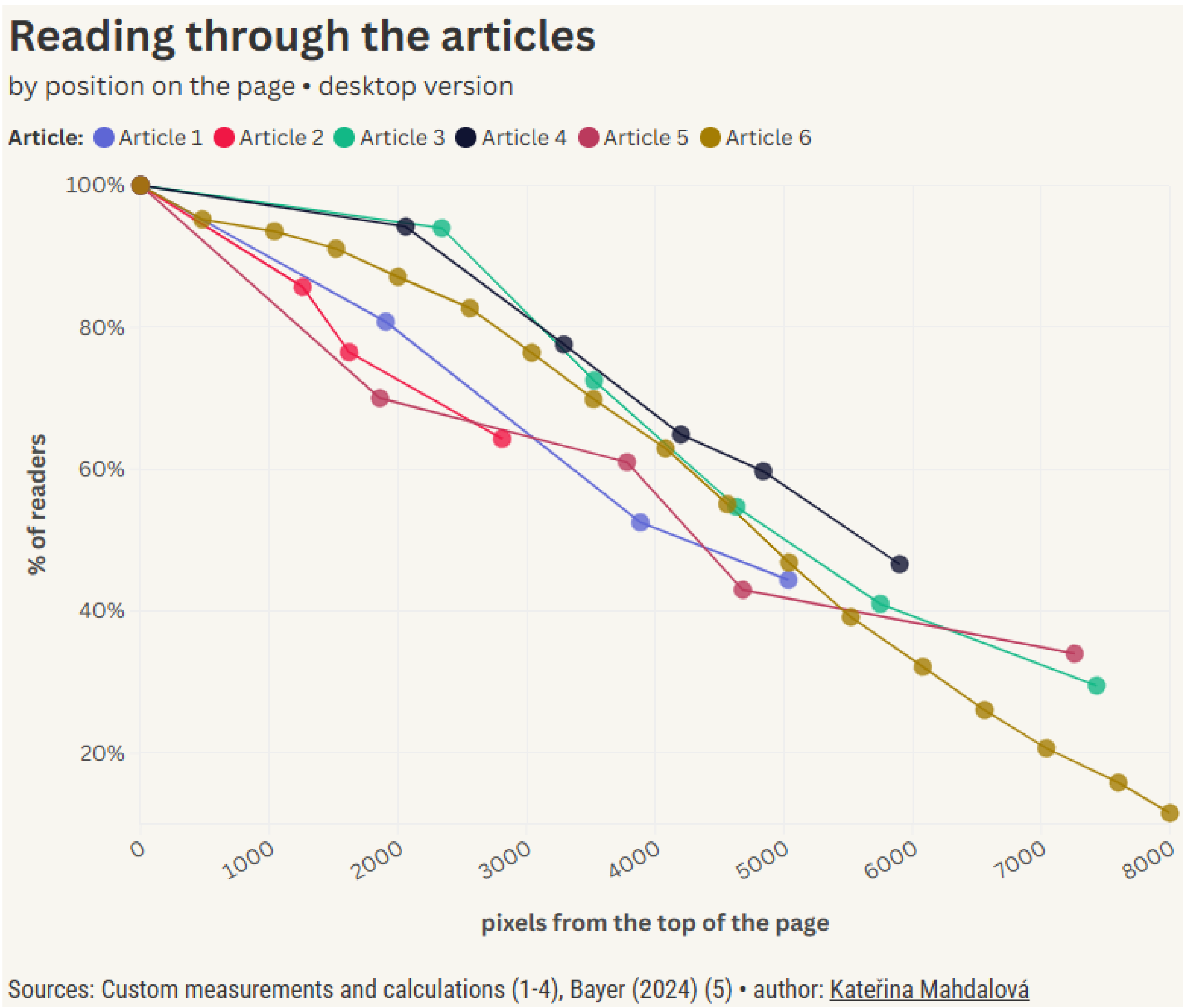
The first graph shows the proportion of readers who reached each subsequent visualization, regardless of the specific article. For all cases, including the Bayer (2024) article, there is an apparent decrease in reader retention depending on the order of visualization, suggesting the consistency of this phenomenon. The loss of readers is particularly evident between the first and second visualization, with each subsequent one losing a smaller but steady share of the audience.



Graph 2 Reading Through the Articles: Number of Viewed Visualizations

The line chart summarizes reader retention across the five articles based on the order of visualizations. A consistent trend of reader drop-off is visible, with fewer readers engaging with later visualizations.

The second graph tracks readership as a function of vertical position on the page, measured in pixels from the top edge. This approach allows for tracking the physical aspect of interaction – that is, which parts of the article were still displayed on the reader's screen, and which were not. The results suggest that across all articles – including the Bayer case – there is a gradual decrease in the proportion of readers as the distance from the beginning of the page increases. The further the content lies from the beginning, the smaller the proportion of users who reach it.



Graph 3: Reading Through the Articles: Reader Retention by Page Position

This chart shows the percentage of readers remaining at different vertical positions (measured in pixels) in the articles. The downward trend highlights how readers gradually leave the page as they scroll further down.

The inclusion of the Bayer (2024) article in both graphs suggests that similar behavior patterns may occur even in a different editorial and cultural context. Although the article uses a different type of scrollytelling structure and different visualization formats, readership retention decreases at a similar rate to that observed in the analyzed Czech articles. This supports the assumption that content position on the page may be a significant predictor of reader retention, and simultaneously that this effect may be observable across different contexts and languages.

4.1.3 Quantification of Decline Rate (Regression Analysis)

For a more precise quantification of the readership decline, a linear regression analysis was performed, modeling the relationship between the vertical position of visualizations and the percentage of readers who reached this position. All points on the y-axis (in pixels) were included in the calculation, except for the very top part of the page (0–200 px), where a large number of immediate reader departures occurred without actual engagement in reading – that is, before any screen movement even took place.

The regression model was applied separately to each article. In each case, the output was a negative slope of the regression line, corresponding to the proportion of readers who "left the page" with each increasing section of article length. The regression results were as follows:

- Article 1: decrease of 11.6% for every 1000 pixels
- Article 2: decrease of 13.8% for every 1000 pixels
- Article 3: decrease of 12.4% for every 1000 pixels
- Article 4: decrease of 12.1% for every 1000 pixels.

The average value across all articles is 12.5% of readers who leave the page with each additional 1000-pixel section of the article. This calculation was performed using a linear regression model in a spreadsheet processor with coefficients of determination (R^2) ranging from 0.78 to 0.89, suggesting a relatively good predictive capability of the model (all p-values < 0.001).

The results of this analysis provide support for hypothesis H1. They suggest that content readership in online articles with visualizations decreases at a relatively regular rate, regardless of the type of visualization or the overall length of the text. This finding may have practical implications for editorial practice, especially when deciding where to place individual visualizations within an article.

While the phenomenon of reader attrition in online articles is well-established in media practice, this research provides precise quantification that can guide editorial decisions. Based on our findings, editors and data journalists can expect to lose approximately 12.5% of their readers with each 1000-pixel downward scroll. This translates to practical recommendations: critical visualizations should be placed within the first 2000 pixels of content to ensure at least 75% of the initial audience will see them. For content requiring engagement with multiple visualizations, a multi-page approach might be more effective than a single long-scroll article. These specific, quantitative insights provide actionable editorial guidelines that go beyond the general understanding that 'readers leave as they scroll'.

4.2 Duration of Visualization Exposure

This section builds on the previous analysis of article navigation (Chapter 4.1) but focuses exclusively on those cases where the reader displayed the visualization on their screen. We therefore track how long the visualization remained in their field of view, and how this duration differs depending on various factors. The analysis is based on time data recorded by the script (see Chapter 3.3), which determined the moment of entry into the visible area of the screen and its departure for each visualization.

The main goal of this section is to find out how much time readers spent with visualizations, and how this time is influenced by variables such as the type of visualization (simple graph vs. multi-slide), the position of the visualization within the article, or the type of device (mobile vs. desktop). The results of the analysis address the research question: *How much time do readers spend on individual visualizations?,* and allow testing of three hypotheses:

- **H2**: More complex (e.g., multi-slide) visualizations maintain attention longer than simple single-image graphs.
- **H4**: Visualizations higher in the article receive more attention time from readers.
- **H3**: Users on mobile devices devote less time to visualizations than users on desktop.

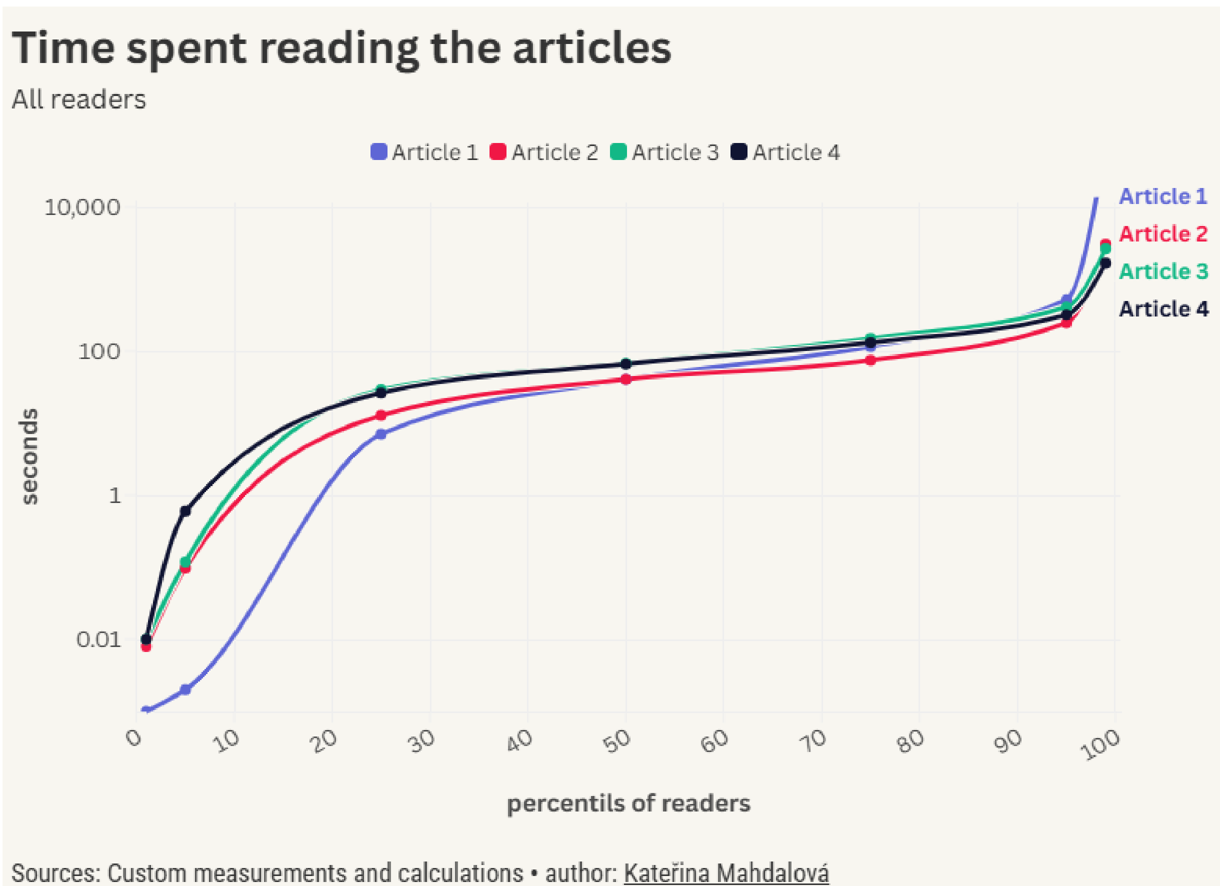
The results of the individual tested hypotheses are presented in the following subchapters.

4.2.1 Differences in Exposure Time Between Articles and Types of Visualizations

The time that readers spent on visualizations varied between articles. For example, the median reading length of Article 3 (which contained complex multi-slide graphs) reached 68.4 seconds, while Article 1, with a simpler structure, had a median of 40.7 seconds. These differences, however, cannot be attributed solely to the type of visualizations – articles also differed in text length, topic, and overall context.

Time (s)	Article 1	Article 2	Article 3	Article 4
25% Quantile	7.1	12.9	29.2	26.4
Median time (s)	40.7	40.7	68.4	66.4
75% Quantile	116.3	74.7	150.4	131.6

Table 5: Distribution of time spent reading the articles, all readers

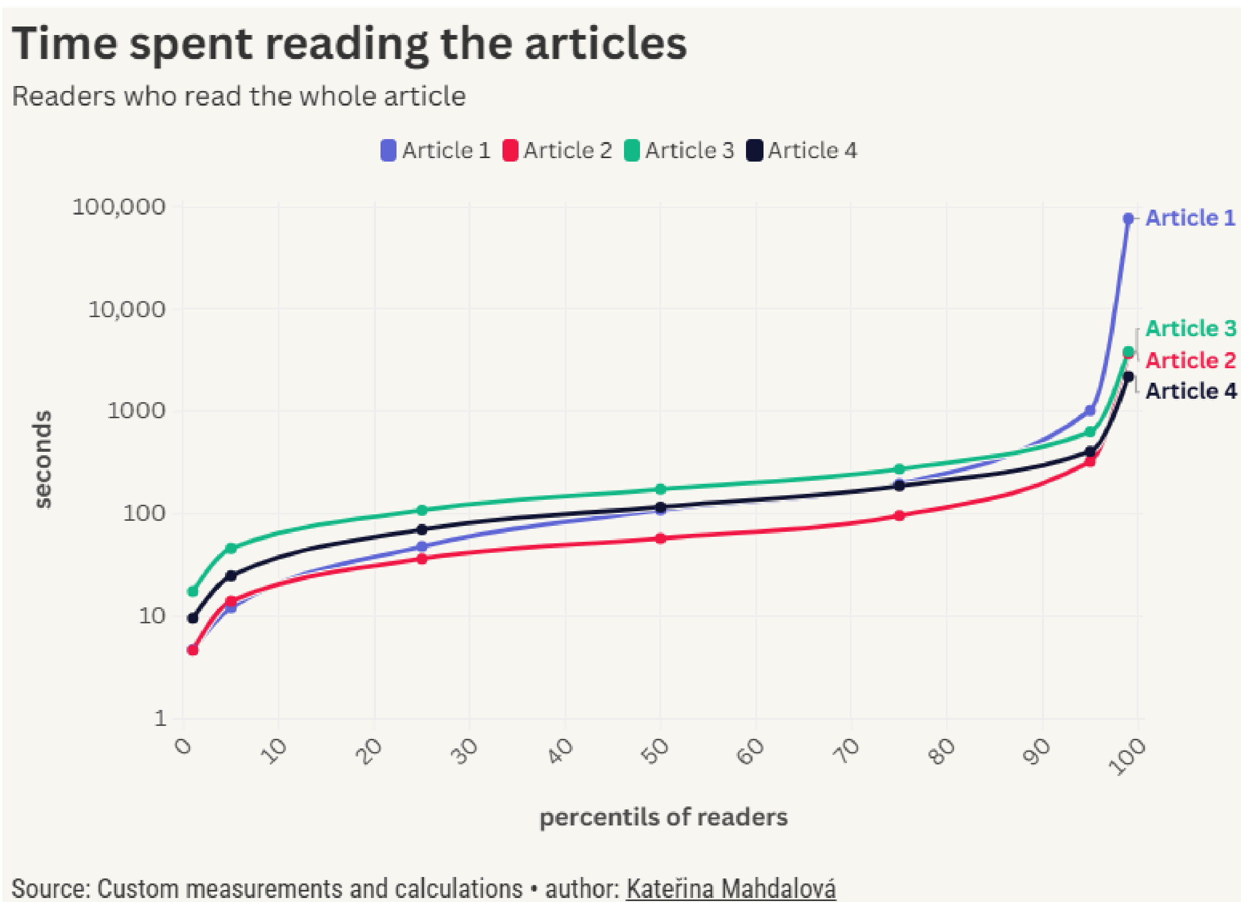


Graph 4: Distribution of time spent reading the articles, all readers

More pronounced differences emerge if we look only at those readers who reached the last visualization of the article (i.e., those who likely paid more attention to the article). For this subgroup, the median times were significantly higher:

Time (s)	Article 1	Article 2	Article 3	Article 4
% of all readers	44.4%	65.6%	31.8%	50.3%
5% Quantile	12.2	14.1	46.1	24.9
25% Quantile	48.0	36.6	108.5	70.4
Median time (s)	109.2	57.7	174.2	116.7
75% Quantile	196.1	96.0	273.9	187.0
95% Quantile	1024.8	325.4	631.6	408.2

Table 6: Distribution of time spent reading the articles, readers who read the whole article

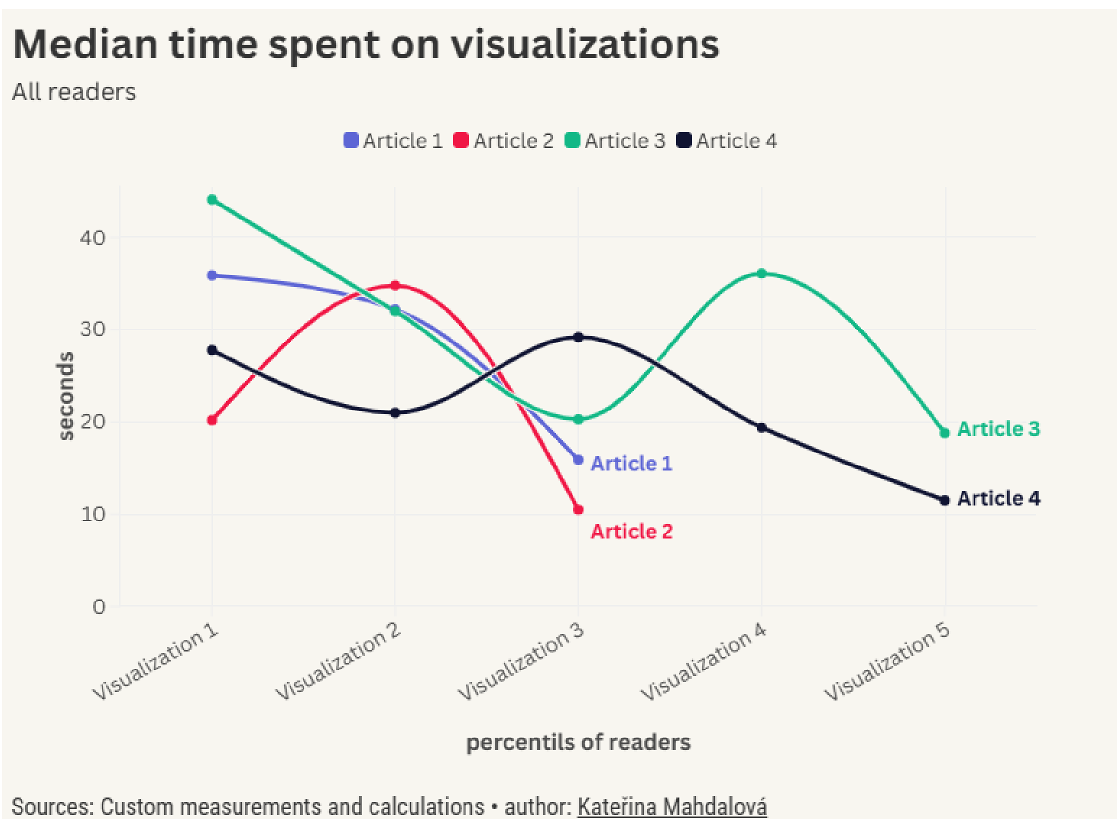


Graph 5: Distribution of time spent reading the articles, readers who read the whole article

This chart focuses on readers who reached the last visualization in each article, providing insight into their engagement time. Articles with more complex visualizations, such as Articles 3 and 4, had generally significantly higher reading times.

4.2.2 Influence of Visualization Type: Simple vs. Multi-slide

Hypothesis H2 assumed that more complex (multi-slide) visualizations would maintain reader attention for a longer time. The data suggests that visualizations composed of multiple slide sequences indeed exhibited longer median display times than simple graphs, which provides some support for this hypothesis. However, the relationship between complexity and exposure time is not linear – for example, visualizations with four slide screens were not viewed four times longer than those with a single slide. These results suggest that beyond a certain level of complexity, adding more slides may not result in a proportional increase in reader time. From an editorial practice perspective, this may mean that extensive multi-slide sequences may not always bring a significantly higher level of reader engagement than concise but well-structured interactive graphs. It should be emphasized that these findings are based on a limited number of cases, especially in the category of multi-slide stories (for example, visualizations 1–3 in articles 1 and 3). The number of analyzed multi-slide visualizations (n=5) is too small for robust statistical analysis, which should be taken into account when interpreting the results.



Graph 6: Median time spent on visualizations, readers who read the whole article

This line chart shows the median time spent on each visualization by readers who completed the entire article. While engagement generally decreases for visualizations lower in the article, time spent also depends on the content and complexity of the visualizations.

The findings thus provide partial support for hypothesis H2 but also point to the importance of optimizing the length and form of interactive visualizations. Effective visualization design appears to be more a matter of finding balance rather than simply maximizing the number of interactive elements.

4.2.3 Influence of Visualization Position in the Article

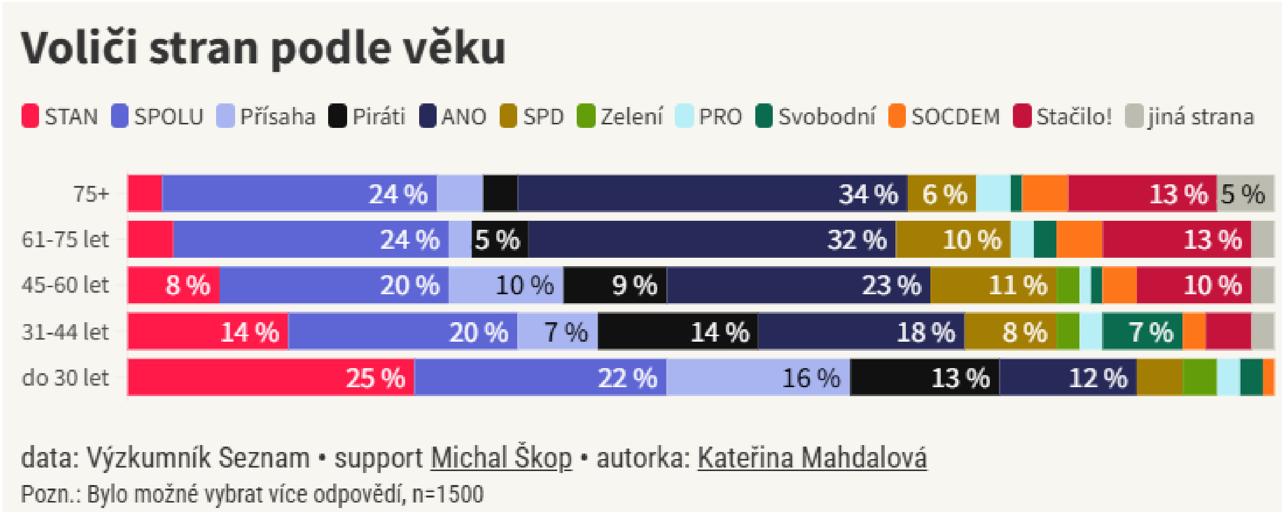
While the previous section showed a potential relationship between visualization type and exposure length, in this section we focus on whether and how exposure time differs depending on where the visualization is placed within the article. This analysis addresses hypothesis H4, which assumes that visualizations higher in the article receive more attention than visualizations placed lower.

The results are shown in the graph in the previous section, which shows the median exposure time for individual visualizations. The data includes only readers who reached the end of the article, thereby minimizing bias caused by selective departure.

From the graph, an overall declining trend is clear, where later (lower placed) visualizations often have shorter display times. However, this relationship is not strictly linear – some later visualizations, especially if they were content-rich or formally distinct, were able to hold readers' attention longer. The influence of position therefore cannot be understood in isolation, but as one of the factors that may influence reader behavior in conjunction with content and visualization type.

Notable Exception to General Trends: Visualization 4 in Article 3

One visualization stood out significantly from the general trends: visualization 4 in Article 3. Readers spent significantly more time with this visualization compared to others in similar positions or with comparable complexity. This visualization apparently managed to capture readers' interest exceptionally well, as shown in the accompanying graphs.



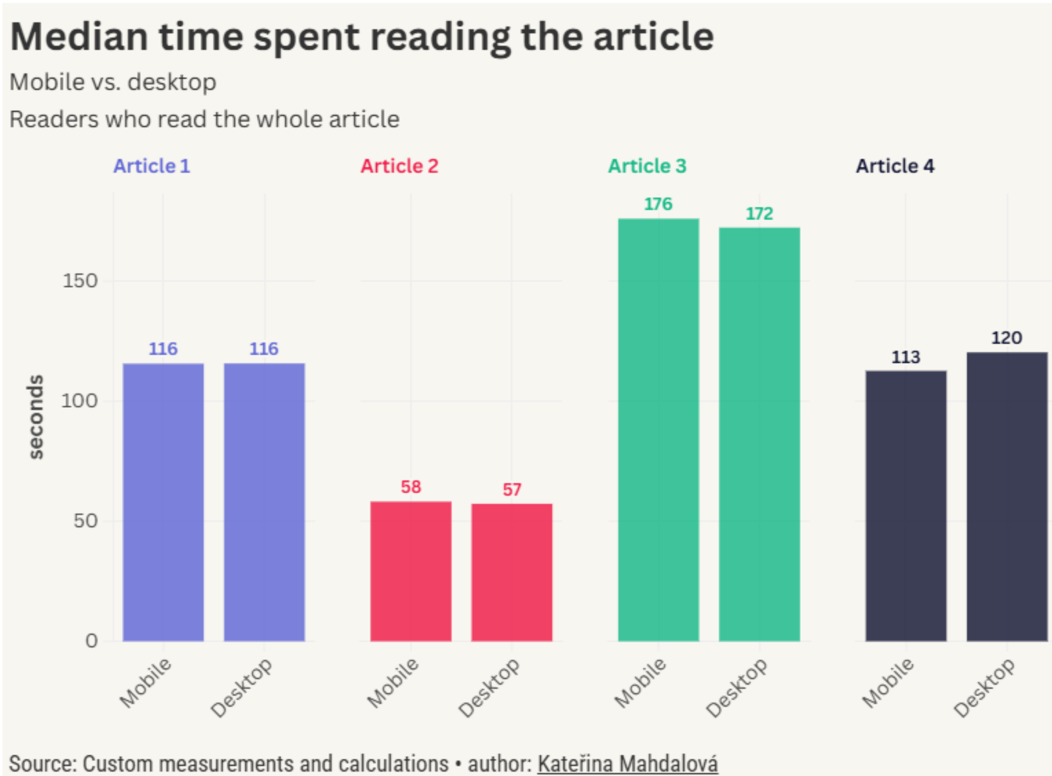
Graph 7: Visualization 4 in article 3: A notable outlier

This visualization broke the general trend, receiving significantly more engagement time than other visualizations in similar positions or with comparable complexity.

These findings provide support for hypothesis H4 and simultaneously point to a practical implication: when designing an article, it may be advantageous to place the most important visualizations closer to the beginning of the text, where they have a higher chance of not only being displayed (see 4.1) but also holding attention longer. At the same time, however, the existence of exceptions such as visualization 4 in Article 3 suggests that attractive content can overcome disadvantages stemming from position.

4.2.4 Influence of Device Type

Hypothesis H3 assumed that users on mobile devices would exhibit a lower level of exposure to visualizations than users on desktop. Surprisingly, data analysis did not provide support for this hypothesis. In all four articles, the median exposure time was very similar between mobile and desktop users, without statistically significant differences.



Graph 8: Median reading time for mobile vs. desktop readers, readers who read the whole article

A comparison of median reading times for mobile and desktop readers. The chart demonstrates minimal differences in engagement between the two groups across all four articles.

This result is interesting given the common assumption that mobile reading leads to accelerated scrolling and lower engagement. In our case, the data did not confirm this expectation. Possible explanations include:

1. Mobile users may scroll more slowly due to a smaller viewport, leading to longer visualization display times.
2. Visualizations may have been well optimized for mobile display, eliminating potential differences.
3. Mobile reader behavior may have changed in recent years and become more similar to desktop users.

This result therefore leads to the rejection of hypothesis H3. Based on the measured data, it cannot be confirmed that device type has a significant influence on the duration of visualization exposure.

Article	Article 1	Article 2	Article 3	Article 4
% of Mobile Readers	46.5%	45.3%	54.9%	53.1%

Table 7: Proportion of mobile readers by article

4.3 Typology of Readers and Level of Engagement

Based on a combination of scrolling patterns and visualization exposure time, several types of readers with varying levels of engagement can be identified. This section builds on the research question: *Are there different types of readers in terms of working with visualizations?* And simultaneously tests hypothesis H5, which assumes the existence of diverse reading strategies and interaction with visualization elements. The analysis revealed three consistently recurring groups across all four articles.

4.3.1 Inactive Readers (non-scrollers)

The first group consists of readers who initiated little or no scrolling after loading the page. In the analyzed data, they represented approximately 10–20% of all article visitors. These are likely users who left the page very early (so-called bounce) or performed only minimal interaction that was not captured by the measurement used.

In terms of working with visualizations, their engagement is very limited. However, if visualizations or other prominent graphic elements are placed entirely in the upper part of the page (above the fold), there is a possibility that the reader might at least notice them at a glance. This finding suggests that for editorial teams who want to reach even very passive audiences, it may be useful to place key visual elements right at the beginning of the article.

4.3.2 Readers with Partial Passage (partial readers)

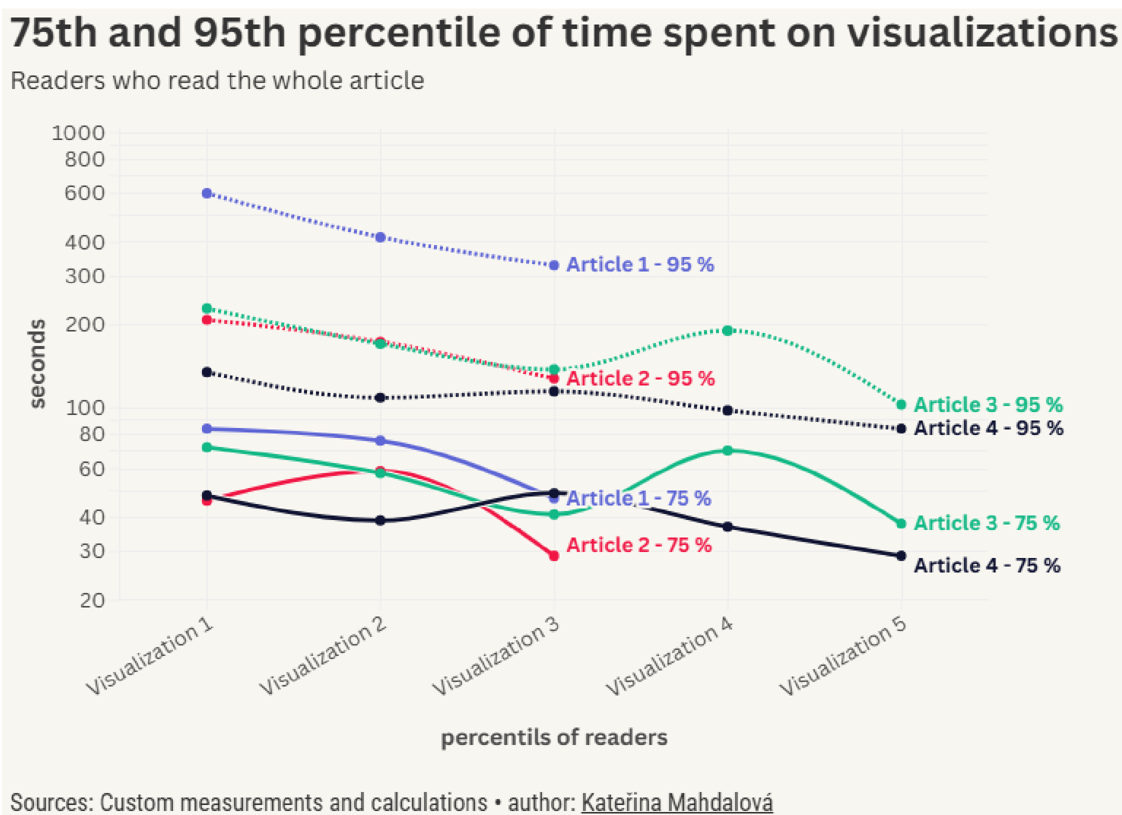
The second and largest group comprises readers who initiate reading the article but do not complete it. They typically display the first one or two visualizations, then leave the page. Their median exposure time is lower than for fully engaged readers but still indicates a certain level of attention paid to visualizations.

This group is important for understanding common user behavior in the online environment. The analysis suggests that the engagement of these readers may be strongly influenced by the position of visualizations and their perceived accessibility. If graphs are placed too low in the article, the probability of their display and processing decreases. The findings also suggest that many readers take away only information from the first part of the article, and therefore it may be effective to include key information in the introduction.

4.3.3 Readers Who Read to the End

The last group consists of readers who went through the entire article, i.e., scrolled all the way to the last visualization. In the analyzed articles, they formed 31.8% to 65.6% of the audience, depending on the length and structure of the given text. These readers are important for evaluating the effectiveness of the overall narrative and interactive content.

Within this group, a segment of so-called highly engaged readers can be identified, who spent significantly more time with visualizations than others – typically above the 75th percentile of the exposure time distribution (see table in section 4.2.1). For some visualizations, these readers stayed for several minutes. In the studied articles, they represented approximately 8-15% of all readers.



Graph 9: Highly Engaged Readers: 75% and 95% Percentile

The graph emphasizes the time spent with visualizations by the most engaged readers (above the 75th and 95th percentiles). More complex visualizations, like those in articles 1 and 3, tend to gain longer engagement, although overall engagement decreases for visualizations placed lower in articles.

This group represents an audience with the highest level of engagement, and its existence suggests that data journalism has the potential to reach readers who are willing to work intensively with graphic content – even in an online environment.

The analysis thus provides support for hypothesis H5. Within the studied articles, it was possible to consistently identify three basic types of readers who differ in their level of engagement, length of interaction, and depth of article passage. This typology may help to better understand the diversity of the data journalism audience and may be a starting point for further exploration of the relationship between article structure and reader behavior.

4.4 Summary of Key Findings

This final part of the chapter summarizes the main results of the analysis, as presented in the individual subchapters, and relates them to the four research questions and five tested hypotheses formulated in the methodological part (Chapter 3). The results are discussed both in terms of their empirical support and regarding possible limitations.

4.4.1 Answers to Research Questions

1) How do readers navigate through the article?

The analysis of scrolling data (see Chapter 4.1) suggests that readers typically leave articles gradually as they move lower in the text structure. As the distance from the beginning of the page increases, the proportion of readers who display that part decreases – on average by 12.5% for each additional 1000 pixels of position. This trend proved consistent across all four analyzed articles. The results are in line with the phenomenon of "reader attrition" described in the literature on web texts.

It should be emphasized, however, that the data measure technical presence on the page, not deeper cognitive processing. Actual understanding of the content, as well as the role of other factors (e.g., length of text between visualizations), were not examined in detail in this research.

2) Which visualizations do readers see? What is the duration of their interactions with them?

Visualization display (see Chapters 4.1 and 4.2.1) depends primarily on their position in the article. Readers typically see the first visualization in 80–90% of cases, while the last visualization is often seen by less than half of readers. When a visualization is shown to a reader, the median exposure time is typically tens of seconds. There are notable variations depending on the type of visualization.

It should be noted that the display metric captures only visual exposure, not the actual level of understanding or interaction with the data. The analysis also did not systematically consider possible influences of the contextual framework of the text or preceding content.

3) What influences whether a reader notices visualizations?

The results suggest that a significant predictor of exposure length and display rate is the vertical position of the visualization (see Chapter 4.2.3). To some extent, the type of visualization also plays a role (e.g., multi-slide formats may attract longer exposure), but this effect is not linear (see Chapter 4.2.2). Surprisingly, no significant difference was observed between mobile and desktop users, either in overall display length or in article completion rate (see Chapter 4.2.4).

These results suggest that although format and technology may play a certain role, the key factor appears to be the structural placement of the visualization. However, other factors cannot be excluded – e.g., visual contrast, narrative context, or reader's previous experience, which were not systematically examined in this study.

4) Are there different types of readers in terms of working with visualizations?

The analysis of scrolling patterns and time data (see Chapter 4.3) revealed that readers can be divided into three consistently occurring groups: 1) inactive readers who practically do not scroll; 2) readers with partial passage through the article; and 3) readers who read to the end. Within the last group, a subset of so-called highly engaged readers was identified, who spent significantly more time with visualizations than other readers.

This typology suggests that the audience for online journalism is not homogeneous, and that engagement with visualizations differs both quantitatively (exposure length) and qualitatively (article completion rate and manner of interaction with content).

4.4.2 Evaluation of Hypotheses

Hypothesis	Result	Summary
H1	Confirmed	The probability of visualization display systematically decreases with their position in the article, on average by 12.5% for every 1000 pixels.
H2	Partially confirmed	Multi-slide visualizations show a tendency to maintain attention longer than simple graphs, but the increase is not linear and is based on a limited number of cases.
H3	Rejected	No significant difference in exposure time was observed between mobile and desktop users, despite expected different behavior.
H4	Confirmed	Visualizations placed higher in the article are typically displayed longer, although there are exceptions likely based on content attractiveness.
H5	Confirmed	Different groups of readers with varied behavior can be identified in the data, including a segment of highly engaged users.

Table 8 Summary of Hypotheses Testing

Overview of the five research hypotheses and their empirical outcomes. The table highlights which assumptions about visualization exposure and user behavior were confirmed, partially confirmed, or rejected based on quantitative analysis of viewport data and reader typologies.

4.4.3 Context and Limitations

The findings of this study are consistent with previous knowledge in the field of web journalism, especially regarding the effect of content position and audience behavior diversity. At the same time, however, the results should be interpreted cautiously given the inherent limits of purely quantitative exposure measurement – the mere display of a visualization does not guarantee understanding, just as longer exposure time does not necessarily mean a higher level of engagement or better comprehension. The results should be interpreted with the awareness that the measurement took place in a natural online environment, where it was not possible to standardize all display conditions.

The analysis also revealed the presence of interesting exceptions – e.g., visualizations that gained a significantly higher level of attention despite their position deeper in the article (see 4.2.3). This suggests the potential influence of other, qualitatively conditioned factors, such as graphic form, connection to text, or specific content, which could be the subject of further research.

The results obtained based on reader data from four articles published on datovazurnalistika.cz were supplemented by a comparative case of the German article CO_2-Rechner published by Bayerischer Rundfunk (Bayer, 2024). This article was included in the analysis as an external reference point, with available aggregated data on readership and scrolling behavior showing similarities to trends observed in Czech texts. As in our primary data set, the Bayer article also showed a gradual decrease in the rate of visualization display depending on their position on the page, and overall reader behavior exhibited a similar degree of audience segmentation.

This similarity strengthens the relevance and potential transferability of findings even beyond the immediate language and editorial context and suggests that some patterns of attention to visualizations may be observable across different implementations of interactive data journalism published online.

A limitation of the study is the relatively small sample of analyzed articles (n=4) and visualizations (n=16), which restricts the possibilities for statistical generalization. Also, the absence of demographic data about readers makes it impossible to analyze whether interaction patterns differ among different age groups or other audience segments. These limitations should be considered in future research, which could combine the methodology used here with qualitative approaches and more detailed demographic data.

Despite the mentioned limits, this study provides an empirically based insight into a hitherto little-explored aspect of data journalism – namely, how readers encounter visualizations in the natural environment of online news, and what factors may influence the probability of this encounter.

5. Discussion

5.1 Interpretation of Results in Relation to Previous Research

This study's findings confirm, extend, or challenge existing knowledge in data journalism, information visualization, and reader behavior research. In this section, they are discussed both in relation to the formulated research questions and hypotheses, and in the context of the methodology used for measurement. Attention is also paid to areas where current literature does not offer direct support – cases that may inspire further research.

5.1.1 Measurement Context

All results of this study are derived from the analysis of actual reader behavior on websites, measured using scrolling data and visualization exposure duration. The methodology used (see Chapter 3) allowed for detailed monitoring of how long a given visualization was on the reader's screen, but not direct measurement of cognitive engagement or content comprehension. The research did not include collection of demographic data, user preferences, or data on connection quality or device specifications. The results thus reflect behavior in a natural online environment, but without the ability to control all potentially relevant factors.

This ecological validity represents both a strength of the study (realism of data) and its inherent limitation (inability to isolate individual variables). As Hullman and Gelman (2021) point out, this trade-off relationship between ecological validity and variable control is characteristic of current research on visualizations "in the wild," outside laboratory conditions.

5.1.2 Interpretation of Results Regarding Research Questions and Scholarly Sources

The first research question (How do readers navigate through articles?) was answered following previous knowledge about the phenomenon known as reader attrition. The observed decrease in content visibility moving downward (H1) corresponds with earlier studies on online reading and the so-called F-pattern (Pernice, 2017; Bärtl et al., 2021). Our results provide quantitative support for the thesis that reader attention in the online environment necessarily decreases with increasing vertical depth of the article – even in the case of interactive visualizations. The quantification of this decrease (12.5% reader loss per 1000 pixels) contributes to a more precise understanding of this phenomenon.

The second question (How much time do readers spend with visualizations?) relates to hypothesis H2, which was partially confirmed. The data suggests that multi-slide visualizations can indeed maintain attention longer, but the increase is not proportional to their length or complexity. This result develops the approaches of authors such as Cairo (2016), who points out the necessity of designing visualizations with regard to the cognitive limits of the audience. Our findings thus contribute to the discussion on the effective scope of visualizations in the journalistic environment – the length of interactive content likely has its limits, beyond which the effect of attention does not significantly increase further.

In relation to the third question (What influences whether a reader notices visualizations?), the key role of position in the article (H4) was confirmed, while the hypothesis about the influence of device type (H3) was not supported by the data. The absence of significant differences between mobile and desktop users challenges existing theoretical assumptions about different patterns of cognitive information processing across device types. This finding contrasts with some previous studies (e.g., Weber et al., 2018; Lee et al., 2021), which pointed to accelerated content use on smaller screens and lower levels of interaction with more complex visual elements.

A possible explanation may be the convergence of user habits across devices due to the growing ubiquity of mobile technologies and adaptation of visualization formats for different screen sizes. As mobile devices become the primary means of accessing news content for an increasing portion of the population (Newman et al., 2023), new reading patterns may be forming that begin to resemble those on desktop devices – and conversely, desktop reading may be adopting some characteristics of mobile-oriented usage.

From a technological perspective, advanced responsive visualization designs that optimize display for specific device types may play a role. Contemporary visualization tools like Flourish (used in this study) offer automatic display optimization, which may eliminate some previous disadvantages of mobile devices. From a cognitive-ecological perspective, this result points to the adaptability of human perception in various technological contexts, suggesting that users have developed compensatory strategies that enable them to effectively process visual information even on smaller devices.

In this regard, our study brings an interesting insight: under certain conditions (e.g., good optimization for mobile devices, specific type of content or audience), mobile access does not necessarily mean lower engagement. This question remains open for further research, which should combine quantitative measurements with demographic and contextual variables.

The fourth question (Are there different types of readers?) was answered affirmatively and provides support for hypothesis H5. The segmentation of the audience into three main groups – inactive readers, partially engaged, and those who read to the end – corresponds with earlier concepts from the field of behavioral reading models (e.g., Pirolli & Card, 1999 and their theory of "information foraging"). A new element that our study brings is the identification of a group of "highly engaged readers," i.e., readers who paid extraordinary attention to visualizations. This subgroup suggests that even in the online media environment, there exists an audience with a significant interest in data content, which may be strategically important for data journalism.

5.1.3 Insights for Future Research

Some findings suggest new research directions that have not yet been systematically developed in the literature. For example, the identification of so-called positive deviations – visualizations that gained above-average display and exposure rates despite being placed deep in the article (see 4.2.3) – suggests the existence of other yet insufficiently explored factors influencing attention. These may include visual contrast, narrative framework, graphic originality, or the type of data presented. However, current literature lacks a comprehensive framework that would systematically quantify these influences.

This raises the question of whether we should pay more attention to other variables, such as demographic characteristics (age, education), frequency of reading data articles, information literacy, or the context in which news content is accessed and interpreted (e.g., environment, time, purpose).

5.2 Implications for Data Journalism Practice

Although the primary aim of this study was to empirically examine the relationship between reader behavior and the characteristics of data visualizations, some findings may have – with an awareness of certain methodological limitations – relevance for the practice of editorial planning, article design, and visualization design. This part of the discussion offers several possible interpretative directions that stem from previous analyses but also reflect the methodological limitations of the research (see 5.3).

5.2.1 Placement of Visualizations within the Article

The results indicate that a significant factor influencing both the probability of display, and the duration of exposure is the position of the visualization within the article. Visualizations placed closer to the beginning achieved higher display rates, and readers also typically stayed with them longer. This suggests that it may be advantageous to place important or interpretatively demanding visualizations higher in the article, where they have a higher chance of being displayed at all. At the same time, however, it should be noted that this interpretation does not consider potential negative effects of cognitive overload at the beginning of the article or the role of gradually building a narrative within the text.

An interesting consideration is the strategy for addressing so-called "non-scrollers," i.e., readers who leave the article very soon after loading. These users can only be reached by visual elements visible without the need for scrolling. If the editorial goal is to reach this group as well, it may be useful to place key visual elements in the "above the fold" area (i.e., the part visible immediately after loading the page).

5.2.2 Complexity and Scope of Visualizations

Findings regarding exposure duration suggest that more complex visualizations (e.g., multi-slide) may be viewed longer, but this effect is not linear and likely has its limits. Visualizations with numerous slides or interactive elements do not attract proportionally more time than simple graphs – often the difference is only in the order of tens of seconds. From this, it can be inferred that the scope and complexity of a visualization have practical boundaries, and the maximum length of engagement is not determined solely by format, but also by other factors, such as the level of comprehensibility, relevance of content, or appropriate contextual placement.

These findings can inspire reflections on finding a balance between the effort to present complex data and ensuring clarity. In some cases, a more concise visualization with a clear context and intuitive design may reach a wider audience than an extensive visual block requiring more complex navigation or interpretation. However, it should be emphasized that this interpretation is not based on direct measurement of cognitive processing or understanding, which represents an area for follow-up research.

5.2.3 Working with Different Types of Readers

Audience segmentation (see 4.3) showed that readers approach articles in markedly diverse ways – from those who leave the article almost immediately after loading, to intensely engaged readers who devote several minutes to individual visualizations. This diversity presents both a challenge and an opportunity for editorial design. A potential solution could be a stratified content structure that offers different entry points to different groups – for example, a brief overview of key findings at the beginning for readers with limited attention and more detailed analysis for those who continue deeper into the article.

However, it should be noted that within this study, it was not possible to systematically examine what factors lead to a reader being categorized into a certain group: it may involve differences in information literacy, motivation, habits in online reading, or thematic relevance. For this reason, concrete recommendations should be approached cautiously, with an awareness of their preliminary nature.

5.2.4 Article Length and Progressive Loss of Readers

The results of scrolling analysis showed that with increasing article length, there is a regular decrease in readers, which is also reflected in lower exposure to later visualizations. This phenomenon was observed across all analyzed texts, being most pronounced in the case of longer articles with a greater number of visualization blocks.

These findings may lead to considerations about the appropriate structuring of data-rich content. In some cases, it might be more effective to divide such content into shorter, independently readable units, for example in the form of a series of articles. Such division could contribute to maintaining reader’s attention while allowing for more thorough development of individual visualizations without the risk that the reader will not reach them at all. However, it should be added that the effectiveness of such a strategy may depend on the specific context of the topic, the target audience, and the technical solution of the publishing platform.

5.3 Research Limitations and Suggestions for Future Directions

This study represents an empirical attempt to gain a deeper understanding of how readers of online journalistic texts work with data visualizations – when they are exposed to them, how long they stay with them, and how these interactions differ depending on various factors. The results suggest that the reader's relationship to visualizations is neither unambiguous nor homogeneous but is influenced by a range of variables – from technical parameters through editorial design to individual user characteristics.

Despite the scope of the analyzed data, the results of this study must be interpreted with an awareness of certain methodological and contextual limitations that necessarily affect their generalizability and depth of interpretation.

5.3.1 Methodological Limitations and Interpretative Framework

The measurement was based on automatic monitoring of user behavior in the natural environment of four online articles. The obtained data allowed determination of whether and how long a given visualization was displayed on the user's screen, but did not provide information on how it was interpreted, whether it was understood, or what value the reader attached to it. Cognitive processing and value reception of the visualization thus remain outside the scope of this research.

The study did not include collection of demographic data on readers, which made it impossible to consider factors such as age, education, professional focus, previous experience with data journalism, or information literacy. Although the volume of interaction data itself was quite extensive, the aggregated nature of the results limits the possibilities for interpretation at the level of individuals or specific reading scenarios.

5.3.2 Sample Size and External Validity

A significant limitation of this study is its reliance on a relatively small sample of four articles containing 16 visualizations. While the total number of user interactions recorded was substantial (over 20,000 events), the limited variety of content types, visualization formats, and thematic areas restricts the generalizability of findings to broader journalistic contexts.

The modest sample size impacts the external validity of the study in several important ways. First, it limits our ability to generalize across different types of news content – for example, we cannot confidently claim that the observed patterns would persist in breaking news, long-form investigative journalism, or specialized financial reporting. Second, the small number of visualization types represented in the sample (particularly for multi-slide formats, where n=5) constrains the statistical power of comparisons between visualization categories. Third, the restricted sample may not adequately capture seasonal variations in reader behavior or responses to different news cycles.

While the consistency of findings across the four analyzed articles suggests some robustness in the observed patterns, particularly regarding the position effect and reader typology, these findings should be considered preliminary and requiring validation through more extensive studies. This limitation, however, does not invalidate the methodological approach or the empirical observations made within the given sample – it rather situates them as a foundation for more comprehensive future research.

Future replication studies would benefit from significantly expanded samples along several dimensions:

- Greater diversity of publishing platforms beyond specialized data journalism sites
- Wider range of content types and thematic areas
- Longitudinal design capturing temporal variations in reader behavior
- Cross-national comparison to assess cultural factors in visualization reception
- Systematic variation in visualization types while controlling for content

Such expanded sampling would allow for more robust statistical analysis and potentially uncover interaction effects that may be obscured in the current study due to the limited sample size.

5.3.3 Selection Limitations and Publication Context

The analyzed articles were published on a single web portal (datovazurnalistika.cz) and exhibited similar genre, visual, and technical characteristics. Although one external article was included in the analysis as a comparative case (Bayer, 2024), the results cannot be uncritically generalized to significantly different editorial environments, languages, audiences, or cultural contexts. It cannot be ruled out that the audience of the specialized website datovazurnalistika.cz has specific motivation, above-average data literacy, and therefore may exhibit different behavior patterns compared to the general population of online news consumers.

5.3.4 Research Impulses for the Future

Despite the mentioned limitations, the study opened several significant research directions that could be further developed:

Complement with a Qualitative Approach: The research provided knowledge about what readers do – but substantially less about why they do so. Interviews, questionnaire surveys, or so-called think-aloud protocols, i.e., monitoring what readers verbalize during reading, could contribute to a deeper understanding of reader strategies. Such an approach could help interpret, for example, cases of short exposure to complex visualizations: do they mean rejection? Lack of understanding? Or, conversely, quick but effective processing of information?

Demographic Segmentation: If it were possible to identify age, education, or other reader characteristics, it would be possible to examine whether and how patterns of interaction with visualizations differ among various audience segments. Particular attention could be given to, for example, the relationship of younger generations with mobile visualizations or differences between professional groups with varying levels of data literacy.

Experimental Testing of Design Elements: The study suggested that some visualizations work better than others, but it was not possible to systematically determine which specific aspects of design (e.g., vertical vs. horizontal arrangement, number of interactive elements, color scheme) are primarily responsible for this. In future research, it would be useful to design controlled A/B tests that would isolate and quantify these influences.

Cross-Genre Comparisons: While this work focused on analyzing articles with a significant data component, future research could compare visualizations across various journalistic genres – from investigative articles through economic analyses to sports statistics or cultural reviews. It would also be interesting to monitor potential differences in user behavior across different platforms – for example, between content published primarily on social media and traditional web articles.

Longitudinal Monitoring: This study measured behavior during one-time exposure to articles. For a more comprehensive understanding of the audience's relationship to data journalism, it would be beneficial to monitor the development of interaction patterns over time – for example, repeated visits to the same content, returns to specific visualizations, or patterns of sharing data articles.

The chosen research approach demonstrated that even without direct questioning of readers or invasive intervention in their natural environment, valuable empirical insights can be gained about how data visualizations in online journalism are actually consumed. This study can thus be perceived as an initial step towards more comprehensive research on visualizations in a media context –

research whose importance is growing with the increasing role of data in contemporary journalistic communication.

6. Conclusion

This master's thesis focused on empirically examining reader interaction with data visualizations in the natural environment of online news. The main goal was to understand the basic prerequisites for the effectiveness of visualizations in a journalistic context – specifically to determine which visualizations readers actually see, how long they spend with them, and what factors influence these parameters. "The research results provide empirically based insight into the still under-explored area of how data visualizations are engaged with in real media operation conditions.

6.1 Summary of Research Findings

As part of the research, a total of 16 data visualizations in four articles on the datovazurnalistika.cz portal were analyzed using a specialized measurement tool that recorded scrolling behavior and exposure to visualizations among real readers. The analysis revealed several significant patterns:

1) Visualization Position is a Significant Predictor of Exposure

The vertical placement of a visualization within an article represents one of the most significant factors influencing whether it will even be displayed on the reader's screen. With every 1000 pixels downward, the probability of display decreases by an average of 12.5%. While the first visualizations in articles typically achieved an 80-90% display rate, the last visualization was often displayed to only 30-50% of readers. This relatively linear decline was consistent across all examined articles, providing support for hypothesis H1 about the systematic decrease in exposure with increasing depth in the article.

2) Visualization Complexity May Affect Exposure Duration

More complex visualizations (e.g., multi-slide formats) showed a tendency to maintain reader attention longer than simple single-image graphs, partially supporting hypothesis H2. However, this relationship is not linear – adding more slides or interactive elements does not bring a proportional increase in exposure time. From a practical perspective, this suggests that there is a certain saturation point beyond which further increases in complexity may not bring a corresponding increase in time spent on visualization.

3) Device Type Has No Significant Impact on Exposure Rate

A surprising finding was the rejection of hypothesis H3, which assumed that mobile users would show a lower rate of exposure to visualizations than desktop users. Despite the common assumption about accelerated scrolling and lower engagement on mobile devices, the data did not show a statistically significant difference in exposure time between these two groups. The proportion of mobile users comprised 45.3-54.9% of the total audience. This result contrasts with some previous studies (e.g., Weber et al., 2018) and raises the question of whether current optimizations of visualizations for mobile devices are leading to the equalization of previously observed differences.

4) Visualizations Higher in the Article Receive More Attention Time

The data provided support for hypothesis H4, that visualizations placed higher in the article receive, on average, more attention time from readers than visualizations placed lower. However, this relationship is not absolute – some visualizations managed to gain above-average attention even in lower positions if they were content-attractive or formally distinct. An exceptional example was visualization 4 in article 3, which, despite its position in the second half of the article, achieved an extraordinarily high exposure time, suggesting that quality content can to some extent overcome the disadvantage stemming from position.

5) Different Types of Readers Exist with Distinct Behavior Patterns

The analysis confirmed hypothesis H5 about the existence of different types of readers in terms of their interaction with visualizations. Three consistent groups were identified:

- **Inactive readers** (non-scrollers) comprising 10-20% of the audience, who performed virtually no scrolling after loading the page.
- **Readers with partial passage** through the article, who formed the largest group and typically displayed the first one or two visualizations before leaving the page.
- **Readers who completed the entire article**, i.e., 31.8-64.3% of the audience depending on the length and structure of the text.

Particularly significant was the identification of a subgroup of so-called "highly engaged readers" (approximately 8-15% of the total audience), who devoted significantly more time to visualizations than average readers – with some visualizations, they stayed for several minutes. This group represents an audience with the highest level of engagement and suggests the existence of a segment of readers who are willing to work intensively with data content even in an online environment.

6.2 Comparative Context and External Validation

Interestingly, the results obtained from the Czech sample showed considerable similarity to data from the German article CO_2-Rechner published by Bayerischer Rundfunk (Bayer, 2024), which was included as a comparative case. Although it was not possible to implement measurements using the same technique, the available aggregated readership data showed very similar patterns of readership retention depending on content position. This similarity strengthens the potential transferability of findings even beyond the immediate language and editorial context and suggests that some identified patterns may have broader validity across different media environments.

Originally, the research was planned on one of the most-read Czech news websites, Seznam Zprávy, but due to changes in editorial policy in autumn 2024, this could not be implemented. In the end, however, it turned out that this limitation probably did not have a significant impact on the validity of the results, as similar research conducted by colleagues from Bayerischer Rundfunk led to comparable findings.

6.3 Methodological Limitations and Their Reflection

When interpreting the results, several methodological limitations must be considered. The research deliberately focused on measuring exposure – that is, whether and how long a visualization was displayed on the screen – not on cognitive processing or understanding. This approach has both strengths (non-invasiveness, ecological validity) and limitations (absence of data on actual understanding of content). In this work, we do not attempt to obscure this limitation, but rather explicitly acknowledge it as a conscious methodological choice that enables the collection of empirical insights into real user behavior without disrupting the natural context in which news content is accessed and engaged with.

Another limitation is the absence of demographic and contextual variables. Although the measurement captured device type (mobile vs. desktop), it was not possible to determine age, education, information literacy, or other reader characteristics that could influence their interaction with visualizations. It was also not possible to measure interactions within the visualizations themselves due to technical limitations given by the use of iframe elements and the Flourish platform.

The sample set of four articles and 16 visualizations from one website represents another factor that may limit broader generalizability of the findings. Although the selection was deliberately constructed to include different types of visualizations and thematic contexts, it cannot be ruled out that the audience of the specialized website datovazurnalistika.cz may differ in some aspects from the broader population of news consumers.

6.4 Theoretical and Practical Implications

Despite the limitations mentioned, the study brings several significant insights for both theory and practice of data journalism.

From a theoretical perspective, the results contribute to the empirical verification of concepts such as:

- **Reader attrition –** linear decrease in readership with increasing depth in the article
- **Visualization blindness** – the phenomenon of overlooking visualizations, especially those placed lower in the text
- **Engagement segments** – the existence of different groups of readers with distinctive interaction patterns

The study also problematizes some existing assumptions, especially those about differences between mobile and desktop users, and opens space for reconsidering positions on the effectiveness of complex interactive visualizations in the context of real-world news use.

From a practical perspective, several potential recommendations for editorial practice can be derived from the results:

- **Strategic placement of visualizations** – key or complex visualizations should generally be placed in the upper part of the article, where they have a significantly higher chance of being displayed.
- **Optimization of complexity** – overly complex multi-slide visualizations may not always bring a proportional increase in time spent with content; in some cases, a more concise but clearer presentation may be more effective.
- **Segmented approach to the audience** – content could be structured to offer different entry points for different groups of readers, from quick summaries for less active readers to more complex analyses for highly engaged audiences.

- **Equivalent approach to mobile and desktop users** – the results suggest that it may not be necessary to assume fundamentally different interaction patterns between these two groups if visualizations are adequately optimized for mobile display.

6.5 Directions for Future Research

This study opens several promising directions for future research in the area of interaction with data visualizations in journalism:

- **Qualitative complement** – Interviews or think-aloud protocols could clarify why readers devote different amounts of time to visualizations and how they interpret them, which would help connect quantitative exposure data with qualitative aspects of understanding.
- **Demographic segmentation** – Linking exposure data with demographic characteristics would allow analysis of how interaction with visualizations differs according to age, education, or level of media and data literacy.
- **Experimental design** – Systematic A/B testing of different types and placements of visualizations could isolate the influence of individual factors on reader behavior and provide more precise answers to questions regarding optimal design.
- **Longitudinal studies** – Monitoring changes in interaction patterns over time would allow capturing potential development of visualization literacy or audience adaptation to new formats of data presentation.
- **Comparison of editorial environments** – Replication of research in various media contexts would help determine which behavior patterns have more universal validity and which are specific to a certain type of media or audience.

Particularly valuable could be research focused on exceptional cases, such as visualization 4 in article 3, which, despite its position, managed to attract above-average attention. Understanding the factors behind such "success" could bring deeper insight into the mechanisms of reader attention in the context of data journalism.

Final Assessment

This master's thesis responded to an identified research gap – the lack of empirical knowledge about how readers actually interact with data visualizations in the natural environment of online news. Although it focused primarily on basic exposure parameters (visibility and time), not on complex cognitive processing of content, it brought several significant findings that can inform both theoretical considerations about visualization effectiveness and practical editorial decisions.

The main contribution of the study is the empirical documentation that the mere presence of a visualization in an article does not guarantee that it will be displayed by the reader, let alone thoroughly studied. Position within the article proves to be a key factor that significantly affects the probability and duration of display, with surprising consistency across the analyzed articles. At the same time, the study revealed that the audience for data journalism is not homogeneous but includes markedly different groups from completely passive to highly engaged readers.

In a time of growing importance of data journalism and visual forms of communication, this study offers empirically based insight into the real conditions of data visualization reception. Although it does not answer all questions related to the effectiveness of visualizations, it provides a solid foundation for further research and stimulates critical reconsideration of some established ideas about how readers work with visual content in the digital environment.

Bibliography

Albers, J. (2006). *Interaction of color* (Vol. 6). New Haven: Yale University Press.

Anderson, C. W. (2018). *Apostles of certainty: Data journalism and the politics of doubt.* Oxford University Press.

Bach, B., Dragicevic, P., Archambault, D., Hurter, C., & Carpendale, S. (2014). A Review of Temporal Data Visualizations Based on Space-Time Cube Operations. In *Eurographics Conference on Visualization (EuroVis 2014)* (pp. 23-41)

Bach, B., Riche, N. H., Carpendale, S., et al. (2018). The emerging role of visualization literacy in data-driven journalism. *Computer Graphics Forum*, 37(3), 417–434. https://doi.org/10.1111/cgf.13412

Bach, B., Wang, Z., Farinella, M., Murray-Rust, D., & Henry Riche, N. (2018). Design patterns for data comics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Paper 38, pp. 1–12). ACM. https://doi.org/10.1145/3173574.3173612

Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191. https://doi.org/10.1177/0165551508095781

Bayer, C. (2024). Interactives [Conference presentation]. Dataharvest Conference, Arena for Journalism in Europe. https://dataharvest.eu/wp-content/uploads/2024/06/interactives-Constanze.pdf (Retrieved June 12, 2024)

Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yeh, C. S., Borkin, D., & Oliva, A. (2015). Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 519-528. https://doi.org/10.1109/TVCG.2015.2467732

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J.-D. (2014). A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1963–1972. https://doi.org/10.1109/TVCG.2014.2346984

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. New Riders.

Cairo, A. (2019). *How charts lie: Getting smarter about visual information*. W. W. Norton & Company.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. Morgan Kaufmann.

Castells, M. (2010). *The rise of the network society* (2nd ed.). Wiley-Blackwell.

Cleveland, W. S., & 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. https://doi.org/10.1080/01621459.1984.10478080

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–185. https://doi.org/10.1017/S0140525X01003922

Dabbish, L., Mark, G., & González, V. M. (2011). Why do I keep interrupting myself? Environment, habit, and self-interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3127–3130). ACM. https://doi.org/10.1145/1978942.1979405

D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

Elmqvist, N., & Yi, J. S. (2015). Patterns for visualization evaluation. *Information Visualization*, 14(3), 250–269. https://doi.org/10.1177/1473871613513228

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 20(5), 325–344. https://doi.org/10.1080/01972240490507974

Few, S. (2013). *Information dashboard design: Displaying data for at-a-glance monitoring*. Analytics Press.

Fink, K., & Anderson, C. W. (2015). Data journalism in the United States: Beyond the usual suspects. *Journalism Studies*, 16(4), 467–481. https://doi.org/10.1080/1461670X.2014.939852

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3), 110–161. https://doi.org/10.1177/15291006211051956

Fry, B. (2004). *Computational information design* [Doctoral dissertation, Massachusetts Institute of Technology].

Gitelman, L. (Ed.). (2013). *Raw data is an oxymoron*. MIT Press.

Google. (2024a). Firebase Realtime Database. https://firebase.google.com/docs/database (Retrieved May 5, 2024)

Google. (2024b). Firebase Data Processing and Security Terms. https://firebase.google.com/terms/data-processing-terms (Retrieved April 15, 2024)

Grammel, L., Tory, M., & Storey, M. A. (2010). How information visualization novices construct visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 943–952. https://doi.org/10.1109/TVCG.2010.164

Grinberg, N. (2018). Identifying modes of user engagement with online news and their relationship to information gain in text. In *Proceedings of the 2018 World Wide Web Conference* (pp. 1775–1784). ACM. https://doi.org/10.1145/3178876.3186180

Grusky, M., Naaman, M., & Artzi, Y. (2018). NEWSROOM: A dataset of 1.3 million summaries with diverse extractive strategies. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 708–719). Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1065

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 203-212). ACM. https://doi.org/10.1145/1753326.1753357

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4), 45–54. https://doi.org/10.1145/2133806.2133821

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3(3), 446–474. https://doi.org/10.1111/j.1756-8765.2011.01150.x

Heravi, B. R., & Lorenz, M. (2020). Data journalism practices globally: Skills, education, opportunities, and values. *Journalism and Media*, 1(1), 26-40. https://doi.org/10.3390/journalmedia1010003

Holmqvist, K., & Wartenberg, C. (2005). The role of local design factors for newspaper reading behaviour: An eye tracking perspective. *Lund University Cognitive Studies*, 127. https://www.lucs.lu.se/fileadmin/user_upload/project/lucs/LUCS-pub/LUCS-127.pdf

Hullman, J., & Gelman, A. (2021a). Designing for interactive exploratory data analysis requires statistical language support. *Harvard Data Science Review*, 3(2). https://hdsr.mitpress.mit.edu/pub/w075glo6/release/3

Hullman, J., & Gelman, A. (2021b). To design interfaces for exploratory data analysis, we need theories of graphical inference. *arXiv*. https://arxiv.org/abs/2104.02015

Hutchins, E. (1995). *Cognition in the wild*. MIT Press.

Hutchins, E. (2010). Cognitive ecology. *Topics in Cognitive Science*, 2(4), 705–715. https://doi.org/10.1111/j.1756-8765.2010.01089.x

Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication & Society*, 19(6), 715–735. https://doi.org/10.1080/1369118X.2016.1153126

Kayongo, P., Sun, G., Hartline, J., & Hullman, J. (2022). Visualization equilibrium. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 465-474. https://doi.org/10.1109/TVCG.2021.3114842

Kim, Y., & Heer, J. (2018). Assessing effects of task and data distribution on the effectiveness of visual encodings. *Computer Graphics Forum*, 37(3), 157–167. https://doi.org/10.1111/cgf.13409

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

Kostkova, P. (2021). Disease surveillance data sharing for public health: The next ethical frontiers. *Life Sciences, Society and Policy*, 17(1), 1-7. https://doi.org/10.1186/s40504-018-0078-x

Laban, P., & Hearst, M. (2017). newsLens: Building and visualizing long-ranging news stories. In *Proceedings of the Events and Stories in the News Workshop* (pp. 1–9). Association for Computational Linguistics. https://doi.org/10.18653/v1/W17-2701

Lagun, D., & Lalmas, M. (2016). Understanding user attention and engagement in online news reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (pp. 113-122). ACM. https://doi.org/10.1145/2835776.2835833

Lankow, J., Ritchie, J., & Crooks, R. (2012). *Infographics: The power of visual storytelling*. John Wiley & Sons.

Lee, B., Dachselt, R., Isenberg, P., & Choe, E. K. (Eds.). (2021). *Mobile data visualization*. CRC Press.

Leroy, S. (2009). Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes*, 109(2), 168–181. https://doi.org/10.1016/j.obhdp.2009.04.002

Levy, D. M. (2008). Information overload. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems*. Springer. https://doi.org/10.1002/9780470281819.ch20

Liu, Y., & Heer, J. (2018). An empirical assessment of quantitative colormaps. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Paper 124, pp. 1–12). ACM. https://doi.org/10.1145/3173574.3174172

Mörth, E., Bruckner, S., & Smit, N. N. (2022). ScrollyVis: Interactive visual authoring of guided dynamic narratives for scientific scrollytelling. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 1101–1112. https://doi.org/10.1109/TVCG.2022.3205769

Murray, J. H. (2013). *Inventing the medium: Principles of interaction design as a cultural practice*. MIT Press.

Pernice, K. (2010). *Eyetracking web usability*. New Riders.

Pernice, K., Nielsen, J. (2017). F-shaped pattern of reading on the web: Misunderstood, but still relevant (even on mobile). *Nielsen Norman Group*. https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content/

Oesch, J., Renner, A., & Roth, M. (2022). Scrolling into the newsroom: A vocabulary for scrollytelling techniques in visual online articles. *Information Design Journal*, 27(1), 102–114. https://doi.org/10.1075/idj.22005.oes

Park, C. S. (2019). Does too much news on social media discourage news seeking? Mediating role of news efficacy between perceived news overload and news avoidance on social media. *Social Media + Society*, 5(3), 1-12. https://doi.org/10.1177/2056305119872956

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Schmidt, A. M., & DeShon, R. P. (2010). The moderating effects of performance ambiguity on the relationship between self-efficacy and performance. *Journal of Applied Psychology*, 95(3), 572–581. https://doi.org/10.1037/a0018289

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages* (pp. 336–343). IEEE. https://doi.org/10.1109/VL.1996.545307

Speier, C. (2006). The influence of information presentation formats on complex task decision-making performance. *International Journal of Human-Computer Studies*, 64(11), 1115–1131. https://doi.org/10.1016/j.ijhcs.2006.06.007

Stoiber, C., Walchshofer, C., Pohl, M., Potzmann, B., Grassinger, F., Stitz, H., Streit, M., & Aigner, W. (2022). Comparative evaluations of visualization onboarding methods. *Visual Informatics*, 6(4), 34–50. https://doi.org/10.1016/j.visinf.2022.07.001

Toffler, A. (1970). *Future shock*. Random House.

Tseng, C., Quadri, G. J., Wang, Z., & Szafir, D. A. (2023). Measuring categorical perception in color-coded scatterplots. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Article 824, pp. 1–14). ACM. https://doi.org/10.1145/3544548.3581416

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Graphics Press.

Vessey, I., & Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2(1), 63–84. https://doi.org/10.1287/isre.2.1.63

Ware, C. (2004). *Information visualization: Perception for design* (2nd ed.). Morgan Kaufmann.

Weber, W., & Rall, H. (2012). Data visualization in online journalism and its implications for the production process. In *Proceedings of the 16th International Conference on Information Visualisation* (pp. 349-356). IEEE. https://doi.org/10.1109/IV.2012.65

Xiong, C., Ceja, C. R., Ludwig, C. J., & Franconeri, S. (2019). Biased average position estimates in line and bar graphs: Underestimation, overestimation, and perceptual pull. *IEEE Transactions on*

Visualization and Computer Graphics, 26(1), 301-310. https://doi.org/10.1109/TVCG.2019.2934400

Yi, J. S., Kang, Y. A., Stasko, J. T., & Jacko, J. A. (2007). Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1224–1231. https://doi.org/10.1109/TVCG.2007.70515

Ziemkiewicz, C., & Kosara, R. (2010). Implied dynamics in information visualization. In *Proceedings of the 2010 International Conference on Advanced Visual Interfaces* (pp. 215–222). ACM. https://doi.org/10.1145/1842993.1843031

List of Charts, Figures and Tables

Annexes

Annex 1 – Technical Methodology of Data Collection

Methodological Implications

This approach aligns with best practices for passive attention measurement in web-based visual journalism. It offers a non-intrusive, scalable, and ethically unobtrusive way to collect behavioral data in live environments.

However, findings must be interpreted in light of the proxy nature of the measurement. Visibility is a necessary condition for engagement, but not a sufficient one. A reader may glance at a chart without understanding it, or leave it on screen while distracted. Likewise, some users may rapidly scroll but still cognitively engage.

Differences in Implementation

For the first article, data collection was limited to a random sample of approximately 1 out of 20 of readers to minimize the risk of technical issues. However, subsequent experience showed that this precaution was unnecessary. As a result, data for all readers were collected for the remaining articles.

This change had no practical impact on the results of the analyses, as the sample size for the first article was still sufficiently large (over 4,000 cases) and the selection of readers was entirely random.

Example code

Example of HTML and javascript used for the second visualization (here "chart 2") on an article page:

```
<!-- 1st measured item: directly above the embedded visualization -->
<span id="chart-2-top" data-track="chart-2-top"></span>

<!-- 2nd measured item: the embedded visualization -->
<iframe src='https://flo.uri.sh/visualisation/18546352/embed' title='Interactive or visual
content' class='flourish-embed-iframe' frameborder='0' scrolling='no' sandbox='allow-same-
origin allow-forms allow-scripts allow-downloads allow-popups allow-popups-to-escape-
sandbox allow-top-navigation-by-user-activation' id='chart-2-iframe' data-track='chart-2-
```

```
iframe'></iframe>

<!-- 3rd measured item: directly below the embedded visualization -->
<span id="chart-2-bottom" data-track="chart-2-bottom"></span>

<!-- Firebase App (the core Firebase SDK) -->
<script src="https://www.gstatic.com/firebasejs/8.10.0/firebase-app.js"></script>
<!-- Firebase Firestore SDK -->
<script src="https://www.gstatic.com/firebasejs/8.10.0/firebase-firestore.js"></script>

<script>
    // Firebase configuration
    const firebaseConfig = {
     apiKey: "API_KEY",
     authDomain: "PROJECT_ID.firebaseapp.com",
     projectId: "PROJECT_ID",
     storageBucket: "PROJECT_ID.appspot.com",
     messagingSenderId: "MESSAGING_SENDER_ID",
     appId: "APP_ID",
     measurementId: "MEASUREMENT_ID"
    };

    // Initialize Firebase
    firebase.initializeApp(firebaseConfig);

    // Initialize Firestore
    var db = firebase.firestore();

    // Function to get or create a new session-specific user ID
```

```
// example: user_hl00tc0vb
function getUserId() {
if (!sessionStorage.getItem('userId')) {
  sessionStorage.setItem('userId', 'user_' + Math.random().toString(36).substr(2, 9));
}
return sessionStorage.getItem('userId');
}

// Function to store an event in Firestore
function storeEvent(eventType, parameter = null) {
var datetime = new Date().toISOString();
var userId = getUserId();
var url = window.location.pathname;

// Information to be stored about each event
// example:
// {
//   datetime: '2024-06-12T10:54:51.953Z',
//   userId: 'user_hl00tc0vb',
//   url: '/EXAMPLE_PATH',
//   event: 'entered-visible',
//   parameter: 'chart-2-top'
// }
var event = {
  datetime: datetime,
  userId: userId,
  url: url,
  event: eventType,
  parameter: parameter
```

```
};

// Element visibility events
function trackVisibility(elementId) {
var element = document.getElementById(elementId);
if (!element) return;

var observer = new IntersectionObserver(function(entries) {
entries.forEach(function(entry) {
if (entry.isIntersecting) {
 storeEvent('entered-visible', elementId);
} else {
 storeEvent('exited-visible', elementId);
}
 });
});

observer.observe(element);
}

// Page loaded event
// NOTE: not worked properly
document.addEventListener('DOMContentLoaded', function() {
 storeEvent('page-loaded');
});

// Page unloaded event
// NOTE: not worked properly
window.addEventListener('beforeunload', function() {
```

```
    storeEvent('page-unloaded');
  });

  // Select all elements with the 'data-track' attribute
  var elements = document.querySelectorAll('[data-track]');

  // Loop over the elements
  for (var i = 0; i < elements.length; i++) {
  // Get the value of the 'data-track' attribute
  var trackValue = elements[i].getAttribute('data-track');

  // Call trackVisibility with the track value
  trackVisibility(trackValue);
  }

  // Device detection
  var isMobile = /Android|webOS|iPhone|iPad|iPod|BlackBerry|IEMobile|Opera
Mini/i.test(navigator.userAgent);
  if (isMobile) {
    var device = 'mobile'
  } else {
    var device = 'desktop'
  }
  storeEvent('device', device);
</script>
```

Specific Consideration: Article 5

Article 5 was measured differently from the others. This data was not tracked using the JavaScript method described above.

Instead, the data was provided directly by BR (Bayerischer Rundfunk). The author received the dataset from Constanze Bayer, who presented it at the Dataharvest 2024 Conference in Belgium (Bayer, 2024). The original source article can be found here: https://interaktiv.br.de/co2-rechner/.

Because the tracking setup was not embedded in BR's server or codebase, standard visibility-based data collection was impossible. This difference should be considered when comparing metrics across visualizations, particularly when interpreting Article 5 in relation to the others.

Annex 2 – Overview of Visualizations

Article 1

URL: https://www.datovazurnalistika.cz/eu-volby-2024-presuny-volicu-analyza/

Summary: A detailed analysis of voters' shifts in the Czech Republic's European Parliament elections reveals that despite record turnout (36.45%), participation remains among the lowest in the EU. Compared to the 2021 parliamentary elections, all parties lost significant portions of their voter bases, with nearly 2.5 million prior voters abstaining this time. KSČM retained the highest percentage of its voters (54%), while Přísaha managed only 16%, indicating a nearly complete turnover of its electorate.

The election also highlighted voter radicalization, particularly among ANO supporters, who shifted toward the communist-led Stačilo! coalition and Přísaha-Motoristé. While parties like Spolu and the Pirates exchanged voters, ANO attracted former Social Democrats, and the Stačilo! coalition gained support from disillusioned ANO and SPD voters. The analysis, using advanced statistical methods, underscores the challenges parties face in mobilizing new voters and retaining their traditional bases.

Picture

The introductory image inserted into the article is presented in a static format, without any interactive elements; it functions solely as a visual preview intended to attract readers' attention.

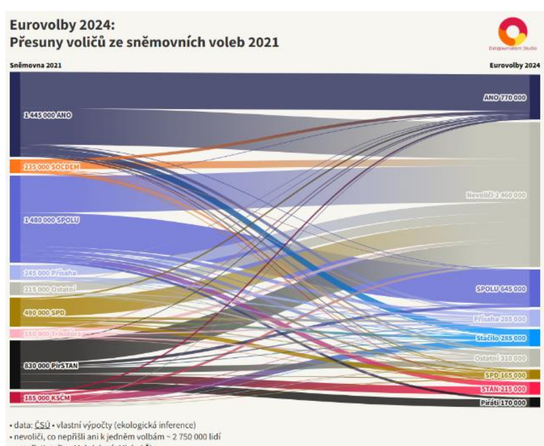


Figure A1: Voters' shifts in the Czech European Parliament elections 2024 from General elections 2021

Visualization 1

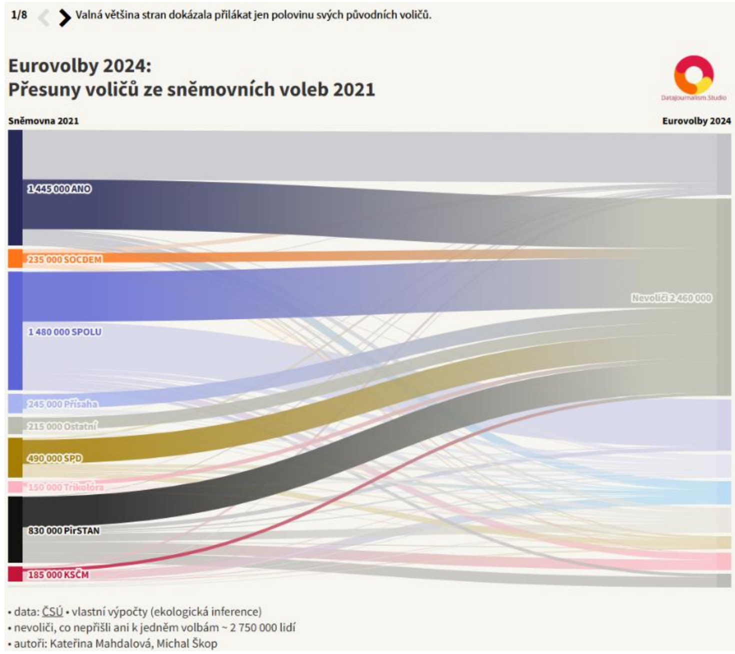


Figure A2: Voters' shifts in the Czech EU Parliament elections 2024 from General elections 2021 I

Visualization 2

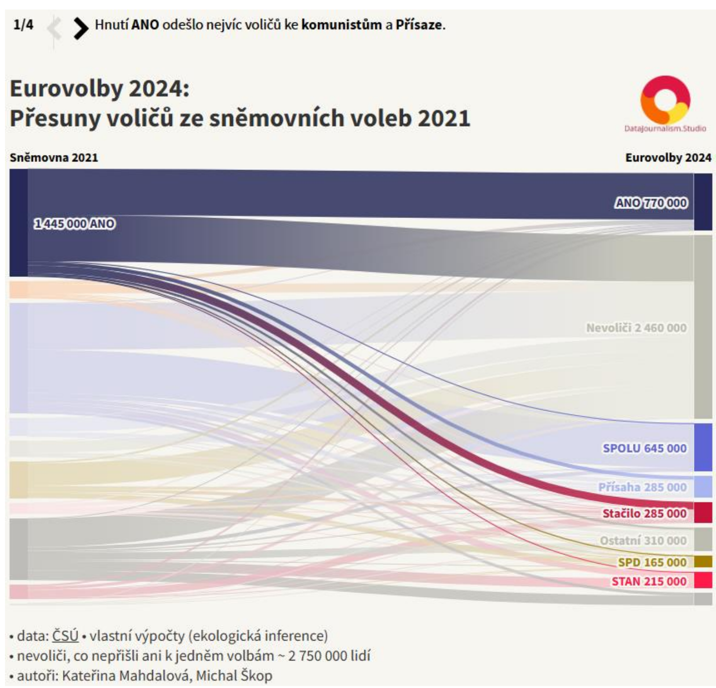


Figure A3: Voters' shifts in the Czech EU Parliament elections 2024 from General elections 2021 II

Visualization 3

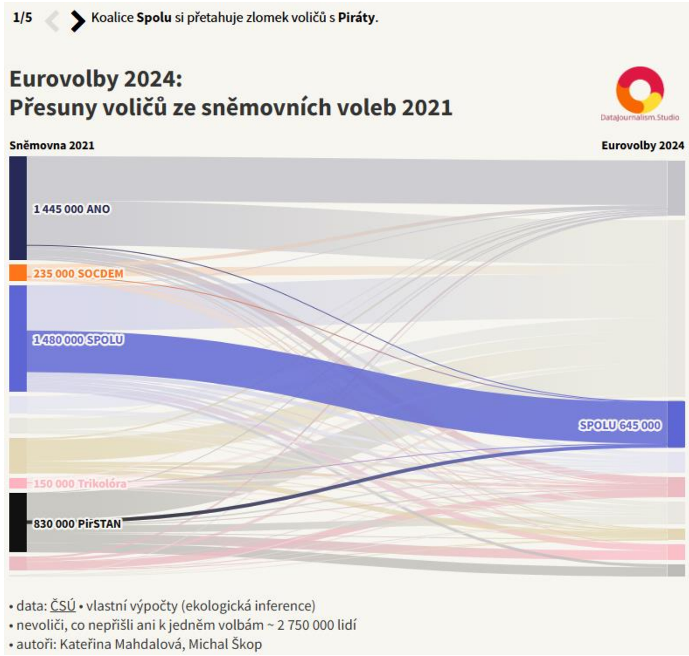


Figure A4: Voters' shifts in the Czech European Parliament elections 2024 from General elections 2021 III

Article 2

URL: https://www.datovazurnalistika.cz/snemovna-volby-mandaty-model/

Summary: A long-term analysis of election models in the Czech Republic highlights evolving voter preferences since the 2021 parliamentary elections. Updated with each new survey, the data reflects public sentiment at the time of data collection, not publication.

From August 2024, projections began focusing on coalitions expected in the 2025 elections, offering a more accurate picture of political dynamics. Weighted averages account for the timing and methodology of surveys, emphasizing recent trends over older data. Surveys show variability due to differing methodologies and the influence of undecided voters, who often sway final results. Coalitions like Spolu and the newly aligned Přísaha-Motoristé reshape the competitive landscape, while undecided voters and non-voters remain crucial in determining electoral outcomes.

Visualization 1

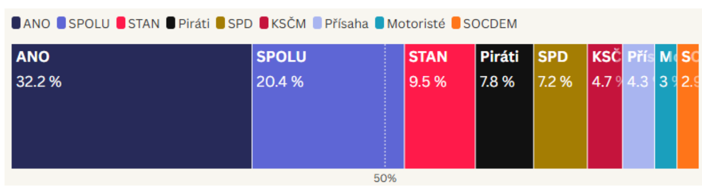


Figure A5: Model Poll-of-polls – Overview

Visualization 2

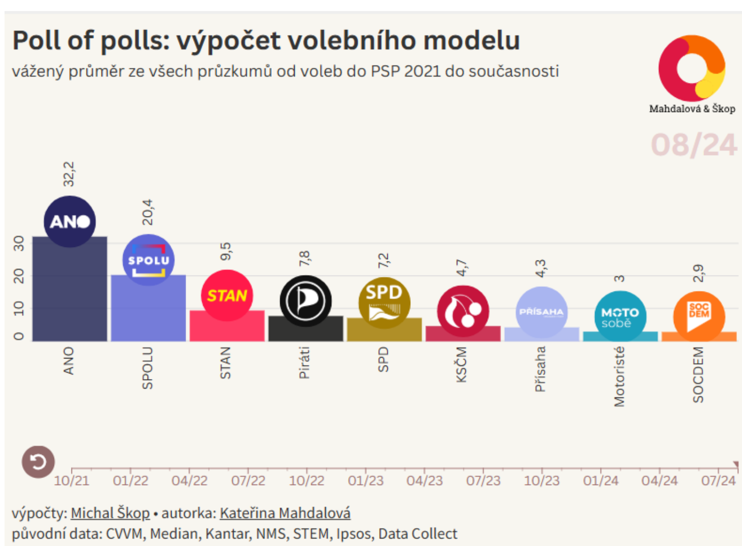


Figure A6: Model Poll-of-polls, race chart, changes in time

Visualization 3

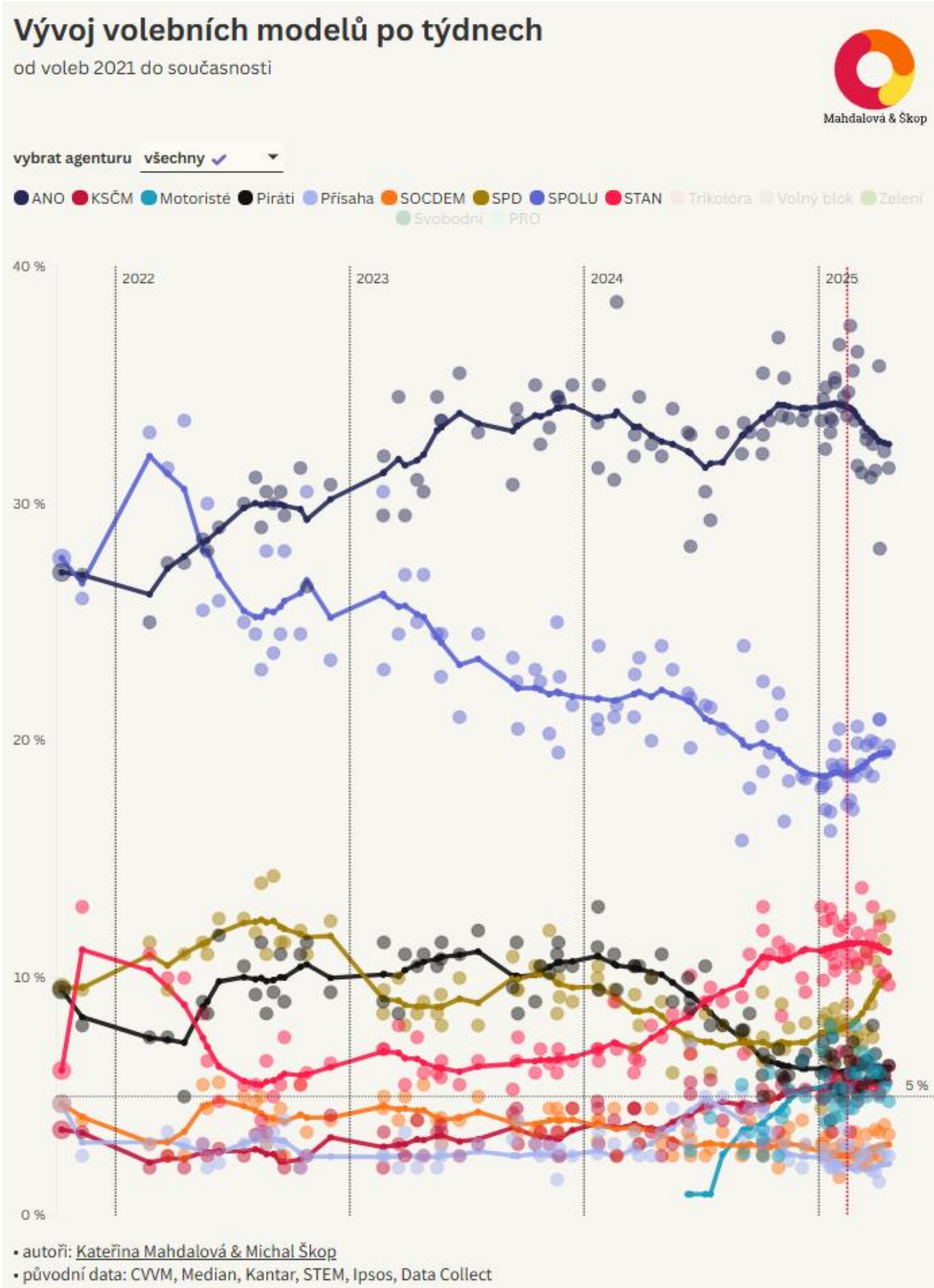


Figure A7: Model Poll-of-polls and individual polls, changes in time

Article 3

URL: https://www.datovazurnalistika.cz/eu-volby-2024-pruzkum/

Summary: A recent analysis of voter motivations in the Czech Republic's record-high European Parliament election turnout (36.45%) reveals a complex mix of protest and support-driven voting. Key motivations included expressing dissatisfaction with the government, opposing EU membership, or seeking change. Protest votes were strongest among far-left and far-right supporters, while optimism for change was notable among voters of Přísaha and Motoristé.

Support for the EU was highest among pro-European parties like Spolu, STAN, and the Pirates, while opposition was concentrated among SPD and communist voters. Age dynamics showed younger voters leaning toward progressive parties, while older demographics favored ANO and traditional parties. Prominent election figures included Filip Turek and Kateřina Konečná, reflecting polarized public opinion. Gender preferences varied, with some parties like Přísaha and Motoristé attracting more male voters.

Visualization 1

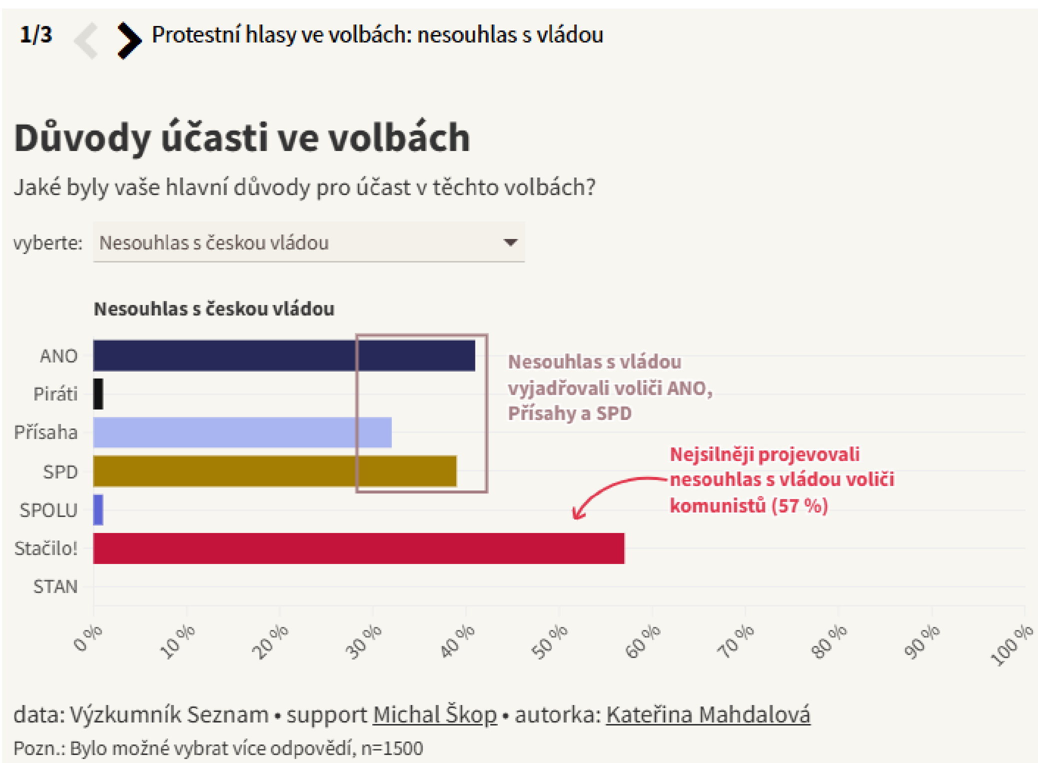


Figure A8: European elections: The reasons for participate I - protest vote

Visualization 2

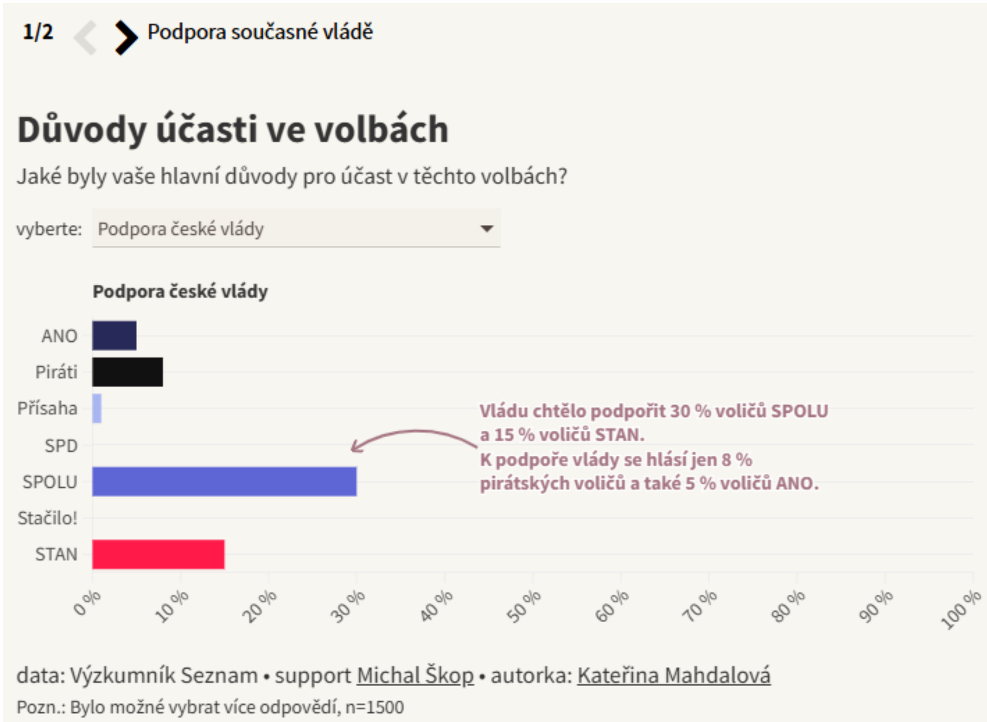


Figure A9: European elections: The reasons for participate II - support for the government

Visualization 3

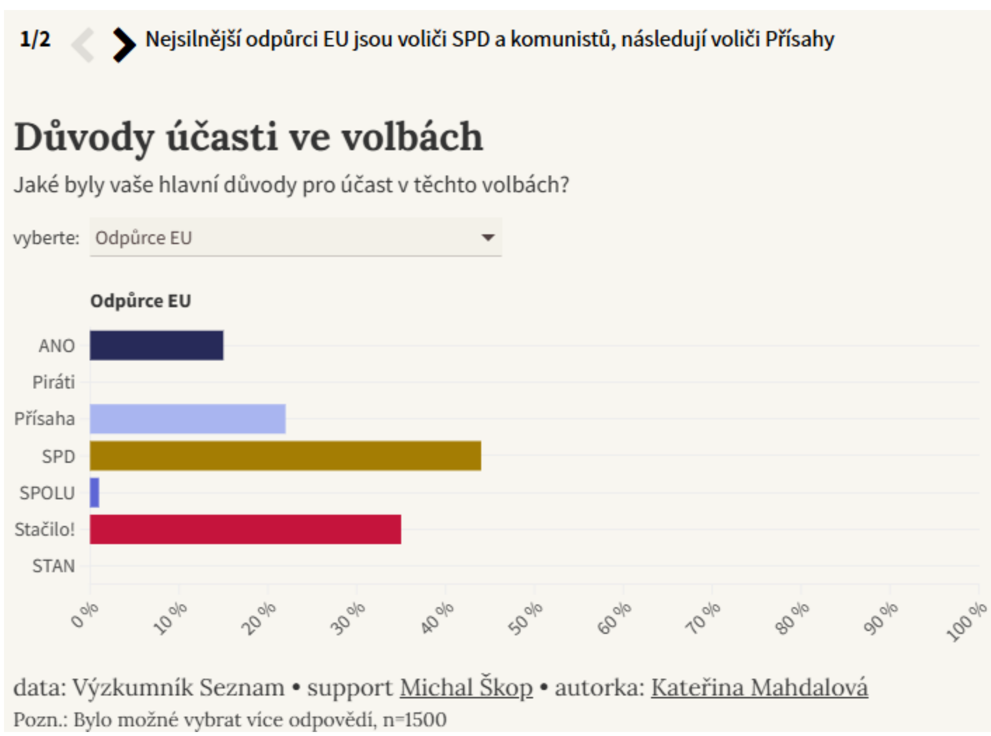


Figure A10: European elections: The reasons for participate III - for and against the EU

Visualization 4

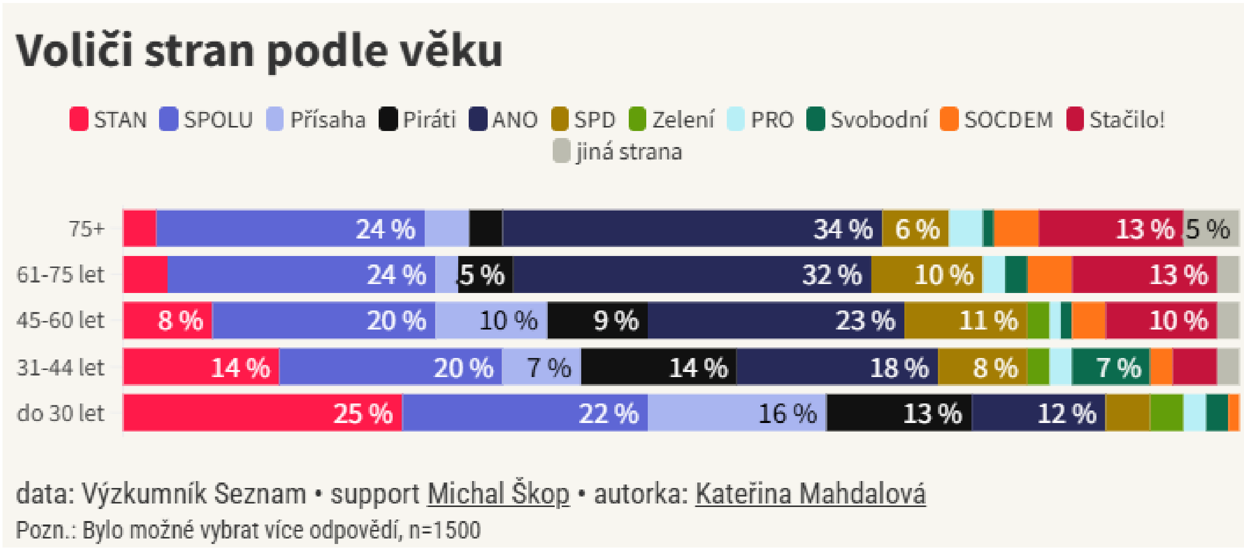


Figure A11: European elections: Voters' behavior by age groups

Visualization 5

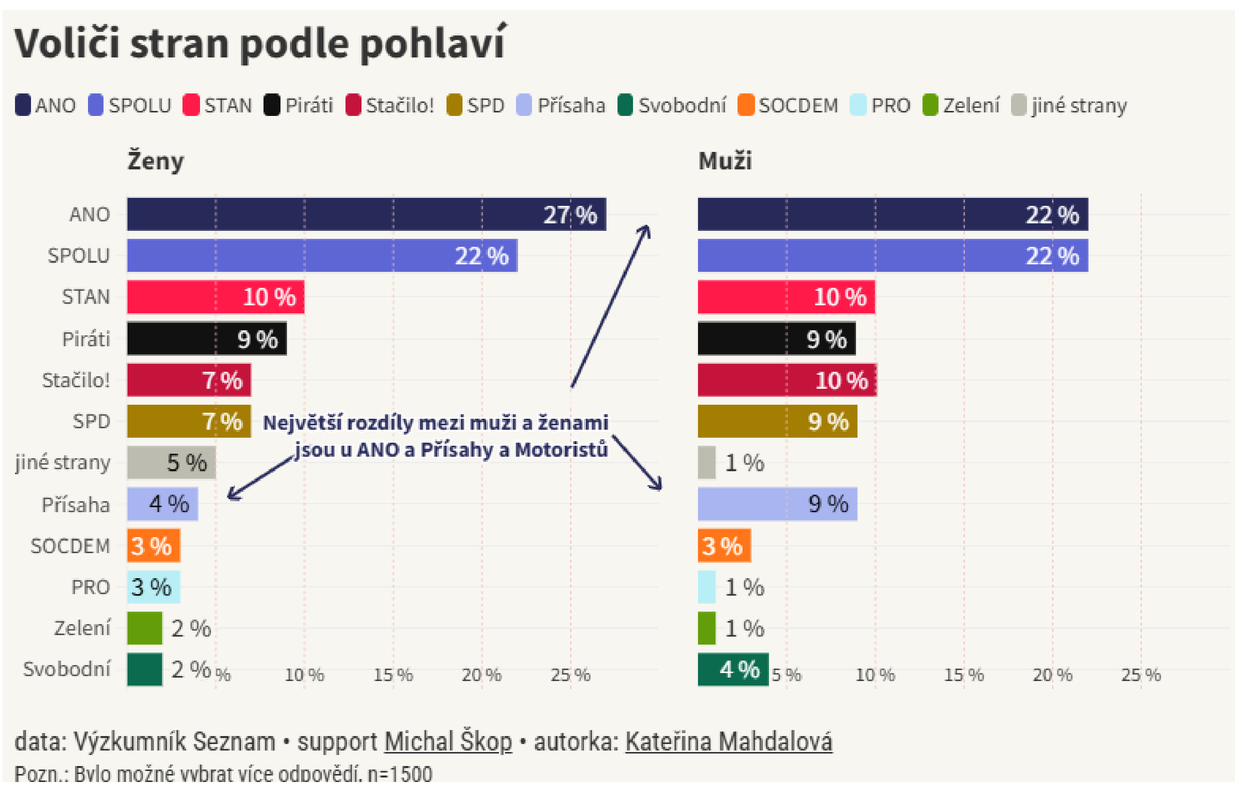


Figure A12: European elections: Voters' behavior by gender groups

Article 4

URL: https://www.datovazurnalistika.cz/cr-a-obrana-pruzkum-2024/

Summary: A recent survey in the Czech Republic revealed that while most citizens doubt the country's ability to defend itself independently in the event of a military attack, they are far more optimistic about collective defense through alliances like NATO and the EU, with 70% considering international cooperation vital for security.

About half of respondents support the reintroduction of mandatory military service for men, but there is significant resistance to extending this to women. Interest in voluntary military reserves has grown since the conflict in Ukraine, yet the number of reservists remains insufficient, with recruitment hindered by an aging demographic. The Ministry of Defense aims to address these challenges, targeting 10,000 reservists by 2030 through incentives and partnerships with universities to attract younger participants.

Visualization 1

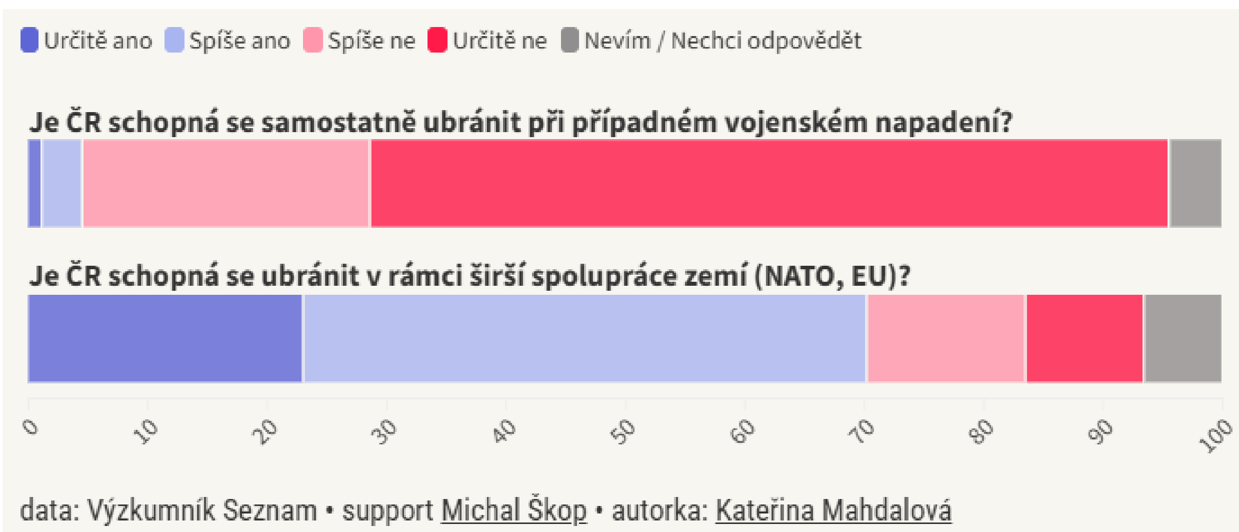


Figure A13: Survey results: defence, NATO, EU I - defence capability

Visualization 2

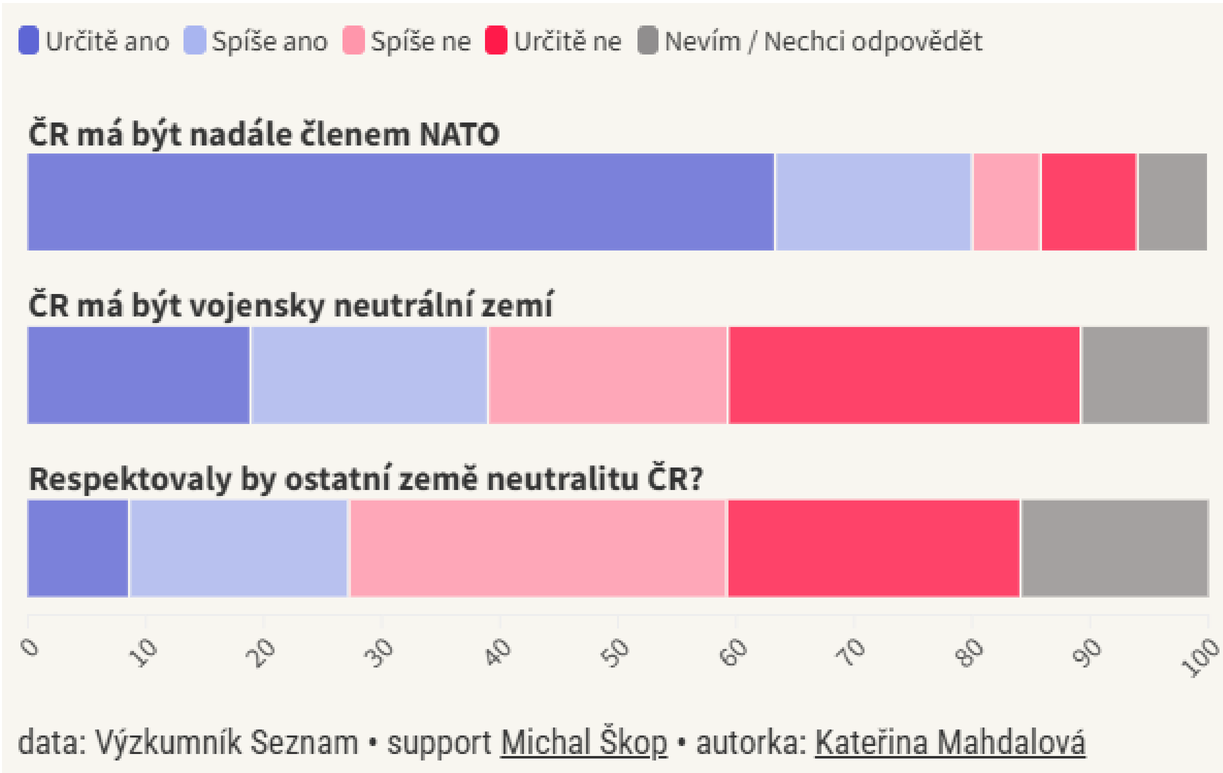


Figure A14: Survey results: defence, NATO, EU II - NATO / neutrality

Visualization 3

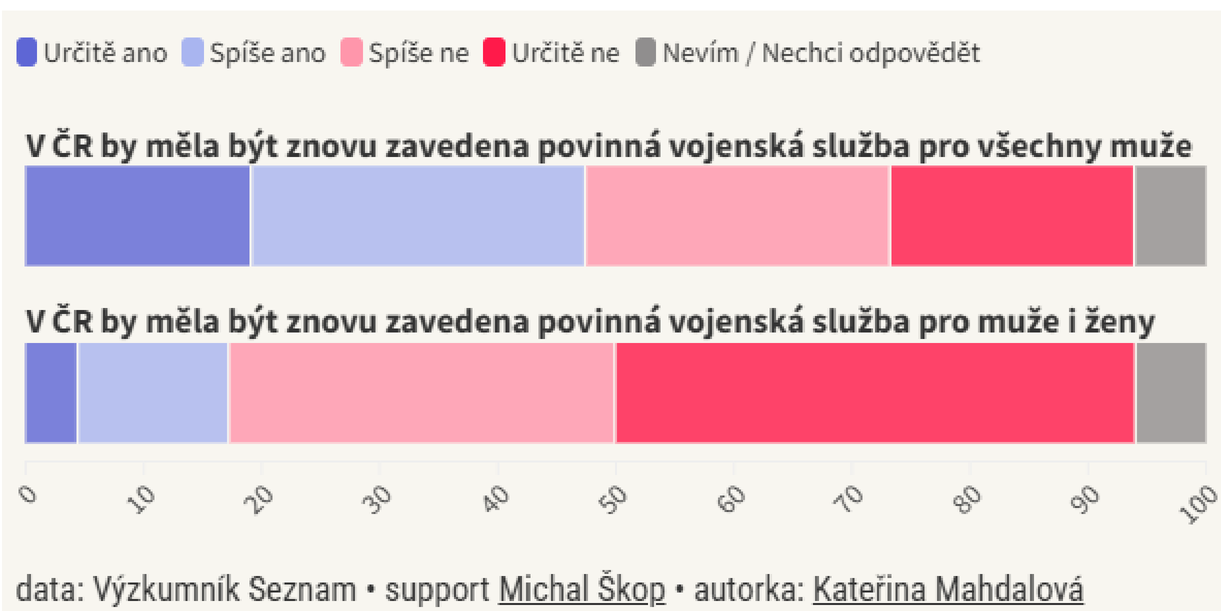


Figure A15: Survey results: defence, NATO, EU III - compulsory conscription

Visualization 4

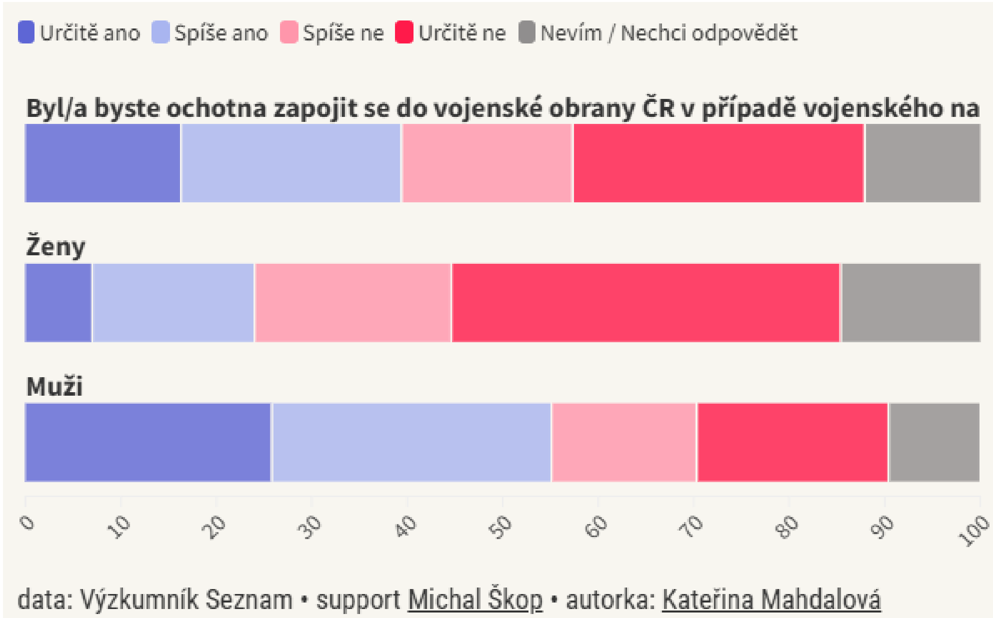


Figure A16: Survey results: defence, NATO, EU IV - personal involvement

Visualization 5

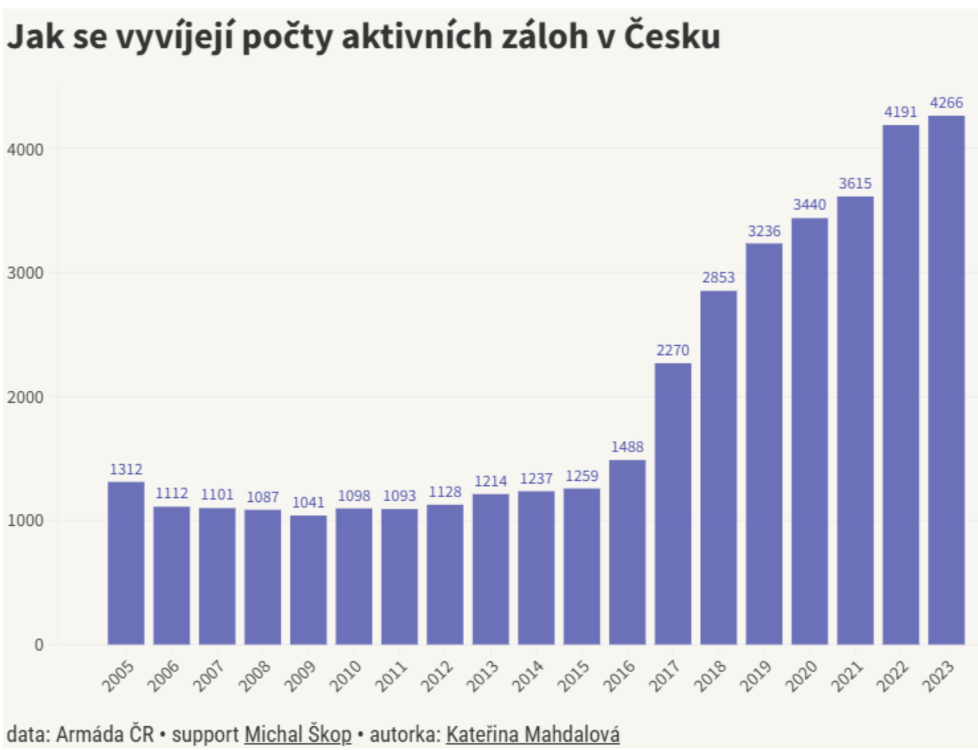


Figure A17: Survey results: defence, NATO, EU V - active army reservists

Article 5

URL: https://interaktiv.br.de/co2-rechner/

Summary: An interactive visualization in the style of scrollytelling developed by Bayerischer Rundfunk enables users to calculate their personal carbon footprint based on various lifestyle choices and activities. Users input data related to their energy consumption, travel habits, dietary preferences, and consumption patterns, with immediate visual feedback indicating the impact of these choices on their annual CO_2 emissions.

The visualization highlights differences in carbon emissions generated by individual decisions, such as frequent air travel, commuting by car versus public transport, dietary choices between meat-based and plant-based diets, and household energy sources. It emphasizes the importance of personal actions in reducing greenhouse gas emissions and mitigating climate change.

Through comparative visualizations, users can identify their largest sources of emissions and explore actionable steps for reducing their personal carbon footprint. The interactive tool also provides contextual information and average benchmarks, allowing users to assess their environmental impact relative to regional and national averages.

It is important to note that the data and measurements for Visualization 5 were obtained directly from Bayerischer Rundfunk's data journalist Constanze Bayer, who presented the findings at the Dataharvest 2024 conference in Belgium. Therefore, the methodology and data accuracy differ from visualizations 1–4.

Visualization 1

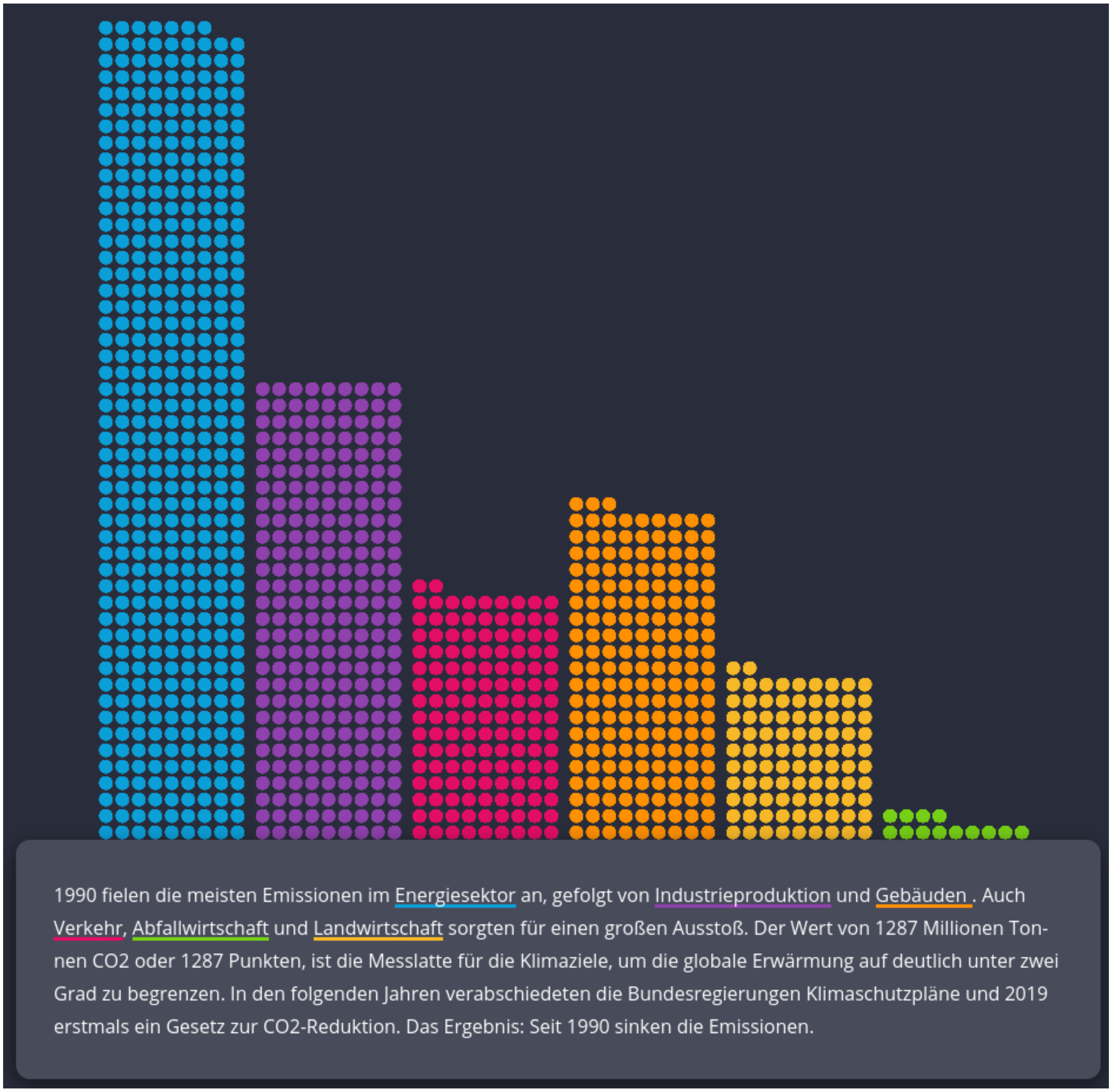


Figure A18: Breakdown of CO_2 emissions by category

Visualization 2 - Personal

Figure A19: Personal CO_2 emissions calculator – Input screen

Visualization 3

Figure A20: Comparison of planned and real emission reduction

Visualization 4

Figure A21: Breakdown of CO_2 emissions by category after reduction

Annex 3 – Desktop vs. Mobile

Article 1 - desktop version	Article 1 - mobile version
	

Graph 10: Diagrams comparing the desktop and mobile versions illustrate differences in the placement and visibility of interactive visualizations relative to text and other elements.

Legend: **Visualization**, **Text**, **Picture**, **Box**, **Additional Elements** (e.g., boxes, headers, advertisements, and links)