

Visualization Use in Process Mining Analysis

Lisa Zimmermann¹[0000-0002-6149-7060], Francesca
Zerbato²[0000-0001-7797-4602], and Barbara Weber¹[0000-0002-6004-4860]

¹ University of St. Gallen, Switzerland

² Eindhoven University of Technology, The Netherlands

Abstract. Process mining analysis relies on visual representations to support analysts in analyzing and improving processes. While a wide range of visualizations has been proposed, little empirical evidence exists on how these representations are actually used in practice. This paper investigates how analysts use visualizations during process mining analysis, based on an observational study with 33 participants. We encoded the analysts' interactions and thought processes and analyzed the use of visualizations over time, and in relation to both the focus of the analysis and the analytical tasks being performed. Our findings are twofold. We first describe the use of 12 distinct visualization types, and highlight a strong dominance of network-based representations, particularly for flow-oriented analysis and core analytical tasks such as exploration and descriptive modeling. Second, we uncover perceived difficulties in visualization use that expose limitations of their current implementation in existing tools. The findings of our analysis provide a sound foundation for the design of more effective, user-centered visualizations for process mining.

Keywords: Process Mining · Visual Process Analytics · Visualizations · Observational Study

1 Introduction

Process mining analysis relies on visual representations to support analysts in exploring, understanding, and improving processes based on event data [2,12]. Over the years, a broad range of visual process representations has been proposed, including process models, performance-oriented views, and data-centric visualizations that support different analytical goals. Most of these are graph-based process models generated by process discovery techniques [20]. Their design was often driven by technical considerations and optimized for aspects such as layout and integration of relevant information, rather than for user-centered qualities [24]. Moreover, insights from visual analytics research and established visualization design practices have been incorporated to a limited extent [6,24]. As a result, today, process analysts are presented with a relatively large set of visualizations they can work with.

At the same time, empirical knowledge about how visualizations are actually used during process mining analysis is still limited. Previous research has

provided valuable insights into visualization use by analyzing case studies and analysis outcomes [6,12,30]. However, such retrospective reviews primarily reveal which visual representations appear in final reports, rather than how analysts work with them throughout the analytical process. In particular, they provide little understanding of (i) when and how visualizations are created and used over time, (ii) why specific visualizations are chosen to analyze processes, and (iii) how evolving analytical goals, tasks, and challenges influence visualization use. This gap is especially relevant in exploratory analysis settings, where analysts frequently browse, compare, and opportunistically switch between visualizations instead of following predefined, goal-driven workflows [29].

However, understanding visualization use at this level of analytical work is a critical prerequisite for both research and design. First, without empirical evidence on how visualizations are used over time and across analysis phases, it remains difficult to assess whether existing visual representations effectively support the analytical tasks they are intended for or whether analysts rely on suboptimal strategies. Second, insights into how analysts use visualizations can reveal difficulties and shortcomings that remain hidden in reports but are crucial for identifying improvement potential in existing tools and for designing visualizations that align with real-world analytical practice.

In this work, we address these gaps by studying visualization use during process mining analysis based on observations. We analyze video recordings and think-aloud data from 33 analysts working in different tools and interacting with a diverse set of visual representations. By observing interactions over time and relating visualization use to analytical focus and tasks, we provide an empirical understanding of which visualizations are used, how and why they are used, when they appear during analysis, and which difficulties are perceived in practice.

Methodologically, this study builds on our prior work that introduced a video-based approach for analyzing visual data analysis processes [34]. Whereas the earlier contribution focused on establishing and validating the analysis method, in this paper, we apply and extend it to generate domain-specific insights into visualization use in process mining. In particular, we incorporate think-aloud data to obtain a deeper understanding of the analysts’ reasoning when selecting and interacting with visualizations.

In summary, this paper makes two contributions to the process mining and visualization communities. We (1) provide the first empirical, process-oriented characterization of how and why visual representations are used, and (2) identify shortcomings in current process mining visualizations and derive actionable implications to improve future visualization design.

In the remainder of this paper, we review the related work (cf. Sect. 2) and describe the observational study and our data analysis approach (cf. Sect. 3). Afterward, we present the results (cf. Sect. 4) and discuss them in light of how visualization design and tool implementation shape analytical workflows (cf. Sect. 5). Section 6 concludes the paper.

2 Related Work

In this section, we review literature on observational studies of process mining analyses and the role and use of visualizations in process mining.

Observational studies in process mining. Observational studies are still rare in process mining research. Most existing work focuses on technical improvements, while the *process of process mining*, i.e., how analysts actually work during analysis tasks, has received comparatively little attention [3,13].

Initial insights into analysts’ behavior were provided by Zerbato et al. [31], who identified general analysis patterns as well as common strategies and goals in process mining analyses. While this work already points to visually driven activities (e.g., “inspect map”), it remains at a high level of abstraction and does not explicitly investigate visualization use. A later study with involvement of the same authors [32] examined the analytical steps before and after arriving at a finding. Although this work reports the use of tool components such as the “Process Explorer” or “Statistics”, it does not reveal which concrete visualizations were used or how they supported hypotheses or findings.

With a focus on characterizing the overall analysis process, Sorokina et al. [25] proposed a framework for modeling process mining analyses as sequences of (cognitive) steps. Their work describes analysis behavior but, similar to prior studies, remains agnostic to the specific visual representations employed during analysis. More recently, Zimmermann et al. [34] transformed experiences from analyzing observational data into a generally applicable method for studying process mining analyses, focusing on methodological guidance rather than on visualization use itself.

While a small body of observational work exists that reveals temporal, procedural, and cognitive aspects of process mining analyses, none of the studies focuses on the use of visualizations or their alignment with analytical tasks.

Visualization use in process mining. The availability and role of visualizations in process mining have been investigated primarily through retrospective analyses, such as tool- and literature-based reviews. For example, Klinkmüller et al. [12] provided an overview of visualization techniques identified from BPI challenge [17] reports, highlighting their diversity and applicability to different domain problems. Yeshchenko et al. [30] surveyed visualization approaches for event sequence data but focused on design choices rather than empirical usage. Häge and Rehse [9] proposed a taxonomy of conformance checking visualizations based on a review of implementations in process mining tools, providing a common ground for describing and researching such visualizations in greater depth. Van den Elzen et al. [6] discussed visualization support for process mining from a visual analytics perspective, emphasizing opportunities for richer interaction and sensemaking.

Additionally, several works assess or improve specific visualization types in a more targeted manner. For example, van der Aalst [26] discussed directly-follows graphs (DFGs) from a practitioner’s perspective, highlighting their strengths

and limitations for analysis. Salas-Urbano et al. [21] also focused on DFGs by providing an interactive tool for working with collections of them to enable direct comparisons. Rubensson et al. [20] explored four layout strategies to align a DFG along a time axis, while Häge and Rehse [10] experimentally assessed whether fixing the position of nodes and edges in the DFG, the so-called ‘mental map’, improves analysis performance, but found no significant benefit. Other studies focused on other visualization types. For example, Schuster et al. [23] proposed a visualization for variant analysis, allowing the display of parallel or overlapping events. Hoogmoed et al. [11] focused on visualizations for resource analysis and Alman et al. [1] introduced a visualization framework that maps processes onto domain-specific backdrops. While these contributions advance visualization design, they consider effectiveness in isolation and do not validate how visualizations are used in conjunction within real-world analysis workflows.

3 Method

To analyze visualization use in process mining analyses, we conducted an observational study. To analyze the resulting data, we followed the method presented in [34], tailored to our research contributions and extended with an analysis of think-aloud data. Figure 1 provides an overview of our methodological approach, which we structure into five steps: observational study, event encoding, think-aloud encoding, event sequence generation, and analysis. Additional details complementing the method outline are provided in the appendix [33].

Observational Study. We designed an observational study to capture screen and audio recordings of process mining analysts performing an analysis task. The task required participants to explore the *Road Traffic Fine Management (RTFM)* event log [16] and identify three circumstances in which fines were not paid and, where possible, verify these circumstances with data evidence.

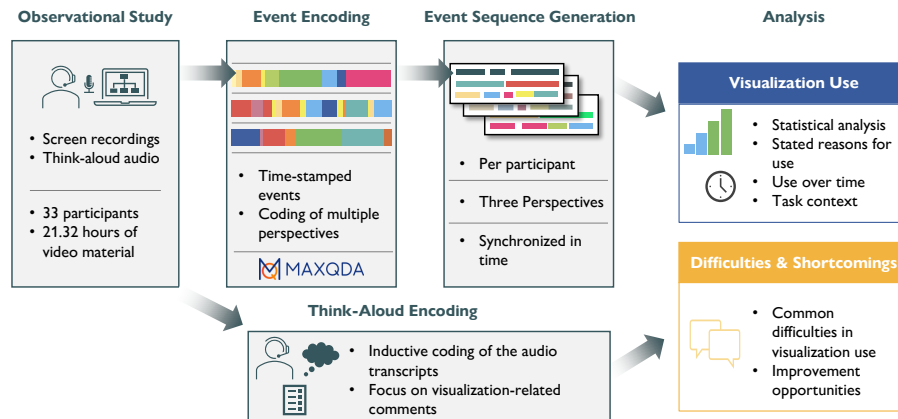


Fig. 1: Overview of our methodological approach.

The RTFM log records traffic fine events, relevant attributes (e.g., offense, fine amount, payment), and alternative execution paths (e.g., appeals). To ensure a shared understanding, participants received written documentation describing the underlying process and relevant business rules. The analysis task was intentionally open-ended, as we aimed to encourage exploratory analysis behaviors.

To collect the data, we conducted remote sessions with participants. We set up a standardized virtual machine (VM) that could be accessed via a web browser, ensuring that all analysts worked under identical technical conditions. The VM contained the RTFM event log, its documentation, and several process mining tools, including ProM [27], Celonis [4], and Disco [8], which enable interactive analysis in environments representative of both research and practice at the time of the study. The participants were free to choose their tools and analysis strategies, enabling the observation of naturalistic behavior and thus increasing ecological validity [7]. During the sessions, participants shared their screen and verbalized their thoughts, allowing us to also learn from their reasoning processes. Both screen activity and audio were recorded, resulting in synchronized video material capturing analysts’ tool interactions and verbal explanations.

We recruited participants through our professional networks and referrals from colleagues. To be eligible for the study, individuals had to: (1) have analyzed at least one real-life event log within the two years preceding the study; (2) be familiar with at least one of the provided process mining tools; and (3) have participated in at least one process mining project involving third-party data. These criteria were intended to ensure a minimum level of practical experience. A total of 33 participants worked primarily in ProM, Celonis, or Disco, generating 21.32 hours of video recordings with an average session duration of 38.7 minutes. Participant details are provided in the appendix [33].

Event Encoding. After data collection, we encoded analysts’ interactions and reasoning processes as discrete events defined by a start time, end time, and event type [34], creating an interpretive abstraction layer between raw observational data and subsequent analysis. Event boundaries were determined based on observable changes in interactions or verbalized shifts in attention and encoded in *MAXQDA* [28]. We captured three complementary perspectives reflecting which visualization was used, which data aspect was attended to, and which analytical task was pursued. Each perspective was coded in a separate pass over the recordings. Consequently, events were segmented independently for each perspective, resulting in distinct temporal intervals and event sequences per perspective. All coding schemes including detailed descriptions are provided in the appendix [33].

First, the *Visualization Perspective* includes events of participants interacting with specific visual representations. Visualization events were identified using visualization names provided by the tools or explicitly verbalized by participants. To enable comparison across tools, we later consolidated semantically equivalent visualizations (e.g., process maps with similar visual encodings) while preserving tool-specific distinctions (e.g., Process Explorer vs. Variant Explorer in Celonis). In the last coding round, the resulting categories were aligned with the categories proposed in [6], but we retained concrete visualization names when higher-level

categories were too abstract to reflect actual interactive artifacts used in practice. In particular, instead of summarizing all “node-based models” [6], we differentiated them in line with the visualization name used in the tools, typically referring to the underlying mining algorithm (e.g., Inductive Visual Miner (IvM) in ProM [15]) to allow for a more fine-grained analysis of their use.

Second, the *Focus Perspective* captures the domain-specific data aspects participants attended to (e.g., control-flow, time, or case attributes), reflecting what they were reasoning about at a given moment. We applied a hybrid coding strategy, starting from established process mining concepts and iteratively adding codes when participants focused on aspects not sufficiently covered by existing categories (e.g., user-defined KPIs or derived metrics).

Third, the *Task Perspective* captures analytical intent by segmenting the analysis process into analytical phases. Here, we followed a deductive coding approach, adapting the phases described by Crisan et al. [5] to the process mining domain. For example, data profiling, which is described as “*the process of assessing attributes [...] to understand the distribution of their values [...]*” [5], was adapted by us to “*Assessing the event log to understand its structure, characteristics, and distribution of timestamps, events, and control-flow patterns*”. Identified tasks include Defining Need, Profiling, Data Wrangling, Experimentation, Exploration, Descriptive Modeling, Diagnostic Modeling, Verification, and Interpretation. This perspective provides a bridge between the low-level interaction data and higher-level analytical strategies.

The coding was primarily conducted by one author, iteratively refined in joint sessions, and validated by at least one more author based on examples. Before, longer segments of the recordings were discussed among the authors to align the interpretations of codes and event boundaries.

Think-Aloud Encoding. In addition to event perspectives, we coded two auxiliary categories capturing reflective aspects of the analysis, labeled *Usage Reasons* and *Suggestions*. They were identified when participants explicitly explained why they used a visualization (e.g., they mentioned that they wanted to obtain an overview or verify a hypothesis) or suggested tool improvements by describing limitations, missing functionality, or desired features.

Unlike event perspectives, these codes do not form continuous event sequences. Instead, they serve as interpretive annotations and we derived them purely from the think-aloud data. We applied a bottom-up coding approach following grounded theory, starting from in-vivo codes and iteratively abstracting higher-level categories [22]. We used the resulting categories to contextualize the observed visualization usage patterns and support the interpretation of patterns emerging from our quantitative analysis of the event sequences.

Event Sequences. Based on the encoded events, we generated structured event sequences for our analysis. First, we created three single-perspective event sequences for each participant. These allow us to study visualization use, focus shifts, and evolving analytical tasks over time, including frequency, duration, and transitions between event types. To enable temporal comparisons across participants with varying session lengths, we added normalized timestamps to

each event sequence by dividing timestamps by the total session duration. Second, we generated two multi-perspective sequences by temporally aligning events across perspectives. The first sequence captures overlaps between visualization use and analytical focus, indicating what data aspects were attended to while using a visualization; the second captures overlaps between visualization use and analytical tasks, indicating during which tasks a visualization was employed. Two events from different perspectives were considered overlapping only if their timestamps showed a direct temporal intersection, i.e., if the start time of one event occurred before the end time of the other. For each such overlap, we generated an event in the multi-perspective event sequence, keeping both event names from the original perspectives (e.g., “[Visualization Perspective] Process Explorer – [Task Perspective] Exploration”) and using the exact timestamps from where the temporal intersection started until it ended.

Details on the analysis methods are outlined together with the results.

4 Results

In this section, we present the results of our analysis of visualization use in process mining, examining how, when, and for what visualizations are used (4.1), and summarizing perceived difficulties (4.2).

4.1 Use of Visualizations

To understand visualization use, we analyzed the *Visualization Perspective* event sequences and the two multi-perspective sequences, describing overlaps with events from the *Focus* and *Tasks Perspectives*. In addition to a quantitative analysis of the sequences, we qualitatively analyzed the *usage reasons* to understand why participants selected specific visualization types (cf. Sect. 3).

Across visualization types, we observed two general, recurring patterns regarding the usage reasons. In *goal-driven* situations, participants pursued a specific analytical objective and deliberately selected a visualization they believed to be well suited to achieve that goal. For example, participant p12 explored different process models until finding a representation they considered satisfactory (e.g., “*Let’s see, can we get a better [normative model] from this representation*”; p12). The second pattern can be characterized as *opportunistic*. Here, participants relied on familiar or easily interpretable visualizations and derived insights during interactions rather than pursuing a predefined goal. For instance, participant p4 selected a causal net because they perceived it as particularly explainable, and then inspected all the observable aspects of the visualization (e.g., “*It’s just I’m trying causal net. I find this very explainable structure*”; p4).

Table 1 presents an overview of all 13 visualization types that our participants used during the analysis. Since both the Process and Variant Explorer show DFGs with similar interaction capabilities [26], they are grouped under the category “Process map”, resulting in 12 visualization types at a more abstract level. Following the work by van den Elzen et al. [6], we grouped the types into **network-based**, **timeline-based**, and **chart-based** visualizations.

Visualization	#P	Tools		min	%	Focus (in %)											
		Celonis	Disco			ProM	w. usage	Activities	Attributes	Cases	Decision Points	Deviations	Flow	Performance	Statistics	Variants	
Network-based																	
Process map	32	x	x	x	321.85	38.64	20.1	-	0.1	-	-	-	39.3	10.4	22.6	7.6	-
→ Process Explorer	31	x	x	x	252.66	30.33	23.1	-	0.1	-	-	-	41.2	9.2	25.7	0.7	-
→ Variant Explorer	11	x	-	-	69.2	24.01	8.9	-	-	-	-	-	32.5	14.7	11.2	32.8	-
Petri net	4	-	-	x	22.59	14.05	12.5	6.0	-	25.2	-	-	30.3	-	25.9	-	-
IvM Model	4	-	-	x	15.56	9.68	14.3	5.3	-	-	-	4.5	39.1	5.2	31.5	-	-
Causal Net	3	-	-	x	6.63	4.12	8.2	23.7	-	-	-	14.5	34.9	-	18.8	-	-
Social Network	1	-	-	x	2.66	1.65	100	-	-	-	-	-	-	-	-	-	-
BPMN	3	x	-	-	1.45	0.50	15.6	-	-	-	-	-	84.4	-	-	-	-
SUM	32				370.75	44.51	19.4	1.1	0.1	1.7	0.5	38.6	9.2	23.0	6.5		
Timeline-based																	
Trace Sequence	12	-	x	x	33.79	6.20	9.1	25.7	10.2	-	-	-	8.8	5.2	5.7	35.2	-
Variant Sequence	5	-	-	x	16.14	10.04	24.2	1.6	-	-	-	-	1.6	-	3.9	68.8	-
SUM	12				49.93	9.17	13.2	19.1	7.4	-	-	-	6.9	3.8	5.2	44.4	-
Chart-based																	
Dotted Chart	5	-	-	x	34.17	21.26	38.1	-	0.4	-	-	-	3.8	44.0	13.7	-	-
Line Graph	8	x	x	-	6.89	1.02	-	-	8.4	-	1.7	-	-	67.1	22.9	-	-
Bar Chart	6	x	x	-	4.20	0.62	-	-	-	-	-	-	16.8	35.4	19.7	28.1	-
Scatterplot	1	x	-	-	0.45	0.16	-	50	-	-	-	-	-	-	50	-	-
SUM	16				45.70	5.49	29.5	0.7	1.5	-	0.2	4.1	46.0	16.0	2.0		

Table 1: Overview of visualization types used during analysis. For each visualization, the number of participants (#P), the tools that implement it, usage duration in minutes (min), weighted usage ratio (%), and the relative (w.r.t. normalized duration) overlaps with focus events are reported. We note that participants were free to choose the tool so visualization availability differed across tools, and time spent using a specific tool was unequal. The weighted usage ratio (%) therefore relates the time analysts interacted with a visualization to the time they spent in tools in which it was actually used.

Overview of Visualization Use and Usage Reasons. Overall, visualizations played a central role during the analysis, with analysts spending a total of 7.77 hours interacting with them, corresponding to an average of 14.13 minutes per participant and accounting for 36.48% of the total analysis time. The remaining time was spent working with less visual data representations, including raw numbers, documentation, or summarizing and planning analysis steps.

Across the entire analysis process, **network-based visualizations** dominated visualization use, accounting for 44.51% of the analysis time spent in the process mining tools (cf. Tab. 1). In particular, the process map emerged as the central visualization, used by 32 of 33 participants and alone accounting for 38.64% of tool interaction time. Participants relied on process maps to obtain an overview of the process, inspect dominant paths and frequent behavior, and identify activities of interest based on the analysis task at hand. Several participants described it as “*the most straightforward option*” (p20) for analysis due to familiarity, indicating both goal-driven and opportunistic usage. Beyond initial exploration, analysts repeatedly returned to process maps to identify relevant control-flow patterns, inspect frequency and performance annotations shown on

arcs, and validate earlier findings. For example, participant p17 used the process explorer to examine whether non-payment was indeed related to late notifications, verifying this by inspecting the resulting process flow filtered for cases with long notification times. Overall, network-based visualizations primarily supported reasoning about the process structure. As shown in Tab. 1, flow-related aspects constituted the dominant focus during their use. They were used to build an understanding of the overall process behavior and to test concrete hypotheses about control-flow dependencies, e.g., with time or the involvement of specific activities. Only for Petri nets and BPMN models, the main usage reasons include building a normative or reference model, as a basis for conformance checking.

In contrast, **timeline-based visualizations** were used substantially less frequently, accounting for 9.17% of tool interaction time (cf. Tab. 1). They were mainly used to inspect specific process variants, but also to analyze attributes and specific activities. Participants emphasized that these visualizations supported early observations (e.g., *“by just looking at these most prominent executions a lot of very interesting observations can be already achieved without going too much into detail”*; p33), but they were also used to compare variants or to identify variants with or without specific activities (e.g., without payment), where visual cues, such as color-coding, supported quick screening. We observed more goal-oriented reasoning about their use. In case of the trace sequence visualization, participants used them to inspect concrete examples to either form hypotheses that they later tested on larger subsets, or to validate an initial gut feeling about the process (e.g., *“I would then actually go through this list and [...] see these random samples. I just check if the fine amount and the payments are in line with each other”*; 33). Overall, timeline-based visualizations, and especially trace sequence visualizations, enabled analysts in our study to reason on the case-level, complementing the more aggregated structure-oriented role of network-based visualizations.

Lastly, **chart-based visualizations** accounted for only 5.49% of tool interaction time. Within this category, the dotted chart dominated, representing more than three quarters of chart-based visualization use. As visible from Tab. 1, participants primarily used charts to explore performance aspects, such as temporal distributions. Oftentimes, they were also used to filter (relevant) subsets of the data. Charts were rarely the primary entry point into the analysis, but instead complemented other visualizations by providing alternative perspectives on the data, such as performance views.

Visualization Use over Time. Figure 2 shows the temporal distribution of visualization use across analysis sessions.

As shown in Fig. 2a, network-based visualizations were used consistently throughout the sessions, with no pronounced differences between the first and second analysis halves. Interestingly, a closer inspection of the individual visualization types in this category reveals that most were used slightly more frequently in the second half, except for the Process Explorer, which was used more in the first half (average events: 7.09 vs. 5.56). Participants reported using it at the beginning to *“see a bit how the process looks like”* (p14) and *“get a general overview”*

(p21). This highlights its central role during initial familiarization, while more specialized network-based visualizations were consulted once a basic process understanding was established. Timeline-based visualizations (cf. Fig. 2b) show a usage peak in the third quarter, mainly driven by the use of Trace Sequence visualizations. This may indicate a shift toward more focused, case-level inspection to validate hypotheses or examine specific example traces. Chart-based visualization use (cf. Fig. 2c) has a similar pattern, with more frequent use in the second analysis half, typically after participants had already explored the process structure, to investigate complementary perspectives.

In general, Fig. 2 shows that the analysis process began with an initial familiarization phase with little visualization use, followed by sustained use of network-based visualizations as a backbone throughout the analysis. Timeline-based and chart-based visualizations were used more intensively in the mid-to-late stages (60%-80% of the analysis progress) to support focused (e.g., case-level) investigation and validation, before analysts again relied on network-based visualizations to consolidate their findings.

Visualization Use Across Analytical Tasks. To analyze visualization use across analytical tasks, we used the multi-perspective event sequence derived from overlaps between the *Task* and *Visualization Perspectives* and normalized overlap durations by the total time analysts spent on the respective task. The resulting heatmap (cf. Fig. 3) thus shows, for example, that while participants worked on understanding the objectives and goals of the analysis (i.e., defining need), 4% of the time was spent using the Process Explorer and 1% using the Variant Explorer, while the remaining time was spent without interacting with any visualization, as analysts were reading the process documentation or understanding the analysis question.

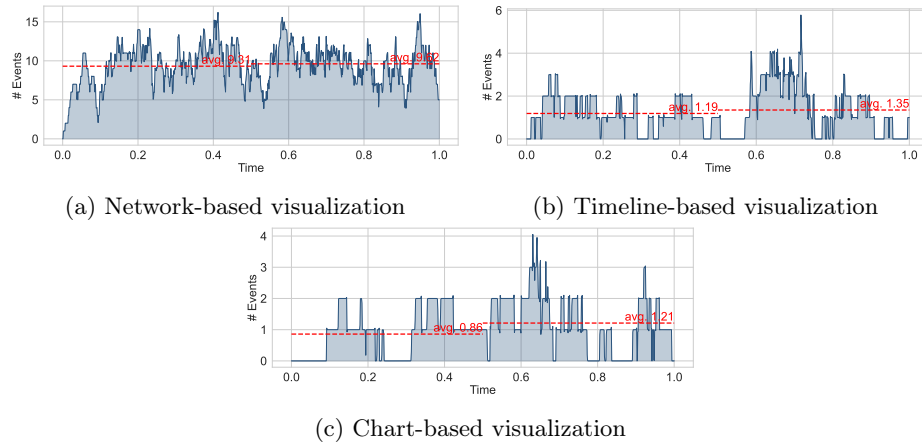


Fig. 2: Visualization use over time for different visualization types. Note that for readability purposes, y-axis scales are not consistent over visualization types.

Figure 3 shows that visualization use varied across analytical tasks but was dominated by network-based visualizations, as they also dominated overall visualization use (cf. Tab. 1). The Process Explorer was the only visualization used across all analytical tasks. It played a central role particularly during exploration (38%) and descriptive modeling (26%), confirming its relevance for understanding process behavior and describing specific aspects of the process. In contrast, the Variant Explorer was used more selectively and predominantly during diagnostic modeling (13%) and interpretation (15%). Instead of using it for general exploration, participants rather referred to it for looking for specific reasons of process behavior by, e.g., comparing variants with each other.

Experimentation and exploration involved the widest range of visualization types, including the Trace Sequence, Variant Sequence, and Dotted Chart visualizations. These visualizations allowed participants to inspect individual cases and temporal dependencies, supporting example-driven reasoning by examining concrete traces to better understand attribute relationships or validate emerging hypotheses. The use of multiple visualization types for these tasks indicates the need to investigate process behavior from multiple perspectives in different visualizations. Conversely, early-stage analytical tasks such as defining needs and data wrangling involved very limited visualization use overall, indicating that these tasks were primarily driven by problem formulation and data preparation rather than visual analysis. Similarly, profiling, i.e., the assessment of the event log to understand its structure, showed only moderate visualization use, mainly relying on the Process Explorer. Finally, verification and interpretation again involved increased use of network-based visualizations. Notably, across all analytical tasks, substantial portions of time were spent without interacting with any visualization. This is partly expected, as analytical work also involves non-visual activities such as thinking, reading, or tool configuration, and not all aspects of

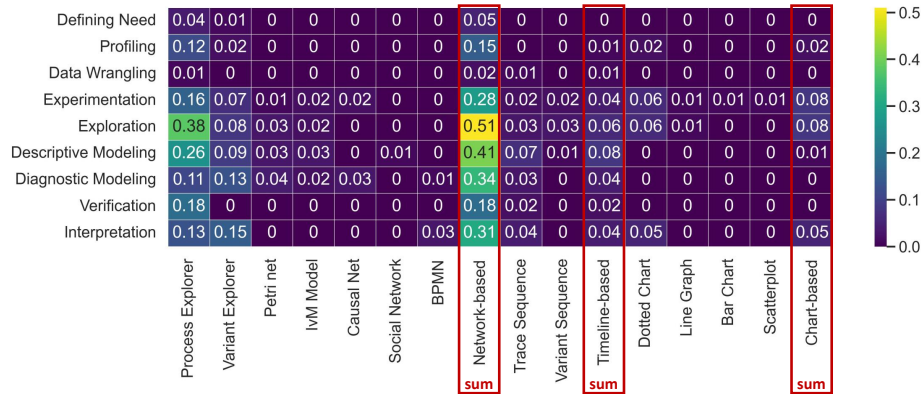


Fig. 3: Overview of visualization use across analytical tasks. Relative interaction time per visualization is shown for each task. The highlighted columns aggregate lower-level visualization types.

analysis can be supported through visualizations. However, the comparatively low visualization use for certain tasks may indicate that suitable visualizations are missing or unavailable in current tools, pointing to opportunities to better support these tasks in process mining.

4.2 Observed Difficulties in Visualization Use

To identify improvement potentials for existing visualizations in process mining, we analyzed the think-aloud data coded as *Suggestions* (cf. Sect. 3). While participants generally succeeded in using network-based, timeline-based, and chart-based visualizations to derive insights, they also repeatedly articulated limitations related to transparency, comparison, aggregation, usability, and consistency across tools.

Transparency of Representativeness and Filtering Effects. A central difficulty of our participants was understanding how representative a visualization was of the underlying event log, especially when filters or relevance thresholds were applied. This issue was most prominent for network-based visualizations, especially the process map. Participants expressed uncertainty about whether the visible structure reflected the “real” process or only a filtered subset. One analyst noted: *“This is actually a bad way to browse through your processes because you never know if what you see is real. You hide some paths and you have no clue how many other or what’s the percentage of the cases this is.”* (p31). Confusion also arose when statistics shown on filtered models (e.g., case counts) did not match expectations, leading to questions about missing activities or paths (*“Where is the other activity? There are other cases that are not [shown] here”*; p20, not aware that not all paths were displayed in the process explorer). Some participants further pointed to modeling limitations, such as the lack of explicit concurrency representation in many network-based visualizations (*“The model cannot represent concurrency [...] probably it has been done before or after”*; p7). These observations suggest that filtering and abstraction simplify models but make it harder for analysts to judge representativeness. Stronger visual cues and interaction mechanisms that communicate (i) active filters, (ii) log coverage, and (iii) underlying modeling assumptions or abstractions, could reduce the need for constant inference about hidden elements and lower the risk of misinterpretation.

Limited Support for Comparisons. Another prominent difficulty was the lack of support for comparing visualizations generated from different subsets of the data, for example before and after applying filters. Participants repeatedly expressed the wish to inspect multiple process models side-by-side (e.g., *“I was thinking, if I can have a side-by-side process model would be good. Yeah, so I can compare the process because it’s hard to remember how it was before.”*; p2). The lack of side-by-side comparison forced analysts to rely on memory or manual artifacts (i.e., participants took screenshots for comparisons), increasing cognitive load and the risk of misinterpretation. This limitation was especially problematic when analysts attempted to contrast cases with and without certain activities in network- or timeline-based visualizations.

Aggregation Levels and Switching Perspectives. Participants encountered difficulties navigating different data aggregations, i.e., when switching between case-level and variant-level analysis. This was especially evident for timeline-based visualizations, where analysts struggled to retrieve aggregated variant-level statistics after selecting individual cases: *“It’s fine if I can select individual cases, but now I want to see not the values for this particular case, but for this variant. Where can I do it?”* (p31). Similarly, analysts struggled when switching between tools to compare results. Implementation differences complicated the validation of findings and the interpretation of data levels. Participants explicitly compared similar visualization across tools to reconcile conflicting impressions and draw reliable conclusions: *“Because in Disco it looks like there is a bottleneck here, but in Celonis it looks like it’s only one percent of the cases that followed this path.”* (p32). Overall, these observations suggest insufficient support for easily switching perspectives between aggregation levels and tools.

Usability, Accessibility, and Visual Clutter. Beyond conceptual issues, participants reported several usability and visual design problems. Timeline-based visualizations were, for example, described as cluttered and visually unappealing (*“I actually don’t like this representation [...] it’s kind of clumsy, not very appealing”*; p31), which hindered rapid comprehension.

For dotted charts, participants emphasized their analytical value, but criticized their current implementation. They reported that configuration is difficult, the UI is not intuitive, and important settings are hard to discover. Additionally, accessibility was problematic, as it did not support color vision deficiency: *“I would really wish to have this in other tools. The dotted chart is a starting point for me [but] Although I know what diagrams I want to see, it’s not easy to find the correct configuration [...] I am colorblind [so] I’m not really aware of the difference between this color [...] This inspector window always stays in front. It’s annoying.”* (p15). Given that dotted charts are not consistently available across process mining tools, a more standardized, accessible, and user-friendly implementation could facilitate broader adoption and accessibility.

Similar concerns were also raised for causal nets, which were described as powerful but difficult to interpret for non-expert audiences, limiting their suitability for communication beyond expert analysts. This indicates that even powerful visualization types may remain underutilized if their interaction design, appearance, or accessibility is inadequate.

5 Discussion

With this work, we have provided a process-oriented characterization of visualization use during process mining analysis. Across the 33 participants and multiple tools, we found that visualizations play a central role throughout the analysis process, accounting for more than one third of the analysis time. Different visualization categories supported distinct analytical perspectives. Network-based visualizations form an important backbone as they are used extensively in our

study setting, primarily for understanding process flow, but also for inspecting all other focus areas. Timeline-based visualizations complement them by supporting case- and variant-level reasoning, while chart-based visualizations are mainly used to investigate performance and to perform filtering operations.

Beyond this, during our analysis, we noticed that visualization use is not solely driven by analytical intent. We observed both goal-oriented and opportunistic patterns, partially influenced by tool implementation. In several cases, visualizations were used not because they were considered most suitable, but because specific functionality was tightly coupled to them. For example, BPMN models in Celonis were only used in the context of conformance checking. Analysts did not select them to explore process behavior, but the tool implementation required them to edit a BPMN model to perform the desired analysis. Similarly, specific filtering operations, such as time-based filtering in Disco, were only accessible via charts. In such situations, visualizations acted as gateways to functionality, rather than freely chosen representations.

At the same time, participants also deliberately selected visualizations to answer specific questions. The tool structure appeared to influence this balance, as in ProM, the tool required explicit function selection, and overall visualization usage appeared more goal-oriented, while prominently exposed visualizations on dashboards encouraged opportunistic use. In general, visualization use reflects an interplay between analytical reasoning and system-imposed constraints. This observation aligns with known concepts in visualization research, indicating that tool affordances and design decisions indeed shape analytical reasoning and interaction patterns [18,19]. Distinguishing analytical preferences from tool-driven effects is, therefore, crucial when interpreting usage patterns and designing future systems.

Our findings also contribute to understanding how process mining is practiced. Although in published reports aggregated results and chart-based summaries are dominantly used [12], our analysis suggests that, in tasks like ours, control-flow exploration in network-based and process mining-specific visualizations (i.e., process map) are central. This discrepancy likely reflects the methodological difference between studying analytical artifacts in isolation and observing analytical processes as they unfold [14]. We observed time-constrained, exploratory sessions, whereas case studies typically focus on longer and more complex projects. The prominence of control-flow exploration resulting from our study may therefore reflect earlier analysis phases rather than different visualization use. However, our results imply that structural understanding is a prerequisite for performance- or case-level investigations, with process maps serving as an anchor that analysts repeatedly use to contextualize their findings.

Finally, our results strongly support the vision of multi-faceted visual process analytics [6]. Different visualization categories address different analytical facets, and analysts rely on their combined use throughout the analytical process. However, our observations reveal that this integration is only partially realized. Limited support for side-by-side comparison, insufficient transparency regarding filtering effects and representativeness, and inconsistencies across tools

hinder multi-perspective reasoning, which is essential when working with processes. Thus, while our findings empirically validate the conceptual need for multi-faceted visual analytics, they also highlight concrete gaps between this vision and current tool support.

Limitations. Although we observed a relatively large number of analysts across multiple tools, our study design may not fully reflect real-world analysis practices. We provided a single and prepared event log, and analysts had limited time depending on their availability. In practice, projects often span several days, involve richer domain knowledge, and are shaped by organizational constraints that may further influence visualization use. Moreover, providing a concrete analysis question and requiring think-aloud may have shaped visualization use, despite encouraging participants to work as they normally would. The observed usage patterns were also influenced by the availability and specific implementations of visualizations in the selected tools at the time of the study, which may limit generalizability. Finally, as with all qualitative and observational research, our analysis involved interpretative coding of empirical data. While we mitigated potential subjective bias through multiple coding rounds and cross-checking, some degree of interpretation remains inherent.

6 Conclusion

To the best of our knowledge, this study is the first to investigate visualization use during process mining analyses from a process-oriented perspective based on empirical observations of analysts’ behavior. By examining analysts’ interactions across different tools and relating visualization use to both the analytical focus and the task context, we show that visualizations are not used in isolation, but as part of an evolving analytical process.

Our findings highlight the central role of network-based visualizations in supporting a structural understanding throughout all analytical tasks, while we found that timeline- and chart-based representations mainly serve complementary purposes such as case-level inspection, performance analysis, and validation. In the context of the open-ended but time-constrained exploratory task studied here, network-based visualizations in particular emerged as an important backbone for building an initial understanding of the process. Additionally, the study reveals important shortcomings in current tool support, particularly regarding transparency, comparison, accessibility, and the seamless integration of multiple analytical perspectives. Beyond characterizing current practice, this work offers an empirically grounded basis for improving the design of process mining visualizations and tools, especially in response to the difficulties we identified in their use. Building on these insights, future work should translate the observed usage patterns and challenges into concrete design implications and visualization concepts that better support analysts in practice.

Acknowledgments. This work was made possible thanks to the ProMiSE project, funded by the Swiss National Science Foundation under Grant No. 200021_197032.

References

1. Alman, A., Arleo, A., Beerepoot, I., Burattin, A., Di Ciccio, C., Resinas, M.: Tiramisù: making sense of multi-faceted process information through time and space. *Journal of Intelligent Information Systems* pp. 1–23 (2024)
2. Badakhshan, P., Wurm, B., Grisold, T., Geyer-Klingenberg, J., Mendling, J., Brocke, J.: Creating business value with process mining. *J. Strateg. Inf. Syst.* **31**, 101745 (2022). <https://doi.org/10.1016/j.jsis.2022.101745>
3. vom Brocke, J., Jans, M., Mendling, J., Reijers, H.A.: A Five-Level Framework for Research on Process Mining. *Business and Information Systems Engineering* **63**(5), 483–490 (2021). <https://doi.org/10.1007/s12599-021-00718-8>
4. Celonis: Process mining platform. <https://www.celonis.com/platform/process-mining/> (2024), accessed: 2025-11-27
5. Crisan, A., Fiore-Gartland, B., Tory, M.: Passing the data baton: A retrospective analysis on data science work and workers. *IEEE Transactions on Visualization and Computer Graphics* **27**, 1860–1870 (2021). <https://doi.org/10.1109/TVCG.2020.3030340>
6. van den Elzen, S., Jans, M., Martin, N., Pieters, F., Tominski, C., Villa-Uriol, M.C., van Zelst, S.J.: Towards Multi-Faceted Visual Process Analytics. *Information Systems* **133**, 102560 (2025). <https://doi.org/10.1016/j.is.2025.102560>, <https://doi.org/10.1016/j.is.2025.102560>
7. Fahmie, T.A., Rodriguez, N.M., Luczynski, K.C., Rahaman, J.A., Charles, B.M., Zangrillo, A.N.: Toward an explicit technology of ecological validity. *Journal of applied behavior analysis* (2023). <https://doi.org/10.1002/jaba.972>
8. Günther, C.W., Rozinat, A.: Disco: Discover your processes. In: *Demonstration Track of the 10th International Conference on Business Process Management, BPM Demos 2012*. pp. 40–44. CEUR-WS. org (2012)
9. Häge, M.C., Rehse, J.R.: Conformance checking visualizations: development, evaluation, and demonstration of a taxonomy. *Process Science* **2**(1), 18 (2025)
10. Häge, M.C., Rehse, J.R.: Mental maps in process mining: Does stabilizing dfgs improve process analysis performance? In: Senderovich, A., Cabanillas, C., Van-derfeesten, I., A. Reijers, H. (eds.) *Business Process Management*. pp. 523–540. Springer Nature Switzerland, Cham (2026)
11. Hoogmoed, A., Vidgof, M., Djurica, D., Rubensson, C., Mendling, J.: Visual representation of resource analysis insights for process mining. In: *International Conference on Business Process Modeling, Development and Support*. pp. 117–128. Springer (2024)
12. Klinkmüller, C., Müller, R., Weber, I.: Mining process mining practices: An exploratory characterization of information needs in process analytics. In: *Int. Conf. on Business Process Management*. Springer (2019). https://doi.org/10.1007/978-3-030-26619-6_21
13. Kubrak, K., Milani, F., Nolte, A.: Process mining for process improvement-an evaluation of analysis practices. In: *International Conference on Research Challenges in Information Science*. pp. 214–230. Springer (2022)
14. Lam, H., Bertini, E., Isenberg, P., Plaisant, C., Carpendale, S.: Empirical studies in information visualization: Seven scenarios. *IEEE transactions on visualization and computer graphics* **18**(9), 1520–1536 (2011)
15. Leemans, S.J., Fahland, D., Van Der Aalst, W.M.: Process and deviation exploration with inductive visual miner. In: *12th International Conference on Business Process Management, BPM 2014*. pp. 46–50. CEUR-WS. org (2014)

16. de Leoni, M., Mannhardt, F.: Road traffic fine management process. Eindhoven University of Technology, Dataset **284** (2015)
17. Lopes, I.F., Ferreira, D.R.: A survey of process mining competitions: the bpi challenges 2011–2018. In: International Conference on Business Process Management. pp. 263–274. Springer (2019)
18. Munzner, T.: A nested model for visualization design and validation. *IEEE Trans on Vis. and Comp. Graphics* **15**(6), 921–928 (2009)
19. Pike, W.A., Stasko, J., Chang, R., O’connell, T.A.: The science of interaction. *Information visualization* **8**(4), 263–274 (2009)
20. Rubensson, C., Kaur, H., Kampik, T., Mendling, J.: Timeline-based process discovery. *Information Systems* **133**, 102568 (2025)
21. Salas-Urbano, M., Capitán-Agudo, C., Cabanillas, C., Resinas, M.: Interactive creation, visualization, and exploration of process model collections. *Software and Systems Modeling* pp. 1–26 (2026)
22. Saldaña, J.: The coding manual for qualitative researchers. Sage publications Ltd (2015)
23. Schuster, D., Zerbato, F., van Zelst, S.J., van der Aalst, W.M.: Defining and visualizing process execution variants from partially ordered event data. *Information Sciences* **657**, 119958 (2024)
24. Sirgmetts, M., Milani, F., Nolte, A., Pungas, T.: Designing process diagrams—a framework for making design choices when visualizing process mining outputs. In: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". pp. 463–480. Springer (2018)
25. Sorokina, E., Soffer, P., Hadar, I., Leron, U., Zerbato, F., Weber, B.: Pem4ppm: A cognitive perspective on the process of process mining. In: Intl Conf on Business Process Management. pp. 465–481. Springer (2023)
26. Van Der Aalst, W.M.: A practitioner’s guide to process mining: Limitations of the directly-follows graph (2019)
27. Verbeek, H., Buijs, J., Van Dongen, B., van der Aalst, W.M.: Prom 6: The process mining toolkit. *Proc. of BPM Demonstration Track* **615**, 34–39 (2010)
28. VERBI Software: MAXQDA: Software for qualitative data analysis (2024), <https://www.maxqda.com/de/>, accessed: 2024-11-06
29. Wongsuphasawat, K., Liu, Y., Heer, J.: Goals, process, and challenges of exploratory data analysis: An interview study. *arXiv preprint arXiv:1911.00568* (2019)
30. Yeshchenko, A., Mendling, J.: A survey of approaches for event sequence analysis and visualization. *Information Systems* **120**, 102283 (2024)
31. Zerbato, F., Soffer, P., Weber, B.: Initial insights into exploratory process mining practices. In: International conference on business process management. pp. 145–161. Springer (2021)
32. Zerbato, F., Zimmermann, L., Vrotsou, K., Weber, B.: From analysis to findings: How do process mining analysts discover results? *Information Systems* p. 102596 (2025)
33. Zimmermann, L., Zerbato, F., Weber, B.: Visualization use during process mining analysis – appendix (2026). <https://doi.org/10.5281/zenodo.19553523>
34. Zimmermann, L., Zerbato, F., Vrotsou, K., Weber, B.: A process-oriented approach to analyze analysts’ use of visualizations: Revealing insights into the what, when, and how. In: Computer Graphics Forum. p. e70104. Wiley Online Library (2025)