

Designing Process Diagrams – A Framework for Making Design Choices when Visualizing Process Mining Outputs

Marit Sirgmetts¹, Fredrik Milani¹, Alexander Nolte¹, and Taivo Pungas²

¹ University of Tartu, Estonia

² Datamob, Tallinn, Estonia

{marit.sirgmetts, milani, alexander.nolte@ut.ee}

{taivo@pungas.ee}

Abstract. Modern information systems can log the executions of the business processes it supports. Such event logs contain useful information on the performance and health of business processes. Event logs can be used in process analysis with the aid of process mining tools. Process mining tools use various diagrams to visualize the output of analysis made. Such diagrams support the visual exploration of the event logs, facilitating process analysis, and usefulness of process mining tools. However, designing such diagrams is not an easy task. Often-times neither the developer nor the end-user know how to visualize the outputs created by process mining algorithms, nor do they know where the interesting information is hidden. Designing diagrams for process mining tools require taking design decisions that, on the one hand allow flexible exploration, and on the other hand, are simple and intuitive. In this paper, we investigate how existing process mining outputs are visualized and their underlying design rationale. Our analysis show that process diagrams, the most common type of diagrams used, are designed with next to no guidance from data visualization principles. Based on our findings, we propose a framework to support developers when designing visualization for process mining outputs. The framework is based on data visualization theory and practices within process mining visualization. The effectiveness and usability of the framework is tested in a case study.

Keywords: process mining, process visualization, process diagram design, framework.

1 Introduction

Data is an essential resource for organizations. The ability of a business to analyze and interpret data and make informed business decisions based on analysis results is crucial for a company to survive. The availability of vast amounts of data from various resources makes it hard for analysts to identify interesting or problematic aspects. Visualizations are crucial for process analysts and decision makers to explore and analyze complex data and make informed decisions. Such visualizations are however often-times created by developers with little knowledge about how to design useful and easy to use visualizations.

One area which is facing this challenge is process mining [1]. In process mining, knowledge is retrieved from execution logs and analyzed from different perspectives such as control flow, resources, and data [2]. Designing visualization of process mining outputs is a complex task [1]. Developers are faced with numerous design questions when composing process diagrams where little or no visualization standards exist. They are also oftentimes left alone when making critical decisions about how to design such visualizations since there is a lack of proper support and guidance. Useful tips are scattered in various resources that process mining developer might or might not be aware about. Moreover, most available resources are presented in a generalized way covering aspects of holistic diagram design [3–5]. This makes such approaches hard to use for developers in the context of process mining due to the necessity to adjust them to the specifics of this domain. Such adjustments are also time-consuming and difficult for developers, who are not professional designers.

For this work we followed a design science approach [6]. We address a real-life problem creating a framework for process mining developers, who are tasked with designing process diagrams. The framework identifies common design issues when visualizing process maps and proposes ideas for solutions. It includes topics related to visual encoding as well as interaction. The framework is based on process mining visualization practices and data visualization theory that is adjusted to process mining. We then conduct a formative evaluation in a real-life project to assess the feasibility of our proposed framework and to identify means for improvement. For the evaluation we recruited developers, who were designing a process diagram as a part of developing a process mining tool. Results of this evaluation point towards the usefulness of this framework for developers of process maps and provide hints for its improvement.

This paper is organized as follows. Section 2 presents the state of the art of visualization within the field of process mining. Section 3 presents our proposed framework. Next, Section 4 presents the evaluation of the framework while Section 5 concludes the paper.

2 State of the Art

We conducted a review to answer three research questions. The first is “which process mining techniques use visualization”. The second is “how are current process mining techniques visualized” and finally, “how do developers decide on how to design visualizations”. We searched for related articles on Scopus and Web of Science using the keywords “visual”, “process”, and “mining”. These electronic libraries were chosen as they constitute the main venues for publication within the field of process mining.

The search yielded more than 2000 results which was filtered in three rounds. In the first round, duplicates were removed. In the second round, papers clearly out of scope, such as those on coal mining or data mining, were excluded. The remaining papers were examined and filtered based on the following criteria. Papers less than 3 pages, not accessible, not in English, or older than 10 years were excluded (exclusion criteria). The remaining list of papers further examined and included if they fell within the domain of process mining, introduced a visualization technique, and mentioned design

choices for visualization (inclusion criteria). The final list consisted of 28 papers¹. Data about the paper (meta data), process mining technique, proposed visualization, platform or tool where visualization is implemented, design process of the visualization, and evaluation of the visualization were extracted from each paper.

2.1 Process Mining Techniques using Visualization

To answer the first research question, we examined the extent to which visualization is used to communicate the output of different process mining techniques. Our review showed that visualization was mostly used for process discovery (generating process models from event logs), process performance (measuring cost, time, and quality aspects of process executions), and process comparison (comparing several processes or checking a model against its event log). Visualization was also used for predictive monitoring (predicting future outcome or upcoming execution paths of a process instance), organizational mining (discovery of organizational structures and communication between units), model repair (improving discovered models based on event logs), deviance mining (uncovering causes of deviant executions of a process), compliance monitoring (surveillance of compliance or violations against regulations in the process execution), and concept drift (changes in the process execution over time). We did not identify any studies visualizing process optimization (identifying improvement opportunities) or process decomposition (clustering models into high-level functions).

A total of 13 papers used visualizations for only one process mining output (single purpose visualization). A few notable examples are process comparison [7, 8], organizational mining [9, 10], performance analysis [11], predictive monitoring [12], and deviance mining [13]. Several studies (a total of 10) concurrently visualized several outputs. An example is the InterPretA tool [14] that visualizes outputs from deviance mining and performance analysis. Another example is “Event Streamer” [15] that visualizes both discovery of declarative processes and concept drift. The remaining studies (5 papers) did not mention any specific process mining technique. Instead, they proposed methods for general exploration of process logs. For example, the tool Event Explorer [1] can be used when an analyst does not know a priori, which specific analysis technique to select.

2.2 Visualization of Process Mining Techniques

The second research question aimed at identifying how process mining outputs are visualized. Our review revealed that node-link diagrams is the most commonly used method for visualizing process mining outputs. For instance, node-link diagrams were used to show the relationship between activities (process diagram) [7, 16] or connections between resources (social network diagrams) [9, 10]. The second most common type of diagrams used were bar-, pie-, and line charts. Such diagrams were often used for visualizing process performance [17, 18]. Performance was also visualized using

¹ The list of papers is available at https://babook.cs.ut.ee/pmviz_framework/

box plots for value distribution [14] and gauge charts [19]. Hierarchical process relationships, such as medical treatment processes and their sub-processes, were commonly visualized using tree maps [1]. Scatterplots were used to visualize correlations of care-process parameters, such as correlation of number of treatment activities and patient’s length of hospital stay [18]. Other chart types used were stream graphs [2] for visualizing live process instance flows and turtle graphic trace map [20] for detecting flow differences amongst process variations.

The charts used followed prevalent and conventional styles. For instance, when using stacked bar charts, the length represented the value and the color hue distinguished the sub-groups [14]. In the case of node-link diagrams, we noted a greater variability in how outputs were visualized. The variability was expressed by unique combinations of visual and interactive elements. In addition to portrayal of the base topology, other visual channels such as shapes, colors, and sizes were utilized to represent additional data elements. The extend of the variability seems to indicate a lack of visualization standards. In summary, we identified eleven different types of diagrams (not necessarily complete) used in the 28 studies reviewed (see Table 1).

Table 1. Diagrams Used for Process Mining Visualization

Diagram Type	Number of Studies	References
Node-link diagram	24	[1, 7, 18–27, 8, 28–31, 9–11, 14–17]
Bar/triangle chart	7	[12, 16, 18, 19, 22, 25, 30]
Pie chart	4	[11–13, 17]
Line/area chart	3	[1, 14, 15]
Tree-map	2	[1, 18]
Scatterplot	2	[13, 18]
Parallel coordinate plot	1	[32]
Box plot	1	[14]
Gauge chart	1	[19]
Instance stream graph	1	[15]
Turtle graphics trace map	1	[33]

2.3 Methods for Visualizing Process Mining Outputs

The noted variety of visualization designs prompted us to identify how design choices were reached. We noted that most papers focused on the proposed algorithm and as such, presented the algorithm outputs without presenting a rationale for design choices taken. Although the identified studies did not explicitly follow a systematic method, they drew inspiration and used input from mainly four sources. These sources are (1)

existing practices, (2) domain expert input, (3) visualization theory, and (4) argumentation.

The first input source refers to critical analysis based on a reviewing process mining related literature and tools. An example is Bachhofner et.al. [11] who noted that existing solutions only visualize one performance metric on process diagrams. To address this limitation, they proposed a tool that allows for concurrently representing several performance metrics. Domain expert input refers to cases where design choices were based on real-life task requirements or user feedback. A notable example is a tool specifically built for users in a hospital by Basole et.al. [18]. The involved domain experts provided feedback to the proposed visualizations. This iterative process resulted in the first versions of visualization being discarded. One study stood out as it employed a systematic design framework based on visualization theory. Wynn et.al. [16] modified design science methodology by using process mining knowledge, visualization principles, and evaluation of visualization as input for design choices. For instance, Wynn et.al. [16], in using size of diagram elements to express continuous variables, grounded this decision in research conducted by Moody [34]. The most common rationale for design choices however, is argumentation. Arguments behind design decisions were generally along the lines of “by watching the displays’ content and simultaneously performing selection on the business process model, ...differences in the selected sets of data become intuitively visible...”. [19] or “we chose this representation because it makes comparisons more natural for the user” [14]. The argumentation was not grounded in common practices, supportive theory, or the result of comparing alternative choices. One could deduce that the arguments were somewhat arbitrarily chosen.

2.4 Summary

Our literature review has shown that most process mining techniques use visualization to present their output. The process mining use cases not using visualization are decomposition and optimization. Decomposition relies on algorithmically re-structuring of processes and thus does not require visualization. Optimization is commonly based on metrics where weaknesses in existing process executions are identified. Such weaknesses might not require specific visualization. Nevertheless, it thus appears that visualization is an integral part of most process mining techniques.

Our review also revealed that a variety of diagram types are used to visualize process mining outputs. The most common is by means of node-link diagrams. The listing of various visual and interactive elements overlaid with node-link diagrams however, seem to indicate a lack of standard or structured way of making design choices. Our review also showed that developers of process mining techniques did not employ a systematic or structured method when making design choices. Design choices are rather oftentimes reached arbitrarily. This is somewhat surprising considering the crucial role visualization plays when exploring and analyzing complex data. Taken together, these results reveal a gap in the visualization design practices within the process mining field. There is thus a need for a specifically tailored visualization framework that supports developers to design useful visualizations for the output of process mining techniques.

3 Framework

This chapter describes our proposed framework for guiding developers of process mining techniques in composing process diagrams. The first part sets the foundation of the framework. The second part moves on to describe its development process and the final part presents the structure and content of the framework.

3.1 Foundation

We propose a framework that is specifically tailored for process mining techniques. The framework serves to support a developer when designing diagrams, it is not a tool that offers suggestions when given requirements as an input. The primary audience is thus developers of process mining algorithms, who do not have professional experience in the design field. The framework aims to aid the aforementioned developers in making informed design decisions. Hence, the output of the framework is a set of decisions options that a developer can decide on to compose a visualization, not a ready-made composition or a mock-up. It should be noted that in this context, design refers to the structure (requirements) of the visualization of process mining output and not for instance its appearance.

The framework focuses on process diagrams because our review showed that they are the most prolific type of diagram. Most process mining techniques require an understanding of the topology of processes, which is usually supported by node-link diagrams, i.e. process diagrams. Moreover, far too little attention has been given to the design of the visualization of process mining tools, resulting in limited guidance in designing process diagrams [1].

Our framework is based on the literature analysis describe in section 2. While the identified studies provided a plethora of aspects to consider, they did not provide a sufficient foundation to shape a framework. There is, therefore, a need for a foundational data visualization theory to build upon. To this end, we chose Munzner’s visualization theory [3] for two main reasons. First, Munzner’s work [3] proposes an overarching framework for designing and analyzing data visualization. The framework considers all aspects of the visualization process, from domain and data analysis to validation. Furthermore, the core of Munzner’s work, how to visualize data, is well aligned with our purpose. Secondly, Munzner’s work is based on well-accepted academic work on data visualization theory (c.f. [35] [36] [37]). Other frameworks such as those by Few [38], Ware [39], Cairo [40], Wilkinson [35], and Tufte [36] were considered but found not appropriate for our purpose. They either mostly focused on dashboards [38], considered presentational rather than explorative data visualization [40], focused on how visualization is perceived [39], or discussed theoretical foundations rather than practical implementation of visualization [35, 36]. Munzner’s work in contrast is user-centric, considers representation and interaction, is systematically categorized and organized, and addresses specifics of network data and node-link diagrams.

Munzner presents the data visualization process as a nested model where the output of one layer serves as an input to the next. Munzner considers four layers, domain situation, data/task abstraction, visual encoding and interaction idiom, and algorithm [3].

The question of “how to visualize”, which is the focus of our framework, lays in the third layer – visual encoding and interaction idiom. Munzner breaks this part into several questions which are then decomposed further into additional sub-questions. Together, the questions form a hierarchical design tree for design choices [3].

The structure of Munzner’s framework is generic. It can therefore be used for a wide range of data visualization cases. This characteristic of the framework enables designers, who aim to expand their awareness of different visualization possibilities, to explore data visualization for a multitude of contexts. However, its generic nature makes it unsuitable for developers who face design choices when visualizing process diagrams. The wide spectrum and the vast materials to consult when designing process diagrams, will most likely be more confusing than constructive for a developer. To address this limitation, we propose a framework that is adjusted and specialized for the context of designing process diagrams within the domain of process mining techniques.

3.2 Development

The framework was developed in three steps. The first step was to identify questions that should be considered when designing process diagrams and set them into a logical sequence. During the second step, the questions were enriched with alternative answers. The third step addressed understandability aspects of the framework. For this step, we developed illustrations to improve the understandability of the used concepts and terminology.

The aforementioned design questions were extracted from Munzner’s theory [3] and mapped against process mining visualization practices. We only included questions that were applicable to process mining techniques. Complementary questions were included where Munzner’s theory failed to cover design aspects essential to process mining techniques. For instance, most process flow diagrams within process mining are directed whereas Munzner’s theory does not cover directed node-link diagrams sufficiently. Therefore, we added for instance, the question of “how is the sequence of the process shown?” This question is derived from process mining practice and not from Munzner’s visualization theory. The final selection contained 62 questions which were considered relevant for our framework.

We then structured the questions using a top-down approach. We identified two main areas – encoding and interaction. Each of these areas was divided into two subcategories. Encoding was divided into *arrange* and *map* and interaction was divided into *reduce* and *change* (see Fig. 1). The remainder of the questions were structured along these four subcategories. After dividing the questions, we identified the dependencies between the questions i.e., one question cannot be answered before some other decisions have already been made. For example, the decision about the basic elements of a diagram must be taken before designing the details of that same diagram. These dependencies defined the sequence and hierarchy of the questions. In cases where there did not exist dependencies, the questions were ordered according to Munzner’s visualization theory [3].

In the second step, the questions were enriched with alternative options. When a developer has to take design decisions, they are not served with alternative options. To

address this limitation, our framework proposes alternative solutions for each question. The alternative solutions were extracted from the visualization theory we selected as foundation [3]. For instance, the first question in the framework is “what is the base diagram?”. Munzner lists three alternative solutions to this question, which are all included in our framework – node-link diagram, adjacency matrix or enclosure [3]. If Munzner’s theory did not provide suitable options for the process mining context or if options were missing, we drew examples from process mining diagrams to identify suitable options. An example is the answer to the question of “where does the embedded data appear?”. The options added are pop-up window or pane that appears on the diagram itself, covering parts of it [17], or in a separate area next to the diagram [14]. When required, we searched for additional supporting theoretical material. For instance, the options to the question of “how are the basic elements ordered” were taken from Colligan et.al. [41] who conducted a comparative study on the effectiveness of hierarchical versus sequential visualization of care-processes. The strengths and weaknesses were extracted together with the specific answers from the visualization theory or inspired by general principles from the theory [3]. In cases, where dualistic pros and cons were irrelevant, common practices with brief reasoning extracted from the literature study were listed instead of theoretical trade-offs.

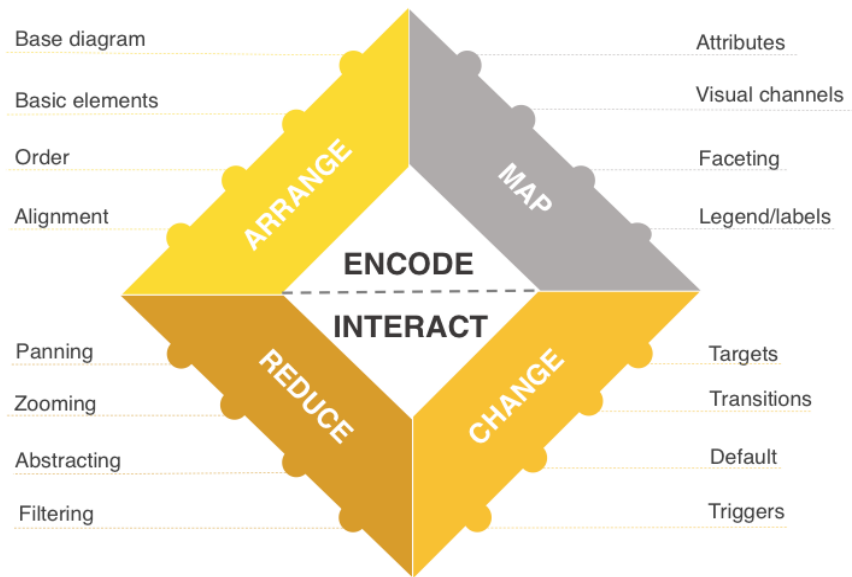


Fig. 1. Reference model of the framework.

The last step in the development of our framework aimed at improving the comprehension of the framework. This was achieved by adding visual illustrations. All visual

illustrations were inspired by examples drawn from state of art studies or data visualization theory. For example, the illustration next to the question of how to solve occlusion in animations was inspired by the work of de Leoni et.al. [17], who propose a process animation tool called Log On Map Replayer (see Fig 2). Examples were also inspired by modelling languages, such as using the pool and lane concepts commonly known from BPMN [42] to give an example of the use of spatial region in process mining (see Fig 3).

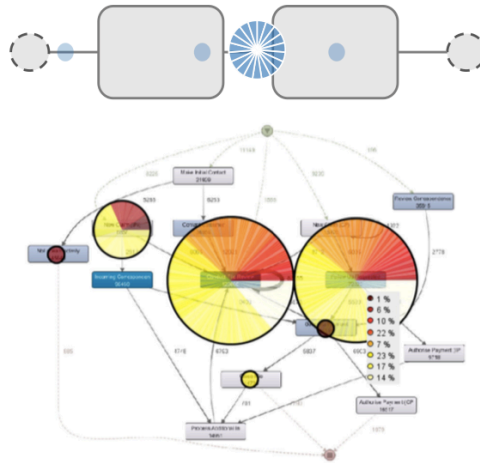


Fig. 2. Illustrations in the framework were inspired by existing tools and visualizations. An illustration for a question about solving occlusion in an animation (upper image) was inspired by Log On Map Replayer solution (lower image) [17].

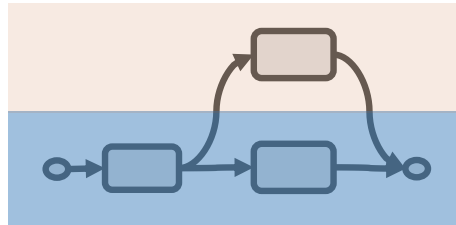


Fig. 3. Illustrations used in the framework were also inspired by modeling languages, such as illustrating the use of spatial region as employed in BPMN [42].

3.3 Overview²

In this section, we introduce the structure and main contents of the framework. The main contents of the framework are illustrated in Fig 1. The model should be read from inside out as the topics are in a hierarchical order. The topics are divided into two building blocks – “encoding” and “interaction”. Encoding contains questions about visual

² The framework and detailed instructions on how to use it are available at <https://owncloud.ut.ee/owncloud/index.php/s/1QxL4P1Iq2Z9ofN>

aspects of the diagram while interaction covers questions about how to manipulate the diagram. In encoding, the developer chooses to visualize the frequency and duration of process activities as two separate layers of a process diagram. In interaction, the developer decides how the user of the diagram can switch between these two layers.

Encoding is further divided into two sub-topics – “arrange” and “map”. Arrange covers questions on the basic structure of the diagram, such as which type of base diagram to use as well as the ordering and alignment of diagram elements. For instance, when encoding, a developer decides on using a node-link diagram where nodes (activities) are ordered sequentially. The direction of the flow of activities is also determined. The direction can be from left to right with a start event placed on the left-hand and the end event on the right-hand side of the diagram. When mapping, the focus is on the aesthetics of the diagram. In mapping, decisions are on which attributes (such as color and shape) are used in a visualization. For instance, a developer might wish to visualize the frequency and duration of the process execution with color saturation – e.g. the darker the shade of the node color, the higher the number of process instances or duration. As two attributes, frequency and duration, are shown on the same diagram, the mapping also guides the developer on faceting the diagrams. In this example, the user can switch between views, one for frequency and one for duration. Using only color saturation is not enough to convey what the encoding means. To facilitate understanding, mapping also contains questions on legends and labels.

Interaction also consists of two sub-topics – “reduce” and “change”. Reduce refers to which data the user can choose to be visualized. The user can e.g. use zoom and pan (scroll) to highlight specific aspects of interest. Filtering and abstraction allow for select subsets of the dataset to be visualized. Change on the other hand, refers to changing the diagram. Change considers what the user can change, how the changes transition from one image to another, and which user actions trigger changes. In mapping, developers chose to facet duration and frequency as two separate versions of the same diagram. In “change”, a developer can determine “how” by considering which is to be the default view and how to switch between the views.

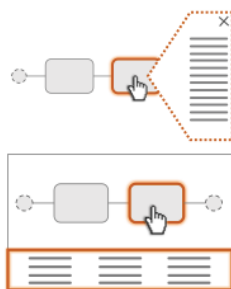
The reference model depicted in Fig. 1 illustrates the main topics covered in the framework. The full framework consists of a set of questions, categorized according to the main components of the reference model. The questions are structured according to the topic they belong to. For instance, the question of “what is the base diagram” is part of “arrange” which in turn, is under “encode”. As such, the framework systematically guides the developer when visualizing process mining outputs through a set of questions. Table 2 provides an example for questions and their structure. For example, on the reference model, encoding is divided to two parts – “arrange” and “map”, which corresponds to the question (level 1) of “how to encode data”. This question is then further divided to two questions (level 2), namely “how to arrange data” and “how to map data”. At the third level, the questions are broken down into detailed sub-questions. The framework also provides further considerations for each alternative answer of the detailed sub-questions questions.

In Fig 4, we illustrate an example of alternative answers to a sub-question at the fourth level, namely for the question of “where does the embedded data appear”. The available alternatives are “on the diagram” and “off the diagram”. The framework also

indicates further consideration (strengths and weaknesses) of each alternative. Embedding the data in the diagram makes it easier to track but requires space which might occlude other relevant parts of the diagram. Also, if it is important to see the diagram while drilling into detailed level of data, “off the diagram” might be a better alternative. “Off the diagram” refers to presenting detailed information on a separate pane which does not cover the diagram. The downside is the space required for the pane will come at the expense of the space allocated for the diagram. Thus, the framework provides the questions and relevant considerations for each of alternative solutions (Fig. 4).

Table 2. Hierarchy of the Questions in Process Visualization Framework

Level 1	Level 2	Level 3
How to encode data?	How to arrange data?	What is the base diagram? What are the basic elements of the diagram? How are the basic elements ordered? How is the diagram aligned?
	How to map data?	Which attributes are shown on the diagram? Which channels express the attributes? How is the data faceted on the diagram? How does the user know the meaning of the channels?
How to design interaction?	How can the user change the visualization?	What can be changed on the diagram? How do the changes appear? What is the default appearance? How can the changes be triggered?
	How can the user reduce data?	Does the diagram need panning? Does the diagram need zooming? Does the diagram need abstracting? Does the diagram need filtering?



1.2.3.7 Where does the embedded data appear?

On the diagram

+ element and embedded data are close – easy for eyes to track.	- pop-up windows occlude parts of the base diagram.
--	---

Off the diagram

+ the full process is in the view when the details are shown.	- space-consuming.
---	--------------------

Fig. 4. An example of lowest level of question and its considerations of the framework.

4 Evaluation

In order to evaluate the usefulness of the framework we conducted a case study in a real-life project. The aim of the evaluation was to identify strengths and weaknesses of the framework as well as means for further improvement. In the following we will describe the design of the evaluation (Section 4.1) before discussing our findings.

4.1 Design

The evaluation was designed as a case study since we aimed to explore its usefulness in a real-life context [43]. Case studies are suitable for answering “why” and “how” questions, particularly in cases where context can provide insightful information and the research requires an observational approach [44].

The aim of the framework is to support developers to create visualizations of process mining diagrams, specifically process maps. It thus has to fulfil the following four main criteria: (1) it has to be understandable by developers of such visualizations; (2) it has to be relevant in the given context; (3) it has to be complete in order to support developers to create visualizations that are useful to and useable by process analysts and (4) it has to be easy to use by the target audience to ensure a balance between the time and effort developers spend using the framework when designing visualizations. The evaluation thus focused on perceived understandability, relevance, completeness, and usefulness of the framework.

The unit of the analysis was defined as follows:

- The effect of the framework on data visualization design tasks executed by developers of process mining tools.

The effect was observed through the lens of the following questions that address the understandability, relevance, completeness and usefulness:

- How is the framework understandable/unclear for developers?
- How is the framework relevant/irrelevant for the process of designing visualizations for process mining diagrams?
- Which aspects are potentially missing from the framework?
- How easy is it to use the framework?

The case study took place in the context of a project that aims to visualize data from a queuing management system used to manage border crossing. A group of developers were building a process mining tool that would help to translate data from aforementioned queuing management system into insightful information to improve and innovate the queuing process. The focus of development was process discovery, performance analysis, predictive monitoring, and deviance mining. The developers of the tool used the framework to design visualizations for process diagrams of the tool the team was developing.

We chose three members (See Table. 3) as participants in the case study – a data scientist and two researchers. All participants have had experience in developing process mining techniques. The data scientist (P1) was currently working on a PhD thesis in process mining field and has worked on industry projects related to process mining. Both of the researchers have about 10 years of experience in developing tools in the context of process mining. In addition, all the participants have had previous experience in data visualization. The first participant had been using data visualization mostly for presentation purposes. The second participant (P2) had become acquainted with data visualization concepts through practice as well as theory. The third participant (P3) had developed process mining tools that include visual presentations – some lectures s/he holds require familiarity with the data visualization literature. None of the participants were professional visualization designers.

Table 3. Case Study Participants

ID	Experience	Field	Project Focus
P1	4 years	Industry Projects Academic Research	Predictive Monitoring
P2	10+ years	Academic Research	Process Discovery and Performance Analysis
P3	10 years	Academic Research Teaching	Deviance Mining

The procedure of the study was conducted as follows. Each participant was invited to an individual session, which was divided to three parts. The session started with a semi-structured interview during which the participants were asked to explain the project they were working on and their role in the project in more detail. They were also asked to explain potential issues they were facing during the project. After the interview the participants were asked to explain their initial visualization ideas before they were introduced to the framework. The participants were then asked to use the framework – which they received in a printed form – for their respective visualization task. The final part of each session was a semi-structured interview during which we asked the participants for their opinions about the framework and their perspective on the impact it had on their visualization task. Each session was audio recorded and the researcher conducting the studies took additional field notes.

4.2 Discussion

The following results are structured along the aforementioned evaluation dimensions of understandability, relevance, completeness, and usefulness.

Understandability. All participants found the framework to be understandable (“*Definitely it was easy*”, P2). This was evident by the participants not struggling with aspects, such as which questions to answer or the meaning of illustrations or tables. Also, all participants were able to identify the purpose of the framework (“*to get a better understanding, to formulate a visualization task better*”, P1). The participants reported

the framework to be useful for tool improvement, making vague visualization ideas more concrete and using it as an inspiration point for designing new visualizations. One participant referred to it as a catalogue of tested ideas (*“sort of a catalogue with some already tested practices”*, P2), which can be revisited several times during the design process. Another participant saw its use in user surveys to identify the solution that the target users would prefer. Two participants pointed out the potential to develop the framework into a mock-up tool, which would turn answers to questions in the framework into sample visualization.

Even though, the basic understandability of the framework was good, all the participants highlighted aspects that could be improved. Two participants found parts of the terminology to be confusing. One participant found the terms easy but added that this is due to his familiarity with literature on visualization theory. One participant suggested a glossary (*“maybe it would be helpful if somewhere were those [definitions], so that they can be immediately looked up”*, P1), where the terms could be easily looked up. Another issue mentioned by two participants was targeting of the. Both mentioned two potential reasons for their issues related to the understandability of the questions: The wording of the questions – it was not clear if the questions are about existing solutions or prospective preferences. For example, “How is the diagram aligned?” refers to something that already exists, while wording such as “How would you like the diagram to be aligned?” is aimed for the designer to think about how s/he would design a future diagram. The second potential reason for aforementioned confusion was the sequence of topics and transitions (*“sometimes it is difficult to follow the sequence of questions”*, P3). The questions move from one topic to another with abrupt transitions and the user may miss that the target of the question has switched.

Relevance. The participants found the framework relevant for their work in particular related to developing new ideas or improving and clarifying existing ideas of process mining visualizations. The purpose of the framework was easy to understand for the participants and it was found relevant for all process mining visualization tasks which include tool improvement as well as inspiration and a guidance for making ideas more concrete. All participants mentioned that the framework helped them to develop new ideas for their respective visualization task and they would recommend it to their colleagues who struggle with similar tasks (*“yes, I think it helps to put ideas together, especially in the initial stage of development”*, P1).

Completeness. Most of suggestions for adding and changing the framework stem from aforementioned understandability issues (*“I don’t know how complete the idea gets, maybe you can add even more alternatives”*, P3). For instance, examples of real tools and a glossary of definitions were suggested to improve the clarity of the framework. Also, the transitions between topics were brought out as a potential place to improve the comprehensibility of the framework. One participant suggested a solution for clarifying the targets of the questions by reducing the topics in the framework. For example, focusing on one of the main topics – representation of data or interactivity – and allowing users to explore the selected topic in more depth, while discarding the other. One

participant saw a possibility to include more questions specifically about embedded data – how to visualize data that is shown in the pop-up windows.

All participants also saw a potential to digitalize the framework (“*if the framework was digital, then it would be very comfortable to see the final result*”, P1). Two participants mentioned the potential to develop the questions in the framework into a mock-up tool that could show an example diagram based on the selected alternatives. One participant suggested to hide the positive and negative aspects in the default view and provide the user with a respective option that allows him to reveal them if necessary.

Usefulness. One participant estimated the level of required focus high, while two others thought it required little effort to use it. The participants also mentioned that terms and targets of questions were the most difficult to understand. The time required for going through the framework on paper varied from 25 to 45 minutes. However, two participants mentioned that it should be used repeatedly during the visualization development process. All participants thought that time and effort they put into using the framework was worthwhile. The framework helped them to develop new visualization ideas and make existing ones more concrete in a relatively short time.

4.3 Limitations

The evaluation of the framework had a number of limitations. First of all, the framework was evaluated on experienced developers with familiarity of visualized outputs from process mining techniques. As such, the framework might require further instructions for novice developers. Furthermore, the evaluation did not include representatives of intended end-users of the visualization such as process analysts. Also, the evaluation focused on developing new node-linked diagrams. Although visualization within this field predominantly uses node-linked diagrams, the suitability of the framework for other types of base diagrams was not covered in the evaluation.

5 Conclusion

During the course of this paper we presented the development and evaluation of a framework that supports developers in designing process mining diagrams. Our work showed the importance of visualizations in process mining field and revealed the complexity of the design tasks developers are facing. Regardless of the importance and complexity of the visualizations, most of diagrams are currently designed by developers with little to no training in developing visualizations and with no systematic support. Design decisions are instead often based on a combination of logical argumentation, existing practices and domain input.

The proposed framework is based on two cornerstones – existing process mining visualizations and data visualization theory. Majority of the topics covered in the framework have their foundation in the visualization theory forwarded by Munzner [27]. However, adjustments were made to the theory to make it relevant to process mining.

The framework consists of questions and alternative answers with strengths and weaknesses. In addition, illustrations that are specific to process mining, were designed and added to the framework to increase the comprehensibility through visual examples.

We evaluated the framework in a case study with three developers. The evaluation revealed that the developers found the framework relevant and balanced in terms of how much effort it requires and how beneficial it is to the task at hand. The main value of the framework was found in making vague ideas concrete, coming up with new ideas, and improving existing ones. The evaluation also revealed potential means for improvement such as clarification of terms.

In the future, we aim to extend the framework to other types of visualizations such as dashboards used in process mining. Another possible venue is improving the format of the framework by developing an online tool.

6 References

1. Gschwandtner, T.: Visual Analytics Meets Process Mining: Challenges and Opportunities. *Data-Driven Process Discov. Anal. SIMPDA 2015*, B. Ser. Lect. Notes Bus. Inf. Process. 244, 142–154 (2017).
2. van der Aalst, W.M.P.: *Process Mining*. Springer-Verlag Berlin Heidelberg (2016).
3. Munzner, T.: *Visualization Analysis and Design*. AK Peters/CRC Press, Boca Raton (2014).
4. Rogers, Y., Sharp, H., Preece, J.: *Interaction Design: Beyond Human - Computer Interaction*. John Wiley & Sons Ltd, West Sussex, UK (2011).
5. Unger, R., Chandler, C.: *A Project Guide to UX Design: For user experience designers in the field or in the making*. New Riders, Berkeley (2012).
6. Bichler, M.: Design science in information systems research. *WIRTSCHAFTSINFORMATIK*. 48, 133–135 (2006).
7. Gall, M., Wallner, G., Kriglstein, S., Rinderle-Ma, S.: Differencegraph - A ProM Plugin for Calculating and Visualizing Differences between Processes. *CEUR Workshop Proc.* 1418, 65–69 (2015).
8. Bolt, A., de Leoni, M., van der Aalst, W.M.P.: A Visual Approach to Spot Statistically-Significant Differences in Event Logs Based on Process Metrics. *Adv. Inf. Syst. Eng. (CAISE 2016)*, B. Ser. Lect. Notes Comput. Sci. 9694, 151–166 (2016).
9. Slaninova, K., Martinovic, J., Drazdilova, P., Snashel, V.: From Moodle Log File to the Students Network. *Int. Jt. Conf. SOCO '13-CISIS'13-ICEUTE'13*, B. Ser. Adv. Intell. Syst. Comput. 239, 641–650 (2014).
10. Jalali, A.: Supporting Social Network Analysis Using Chord Diagram in Process Mining. *Perspect. Bus. Informatics Res. BIR 2016*, B. Ser. Lect. Notes Bus. Inf. Process. 261, 16–32 (2016).
11. Bachhofner, S., Kis, I., Di Ciccio, C., Mendling, J.: Towards a Multi-parametric Visualisation Approach for Business Process Analytics. *Adv. Inf. Syst. Eng. (CAISE 2017)*, B. Ser. Lect. Notes Bus. Inf. Process. 286, 85–91 (2017).

12. Jorbina, K., Rozumny, A., Verenich, I., Di Fancescomarino, C., Dumas, M., Ghidini, C., Maggi, F.M., La Rosa, M., Raboczi, S.: Nirdizati: A Web-based Tool for Predictive Process Monitoring. *CEUR Workshop Proc.* 1920, (2017).
13. Bose, R.P.J.C., van der Aalst, W.M.P.: Discovering Signature Patterns from Event Logs. *Proc. 2013 IEEE Symp. Comput. Intell. Data Mining, CIDM 2013 - 2013 IEEE Symp. Ser. Comput. Intell. SCI 2013.* 111–118 (2013).
14. Dixit, P.M., Caballero, H.S.G., Corvo, A., Hompes, B.F.A., Buijs, J.C.A.M., van der Aalst, W.M.P.: Enabling Interactive Process Analysis with Process Mining and Visual Analytics. *Proc. 20th Int. Jt. Conf. Biomed. Eng. Syst. Technol.* 5, 573–584 (2017).
15. Burattin, A., Cimitile, M., Maggi, F.M.: Lights, camera, action! Business process movies for online process discovery. *Lect. Notes Bus. Inf. Process.* 202, 408–419 (2015).
16. Wynn, M.T., Poppe, E., Xu, J., ter Hofstede, A.H.M., Brown, R., Pini, A., van der Aalst, W.M.P.: ProcessProfiler3D: A visualization framework for log-based process performance comparison. *Decis. Support Syst.* 100, 93–108 (2017).
17. de Leoni, M., Suriadi, S., ter Hofstede, A.H.M., van der Aalst, W.M.P.: Turning event logs into process movies: animating what has really happened. *Softw. Syst. Model.* 15, 707–732 (2016).
18. Basole, R.C., Park, H., Gupta, M., Braunstein, M.L., Chau, D.H., Thompson, M.: A Visual Analytics Approach to Understanding Care Process Variation and Conformance. *ACM Int. Conf. Proceeding Ser.* 6, (2015).
19. Gulden, J., Attfield, S.: Business Process Models for Visually Navigating Process Execution Data. *SAP Bus. Process Manag. Work. B. Ser. Lect. Notes Bus. Inf. Process.* 256, 583–594 (2016).
20. Cordes, C., Vogelsang, T., Appelrath, H.-J.: A Generic Approach for Calculating and Visualizing Differences Between Process Models in Multidimensional Process Mining. *Bus. Process Manag. Work. BPM 2014, B. Ser. Lect. Notes Bus. Inf. Process.* 202, 383–394 (2015).
21. Lucas, W., Xu, J., Babaian, T.: Visualizing ERP Usage Logs in Real Time. *ICEIS Proc. 15th Int. Conf. Enterp. Inf. Syst.* 3, 83–90 (2013).
22. Pini, A., Brown, R., Wynn, M.T.: Process Visualization Techniques for Multi-perspective Process Comparisons. *Asia Pacific Bus. Process Manag. AP-BPM 2015, B. Ser. Lect. Notes Bus. Inf. Process.* 219, 183–197 (2015).
23. Mannhardt, F., de Leoni, M., Reijers, H.A.: The Multi-perspective Process Explorer. *CEUR Workshop Proc.* 1418, 130–134 (2015).
24. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P.: Exploring Processes and Deviations. *Bus. Process Manag. Work. BPM 2014, B. Ser. Lect. Notes Bus. Inf. Process.* 202, 304–316 (2015).
25. Vogelsang, T., Appelrath, H.-J.: Multidimensional Process Mining with PMCube Explorer. *CEUR Workshop Proc.* 1418, 90/94 (2015).
26. Hipp, M., Strauss, A., Michelberger, B., Mutschler, B., Reichert, M.: Enabling a User-Friendly Visualization of Business Process Models. *Bus. Process Manag. Work. BPM 2014, B. Ser. Lect. Notes Bus. Inf. Process.* 202, 395–406

- (2015).
27. Slaninova, K., Vymetal, D., Martinovic, J.: Analysis of Event Logs: Behavioral Graphs. *Web Inf. Syst. Eng. - WISE 2014 Work. B. Ser. Lect. Notes Comput. Sci.* 9051, 42–56 (2015).
 28. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P.: Process and Deviation Exploration with Inductive visual Miner. *CEUR Workshop Proc.* 1295, 46–50 (2014).
 29. Kecman, P., Goverde, R.M.P.: Process mining of train describer event data and automatic conflict identification. *WIT Trans. Built Environ.* 127, 227–238 (2012).
 30. van Dongen, B.F., Adriansyah, A.: Process Mining: Fuzzy Clustering and Performance Visualization. *Bus. Process Manag. Work. BPM 2009, B. Ser. Lect. Notes Bus. Inf. Process.* 43, 158–169 (2010).
 31. Gall, M., Wallner, G., Kriglstein, S., Rinderle-Ma, S.: A Study of Different Visualizations for Visualizing Differences in Process Models. *Adv. Concept. Model. B. Ser. Lect. Notes Comput. Sci.* 9382, 99–108 (2015).
 32. Gupta, N., Anand, K., Sureka, A.: Pariket: Mining Business Process Logs for Root Cause Analysis of Anomalous Incidents. *Databases Networked Inf. Syst. DNIS 2015, B. Ser. Lect. Notes Comput. Sci.* 8999, 244/263 (2015).
 33. Stofa, J., Stofa, S., Kopka, M., Snashel, V.: Adaptation of Turtle Graphics Method for Visualization of the Process Execution. *Afro-European Conf. Ind. Adv. AECIA2014 B. Ser. Adv. Intell. Syst. Comput.* 334, 327–334 (2015).
 34. Moody, D.: The “Physics” of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Trans. Softw. Eng.* 35, 756–779 (2009).
 35. Wilkinson, L.: *Grammar of graphics*. Springer, New York (1999).
 36. Tufte, E.: *The Visual Display of Quantitative Information*. Graphics Press, Cheshire (1983).
 37. Ware, C.: *Information Visualization: Perception for Design*. Morgan Kaufman, Hampshire (1999).
 38. Few, S.: *Information Dashboard Design*. (2006).
 39. Ware, C.: Visual Thinking: For Design. *Ergonomics.* 53, 138–139 (2008).
 40. Cairo, A.: *The Functional Art: An introduction to information graphics and visualization*. New Riders (2013).
 41. Colligan, L., Anderson, J.E., Potts, H.W.W., Berman, J.: Does the process map influence the outcome of quality improvement work? A comparison of a sequential flow diagram and a hierarchical task analysis diagram. *BMC Health Serv. Res.* 10, (2010).
 42. Object Management Group: *Business Process Model and Notation (BPMN)*.
 43. Baxter, P., Jack, S.: Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *Qual. Rep.* 13, 544–559 (2008).
 44. Yin, R.K.: *Case study research and Applications: Design and methods*. SAGE publications, Los Angeles (2018).