

Towards a Multi-parametric Visualisation Approach for Business Process Analytics

Stefan Bachhofner, Isabella Kis, Claudio Di Ciccio, and Jan Mendling

Vienna University of Economics and Business, Austria
s.bachhofner@me.com, isabella.kis93@gmail.com,
{claudio.di.ciccio, jan.mendling}@wu.ac.at

Abstract. Visualisation is an integral part of many scientific areas and is reportedly an important tool for learning and teaching. One reason for this is the picture superior effect. Nevertheless, little research endeavour has been carried out so far to effectively apply visualisation principles to the emerging field of business process analytics. In this paper a novel multi-parametric visualisation approach is proposed in such a context. General visualisation principles are used to create, evaluate, and improve the approach in the design process. They are drawn from a wide range of fields, and are synthesised from theory and empirical evidence.

Keywords: visualisation, business process analytics, business process management, process mining

1 Introduction

Visualisation is a powerful tool for understanding data. In statistics an exploratory data analysis is performed before any statistical method is applied. Any data science process includes a step where the data is explored visually. Medicine and cartography pay particular attention to the colour scheme too. However, in the context of Business Process Management (BPM), little has been done in research to develop visualisation frameworks that effectively help domain experts and process analysts understand the performance of the examined processes. A missed opportunity because information represented visually is more likely to be remembered due to the picture superior effect [7,11]. Business Process Management Systems (BPMSs) play an important role for process-aware organizations. However, BPMS fall short on powerful process analysis tools, especially from the perspective of visualisation. At times, pie charts are used instead of representations that convey the information more accurately.

In this position paper, the importance of powerful visualisation tools in process science is emphasised. In particular, a set of general visualisation principles is presented. Thereupon, we design an unprecedented multi-parametric approach that visually depicts process execution dynamics on a process model, with the representation of multiple performance metrics at once. The presented framework is based upon the results of a research project held in collaboration with PHACTUM Softwareentwicklung GmbH.

The remainder of the paper is organised as follows. [Section 2](#) introduces the preliminaries from BPM and visualisation. [Section 3](#) proposes our novel multi-parametric visualisation approach for process analytics. [Section 4](#) concludes the paper and outlines further research.

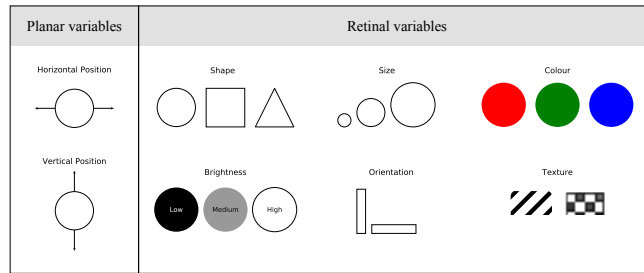


Fig. 1: Visual variables [12]

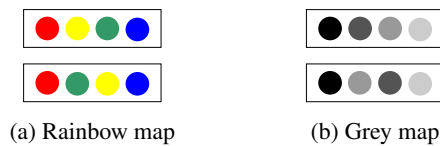


Fig. 2: Colour perceptual ordering [4]

2 Background

BPM is the art and science of overseeing how work is performed in an organisation to ensure consistent outcomes and to take advantage of improvement opportunities [5]. To that extent, the Internet of Events (IoE) [1] opens up new opportunities to process analysts who can rely on the efficient treatment of big data and various sources. Such opportunities include the automated processing of data by means of machine learning techniques and statistical methods, which benefit from the availability of large data sets.

Visualisation is graphical representation of data or concepts [17]. Atomic building blocks of this representation are visual variables, as first described by Bertin et al. [3] and successively clarified by Moody [12] (Fig. 1). Together they form the set of possible visual combinations, i.e., the design space [12]. The chosen visual variable has to preserve the structure of the underlying data [15]. For example, assume a categorical ordinal variable is given such as quantiles or age categories, e.g. young, middle-aged and elderly. In both cases the categories imply an order that has to be maintained. Therefore, a visual variable has to be chosen where perceptual ordering is possible, as shown in Fig. 2. For example, the grey colour map is perceptually ordered because it only varies in brightness (Fig. 2(b)). In contrast, the widely used rainbow colour map is not perceptually ordered because no intuitive sorting of colours is usually sensed by readers (Fig. 2(a)). Another important principle is contextualisation, namely the context+focus paradigm [9,18], which applies when one wants to focus on a part of a system while showing the context of the system as a whole.

Process mining tools such as Minit [14] and ProM [2]¹ use visualisation extensively to display values related to activities' performance metrics such as the frequency with which tasks are carried out – see e.g. the inductive visual miner of ProM [10]. Recently

¹ www.minitlabs.com, www.promtools.org

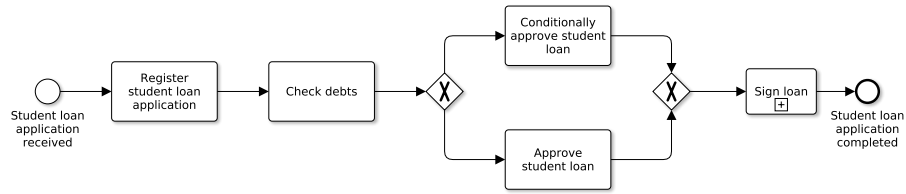


Fig. 3: Example of a student loan application in BPMN [5]

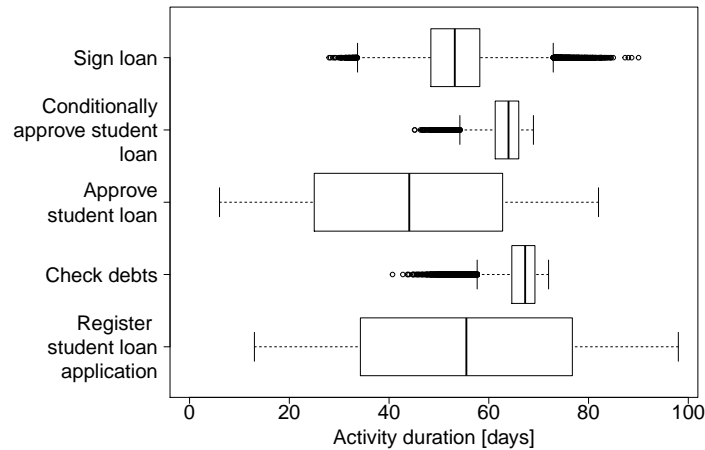


Fig. 4: Box plot of the simulated activity durations







also BPMSs such as Camunda² have begun to show measured metrics on process models. However, little has been done in research and practice to visualise more than one performance metric. Visualising two or more metrics at once can prove beneficial because the user can identify patterns and relationships, as it happens with mosaic plots, a multivariate visualisation for categorical data in statistics [16]. The following section clarifies this assumption with a use-case example.

3 Outline of the approach

To illustrate our approach a student loan application process extracted from [5] is used (Fig. 3). We assume that the process has been completed 100,000 times. The process starts when a loan application is received. First, the application is registered and then the applicants credit-worthiness is checked. Then, the application is either conditionally approved or approved. “Conditionally approve student loan” has been completed 80,000 times and “Approve student loan” 20,000 times, respectively. Finally, the complex activity “Sign loan” is activated and the process is completed. The box plot of the simulated activity durations in days are reported in Fig. 4.

² docs.camunda.org/manual/7.6/webapps/cockpit/bpmn/process-history-views/#heatmap

Table 1: Colour codes for categories c_i

Category	Colour
Outlier below the 1 st quantile (c_1)	
Between 1 st and 2 nd quantile (c_2)	
Between 2 nd and 3 rd quantile (c_3)	
Between 3 rd and 4 th quantile (c_4)	
Between 4 th and 5 th quantile (c_5)	
Outlier above the 5 th quantile (c_6)	

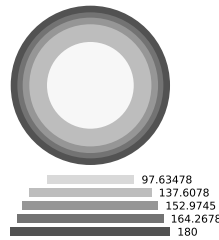


Fig. 5: Visual representation of frequency and duration for activity “Check debts”

In our approach, the first step is to identify variables that are of interest for the analysis purpose. In this example, we consider (i) the number of times they were executed, namely their frequency, and (ii) the time the activities need to complete. In addition, we want to show outliers with respect to time, to point out where exceptionally long- or short-lasting tasks took place in relation to the others. To detect the outliers, we classify the registered absolute time values into N categories for every activity, based upon the corresponding quantiles. In the following, we will refer to these categories as c_i where $c_i \in \{c_1, \dots, c_N\}$. c_i is the category of values between the $(i - 1)$ -th and the i -th quantile. c_1 and c_N refer to the outliers. In our example, we consider $N = 6$.

As previously stated, maintaining the consistency between the underlying structure of data and the visual representation is essential. In our example, both duration quantiles and frequency are data for which a total order exists. To depict their values, we therefore choose two visual variables which allow for a perceptual ordering: The grey colour map to encode quantiles and the size to encode frequency. A third visual variable is implicitly considered because the information is displayed on the process model, hence the additional parameter is the activity for which the metrics are measured. In our example a radial representation of data overlaps the activity boxes of the model to that extent.

Table 1 lists the colour codes assigned to c_i . In the following, we provide an example of how the described categories c_i can be visually translated into the diameter of circles over activities, taking into account the execution frequency. Since the information is displayed on top of a process model, the maximum allowed diameter for each category has to be pre-calculated on the basis of the box size for the activity label containers, due to clear readability reasons. We name such a parameter as \bar{d} . For example, assume that the maximum allowed diameter is equal to 180 units³. The chosen maximum diameter d for an activity should not overcome the activity box. We recall that the diameter of circles here represents the frequency of executed activities. Therefore we scale it by the maximum frequency among all the activities in the process (in this example, 100,000). For activity “Approve student loan”, e.g., we have that $d = \frac{20000}{100000} \cdot \bar{d} = 0.2 * 180 = 36$ units. For “Check debts” $d = 1.0 * 180 = 180$ units instead. The following equation is

³ By “unit” we mean any display or printing unit of measurement, such as pixels, centimetres, and the like.

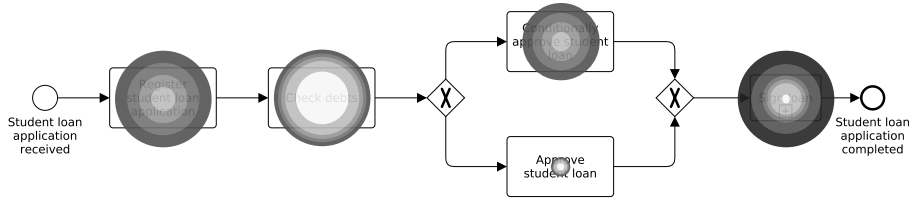


Fig. 6: Multi-parametric visualisation of activities duration and execution frequency

then used to determine the diameter d_{c_i} of every category c_i .

$$d_{c_i} = \frac{\lambda_i - a}{b - a} \cdot d \quad (1)$$

where λ_i is the upper bound of category c_i , i.e., the i -th quantile, a is the minimum activity duration (i.e., the 0-th quantile), and b is the maximum activity duration (i.e., the 6-th quantile). The formula applied to activity “Check debts”, e.g., results in the following diameters:

- $d_{c_2} = \frac{57.66257 - 40.69231}{71.9848 - 40.69231} \cdot 180 = 97.63478$
- $d_{c_3} = \frac{64.61496 - 40.69231}{71.9848 - 40.69231} \cdot 180 = 137.6078$
- $d_{c_4} = \frac{67.28645 - 40.69231}{71.9848 - 40.69231} \cdot 180 = 152.9745$
- $d_{c_5} = \frac{69.24977 - 40.69231}{71.9848 - 40.69231} \cdot 180 = 164.2678$

The results are depicted in Fig. 5. Observe that only four diameters were calculated because the diameter for the last category is always equal to d . Repeating this calculations for each activity and projecting the results on the process model leads to the result drawn in Fig. 6. Examining the figure, the outliers can be easily identified by visually extracting the lowest and highest brightened areas. Both activities “Check debts” and “Sign loan” present outliers, but the latter stands out for the ratio of long-lasting executions, as opposed to the former. However, the frequency plays no role in that, as it can be noticed by the correspondence of the diameter of the superimposed circles.

4 Conclusion

This paper has positioned our research endeavour in the visualisation of business process analytics using general visualisation principles based on theory and empirical evidence. In this context, an example has been proposed that deals with the activities’ execution times and their frequency simultaneously visualised on a process model. Beyond the proposed example, a multi-parametric visualisation might be improved by considering additional parameters, e.g., a cost matrix depending on actual costs from accounting, or the extent to which a category is considered to be the least favourable to the business purposes. This matrix can then have an influence of the visualisation, e.g., scaling the size of the graphical elements or modifying the the colour scheme.

For our future research, we aim at implementing a prototype applying those principles in practice, so as to perform experiments on case studies with researchers and

practitioners in the area. Theoretical concepts to compute Process Performance Indicators (PPIs) on the basis of registered process data have been recently proposed in [8]. We will work to integrate the metrics devised in [8] with our approach. Studies on the influence of virtual and augmented reality on visualisation and how BPM can benefit from this new technologies are in our future plans too, also in the light of the recent advancements in the area [6,13].

References

1. van der Aalst, W.M.P.: Data Scientist: The Engineer of the Future, pp. 13–26. Springer International Publishing, Cham (2014)
2. van der Aalst, W.M.P., van Dongen, B.F., Günther, C.W., Rozinat, A., Verbeek, E., Weijters, T.: Prom: The process mining toolkit. In: BPM (Demos). CEUR Workshop Proceedings, vol. 489. CEUR-WS.org (2009)
3. Bertin, J.: *Semiology of Graphics*. University of Wisconsin Press (1983)
4. Borland, D., Taylor, R.M.: Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications* 27, 14–17 (2007)
5. Dumas, M., Rosa, M.L., Mendling, J., Reijers, H.A.: *Fundamentals of Business Process Management*. Springer Publishing Company, Incorporated (2013)
6. Filonik, D., Rittenbruch, M., Foth, M.: Datachopin - designing interactions for visualisation composition in a co-located, cooperative environment. In: CDVE. pp. 126–133 (2016)
7. Goolkasian, P.: Pictures, words, and sounds: From which format are we best able to reason? *The Journal of General Psychology* 127(4), 439–459 (2000)
8. Kis, I., Bachhofner, S., Di Ciccio, C., Mendling, J.: Towards a data-driven framework for measuring process performance. In: BPMDS (2017)
9. Lamping, J., Rao, R.: The hyperbolic browser: A focus+context technique for visualizing large hierarchies. *Journal of Visual Languages & Computing* 7(1), 33 – 55 (1996)
10. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P.: Process and deviation exploration with inductive visual miner. In: BPM (Demos). vol. 1295, p. 46 (2014)
11. Lidwell, W., K.H., Butler, J.: *Universal principles of Design: A Cross-Disciplinary Reference*. Gloucester MA: Rockport Publishers (2003)
12. Moody, D.: The physics of notations: Toward a scientific basis for constructing visual notations in software engineering. *IEEE Transactions on Software Engineering* 35(6), 756–779 (Nov 2009)
13. Poppe, E., Brown, R., Recker, J., Johnson, D., Vanderfeesten, I.: Design and evaluation of virtual environments mechanisms to support remote collaboration on complex process diagrams. *Information Systems* (2017)
14. Puchovsky, M., Di Ciccio, C., Mendling, J.: A case study on the business benefits of automated process discovery. In: SIMPDA. pp. 35–49 (2016)
15. Rogowitz, B.E., Treinish, L.A., Bryson, S.: How not to lie with visualization. *Comput. Phys.* 10(3), 268–273 (Jun 1996)
16. Theus, M.: Mosaic plots. *Wiley Interdisciplinary Reviews: Computational Statistics* 4(2), 191–198 (2012)
17. Tory, M., Möller, T.: Human factors in visualization research. *IEEE Transactions on Visualization and Computer Graphics* 10(1), 72–84 (Jan 2004)
18. Turetken, O., Schuff, D., Sharda, R., Ow, T.T.: Supporting systems analysis and design through fisheye views. *Commun. ACM* 47(9), 72–77 (Sep 2004)

This document is a pre-print copy of the manuscript
([Bachhofner et al. 2017](#))
published by Springer
(available at link.springer.com).

The final version of the paper is identified by DOI: [10.1007/978-3-319-60048-2_8](https://doi.org/10.1007/978-3-319-60048-2_8)

References

Bachhofner, Stefan, Isabella Kis, Claudio Di Ciccio, and Jan Mendling (2017). “Towards a Multi-parametric Visualisation Approach for Business Process Analytics”. In: *CAiSE Workshops*. Ed. by Andreas Metzger and Anne Persson. Vol. 286. Lecture Notes in Business Information Processing. Springer, pp. 85–91. ISBN: 978-3-319-60048-2. DOI: [10.1007/978-3-319-60048-2_8](https://doi.org/10.1007/978-3-319-60048-2_8).

BibTeX

```
@InProceedings{ Bachhofner.etal/COGNISE2017:MultiparametricVisualisationBusinessProcessAnalytics,
  author      = {Bachhofner, Stefan and Kis, Isabella and Di Ciccio,
                Claudio and Mendling, Jan},
  title       = {Towards a Multi-parametric Visualisation Approach for
                Business Process Analytics},
  booktitle   = {CAiSE Workshops},
  year        = {2017},
  pages       = {85--91},
  publisher    = {Springer},
  crossref    = {CAiSE2017Workshops},
  doi         = {10.1007/978-3-319-60048-2_8}
}
@Proceedings{ CAiSE2017Workshops,
  title       = {Advanced Information Systems Engineering Workshops: CAISE
                2017 International Workshops, Essen, Germany, June 12--16,
                2017, Proceedings},
  year        = {2017},
  editor      = {Metzger, Andreas and Persson, Anne},
  volume      = {286},
  series      = {Lecture Notes in Business Information Processing},
  isbn        = {978-3-319-60048-2},
  doi         = {10.1007/978-3-319-60048-2}
}
```