Guiding the Visualization of Time-Oriented Data

Davide Ceneda, Wolfgang Aigner, Markus Bögl, Theresia Gschwandtner, Silvia Miksch*†

Abstract

The analysis of industrial processes allows quality assessment and production monitoring. Usually these operations are carried out exploiting time-series data. In this work, we analyze a concrete design study of space efficient and time-aggregating visualizations for the analysis of high-frequency time-series. We derive recommendations to enhance the design process and demonstrate their applicability to our case study.

1 Introduction

Many industrial processes, like oil extraction, require constant assessment and quality checks of the whole production phase, to avoid unnecessary and costly maintenance operations subsequent to machine failures. This kind of assessment is useful also to maintain adequate production levels, and optimize the process.

A common way to inspect the production of industrial goods is to analyze the data measured by numerous sensors placed on the physical items. Such sensors produce a huge amount of data by logging measurements at fixed time intervals. The visualization and the analysis of this high-frequency quality data add challenges that have been studied in numerous works (for a comprehensive view, refer to [1]). However, many constraints depend also on the particular context in which the analysis takes place. Aside the challenges imposed by the data itself, further constraints derive directly from the users of the visualization: their inclusion in the exploration cycle contributes to the subjectivity of the analysis [10], and requires additional effort in the design phase. For these reasons, adapting or creating new visualizations to analyze long time-series is highly context-sensitive, and usually requires many design loops to reach satisfying results. In the following, we 1) describe a concrete case study project of space efficient and time-aggregation designs for high-frequency sensor data using data sketching. By analyzing the issues in the case study, we 2) propose the introduction of guidance techniques to enhance, lean and, simplify the design.

2 Visual Analysis of Drilling Measurements

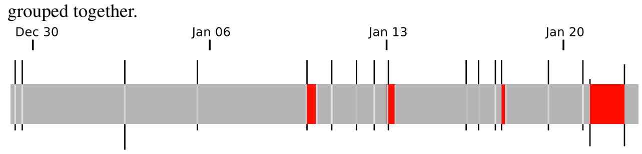
Time-series visualization comprises a lot of interesting problems that have been studied for long, like data-aggregation [3], or space-efficient approaches [5]. Despite the great number of studies [1], usually the results strictly rely on the application context. For this reason, we derive recommendations for the design of space-efficient and data-aggregating time-series visualizations. Our work draws on a paper by Alsallakh et al. [2], who presented an approach for segmenting and labeling multivariate time-series, to reconstruct drilling processes. The resulting process is then analyzed by experts for optimization purposes. Starting from this scenario, we designed a visualization specifically tailored to highlight critical events occurring during the excavation. We describe the problem context in terms of data, users, and tasks [9].

*Davide Ceneda, Theresia Gschwandtner, Markus Bögl, and Silvia Miksch are with Vienna University of Technology, Austria. Wolfgang Aigner is with St. Pölten University of Applied Sciences.

†Emails: {davide.ceneda, theresia.gschwandtner, markus.boegl, silvia.miksch}@tuwien.ac.at; wolfgang.aigner@fhstp.ac.at

(a) Critical measurements are represented with a centered shape filled with red proportionally to the number of critical events. Data is aggregated in periods of fixed length i.e. 4 days.

(b) The color and the saturation are proportional to the percentage of critical events in a variable time period. Events falling within a fixed time span (e.g., 1 hour) are grouped together.



(c) The bar visualization is enriched with aggregated data, shown on an additional y-axis above and below the main view. Vertical space is used to include details about the time period, like the average events' length, etc., to support the evaluation.

Figure 1: Visual sketches of critical events in a drilling process.

Data: The dataset comprises quantitative measurements collected (one each second) by tens of sensors mounted on the drilling rig. Critical events are related to values passing a certain threshold and are symptoms of problems. The dataset comprises multiple time-series, one for each sensor. In a first phase we were asked to support just the visualization of single time-series. However, the huge number of collected data makes the problem challenging, in terms of space-efficiency, also if considering single sensors. **Users:** The visualization should address the needs of three kinds of users: managers, engineers, and the drilling crew. **Tasks:** According to the roles, three tasks are identified: the managers have the responsibility to *supervise* the overall trend of the excavation. If a problem is noticed, the engineers should *trace the causes* of the fault comparing the measurements with the raw data. The drilling crew needs a *real time indicator* of the mining process to prevent the interruption of work and the disruption of machines. The opposite necessities of the users add complexity to the design. This stimulated us to think whether it is possible to limit the design cycles, by integrating such knowledge (data, users, tasks) into assistive methods to lean the design. This consideration is exposed, together to the issues that lead to it, in the next section.

3 Visual Design

For sake of brevity, we describe just the sketches designed to support the managers. Managers need a high-level overview of the drilling process. They need to understand in a glimpse and with minimum interaction effort the overall process quality. Besides the large number of samples in the dataset, the available screen space for visual elements is very limited and space efficient solutions are needed. For this reason, although the position is a better visual variable to encode quantitative data [8], we chose color, saturation, (Figures 1b, 1c) and length (Figures 1a, 1c), to encode the values. The large number of samples, as well as the task itself (supervise/overview), imposed an aggregation of the data.

Aside these visual constraints, we had to handle the haziness of the 'bad events' concept. Although, the concept of single bad measurements is easy to understand, it was not clear how a series of bad events should be visually aggregated. Moreover, we had to identify suited time spans for data aggregation, and suited thresholds for distinguishing critical intervals from the non-critical ones (i.e., reflecting the amount of bad events within an interval). To meet the constraints we aggregated the data in several ways, and proposed the resulting visualizations to our partners.

Because of these uncertainties, we chose an approach based on sketches, to sustain a frequent feedback loop between us and our partners. Some sketches represented in Figure 1: these are just a subset of the produced ones. From a visualization perspective, our partners chose the more simple visualizations out of a range of possibilities, because of their effectiveness with respect to the task. The sketch in Figure 1a was chosen because of its familiarity: it is in fact, one of the most common ways to encode time-series data. Figures 1b and 1c were chosen because they effectively give a rapid overview of the process. Figure 1c stands out because it consents to encode additional information on the vertical space, providing additional values for the analysis.

When analyzing the benefits and shortcomings of these visualizations, we derived some general rules for designing effective and expressive visualizations by taking (1) task specifications, (2) dataset properties, and (3) user abilities into account. Considering these types of information, these rules act as guidance [11] for the selection of suited visualization techniques. The idea of guidance is not new, however a formalization is not yet available for visualization. These thoughts are resumed in the next paragraph, in the form of recommendations for the future work.

Towards Task-Oriented Guidance During the case study, we faced different uncertainties which lead to a number of trial-and-error iterations. While a user evaluation is essential, it is possible to design better suited visualizations right from the start and thus, limit the number of design cycles by asking the right questions and considering the right information. We formalize our thoughts as recommendations, and demonstrate their applicability to the case study.

Apply Guidance to the InfoVis Pipeline. In order to support or even automate the design we need to tackle each step of the information visualization pipeline [4]. Thus, we identify information that has to be considered in visualization design for each step of the pipeline. This information can be derived from a detailed task specification, describing which data subset is needed, in which way the data needs to be manipulated as well as the workflow of composed tasks. In the data transformations step interaction techniques provided by the visualization need to support the specified tasks. In the visual mapping and view transformation steps these tasks may impose particular visual cues, and a proper arrangement in respect to the whole workflow of composed tasks. Furthermore, the perceptual abilities of the user, and domain specific constraints need to be considered. We identify task-taxonomies and user-categorizations as a good input to effectively support these steps.

Exploit Taxonomies and Use Degree of Interest Functions. One of the main issues raised during the design, regarded the identification of interesting data cases. In particular, it was not clear whether the properties of the time domain should be taken into account to extract interesting insights. The importance of a data case depends on the tasks. A way to formalize tasks is using taxonomies. Because of their generality, they may be exploited to generalize approaches aimed at solving specific problems for a complete range of tasks. Currently, just a few of them exist for time-oriented data [7, 1], however it is not straightforward to apply such knowledge to specific contexts. However, it is not always possible to define tasks in a definite way. Approximate interestingness indicators, like Degree of Interest (DOI) [6] functions may lead to better results, by taking into account the characteristics of the time domain. We think that the use of taxonomies and DOI functions may be generalizable to various domains. DOI functions may contribute where a formalization of tasks is not possible.

Application to the case study. In our use case, the data transformation can be supported by the definition of a DOI function: in the field of oil extraction, as in other fields, it is possible to find correlations between parameters and sensors. This cause-effect relation is useful for the discovery of non-trivial problems. It may be mapped to numerical values in order to assign to each data case a score, related to its importance. Timings among measurements are crucial, as time implies a relationship. Consequently, the assignment of different scores varies, depending on the temporal distance. For instance, this score may be directly mapped to the opacity of the color (Figure 1b, 1c). The monitoring task implies an aggregation of the data at different time scales in combination with interactive exploration (zoom-in on the time-line) of interesting areas following the importance score of the data. This is useful for engineers. Although quite simple, these observations, derived from the knowledge of tasks, domain, users, and time characteristics, would have decreased the number of produced sketches, thus demonstrating the need to incorporate such knowledge into assistive methods.

4 Conclusion and Future Work

In this work, we described a case study project of space efficient and time-aggregation designs for high-frequency sensor data using data sketching. We locate different issues affecting the design phase, and propose different solutions, in form of recommendations. We discuss the introduction of guidance approaches to enhance the design phase. To this aim, task taxonomies and DOI functions may help. However, further work is required to structure this knowledge for time-oriented data. Guidance itself is not yet a defined field. Many facets of guidance approaches require further investigations.

Acknowledgements

This work was supported by Centre for Visual Analytics Science and Technology CVAST, funded by the Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative (#822746).

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer, 2011.

[2] B. Alsallakh, M. Bögl, T. Gschwandtner, S. Miksch, B. Esmael, A. Arnaout, G. Thonhauser, and P. Zllner. A Visual Analytics Approach to Segmenting and Labeling Multivariate Time Series Data. In M. Pohl and J. Roberts, editors, *EuroVA*. Eurographics, 2014.

[3] L. Berry and T. Munzner. Binx: Dynamic exploration of time series datasets across aggregation levels. In *InfoVis 2004. IEEE Symp. on*, pages p2–p2. IEEE, 2004.

[4] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Kaufmann, 1999.

[5] M. Chuah. Demystifying venture capital investing. In *InfoVis 2002. IEEE Symposium on*, pages 161–164. IEEE, 2002.

[6] G. W. Furnas. *Generalized fisheye views*, volume 17. ACM, 1986.

[7] A. MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 1995.

[8] J. Mackinlay. Automatic design of graphical presentations. Technical report, Stanford Univ, 1987.

[9] S. Miksch and W. Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38:286–290, 2014.

[10] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *TVCG*, 20(12):1604–1613, 2014.

[11] H.-J. Schulz, M. Streit, T. May, and C. Tominski. Towards a characterization of guidance in visualization. *Poster at IEEE VIS*, 2013.