

On Visual Analytics and Evaluation in Cell Physiology: A Case Study

Fleur Jeanquartier and Andreas Holzinger

Research Unit Human-Computer Interaction, Institute for Medical Informatics,
Statistics and Documentation, Medical University Graz
{f.jeanquartier,a.holzinger}@hci4all.at

Abstract. In this paper we present a case study on a visual analytics (VA) process on the example of cell physiology. Following the model of Keim, we illustrate the steps required within an exploration and sense-making process. Moreover, we demonstrate the applicability of this model and show several shortcomings in the analysis tools functionality and usability. The case study highlights the need for conducting evaluation and improvements in VA in the domain of biomedical science. The main issue is the absence of a complete toolset that supports all analysis tasks including the many steps of data preprocessing as well as end-user development. Another important issue is to enable collaboration by creating the possibility of evaluating and validating datasets, comparing it with data of other similar research groups.

Keywords: visual analytics, evaluation of visualization, human computer interaction, biomedical science.

1 Introduction

From the first data analysis attempts to exploratory data analysis, up to information visualization, today we are facing the possibilities of visual analytics (VA). With VA several analysis processes may be transformed and become more effective and efficient through integrating automated analysis results and reasoning [1]. There is ongoing research in a variety of application areas ranging from document analysis over network security to molecular biology. Applying visual analysis techniques within these areas bring up certain limitations [2]. Dealing with the complexity of biological data requires sophisticated visualization technologies. Prominent examples of visualization for exploration and analysis in the domain of biology come from systems biology and include, among many others, the visualization of biological networks and omics data [3] such as protein structures [4], visual analysis of gene expression data [5], but also visual analysis of cell signaling networks [6]. For populating such databases for network analysis biologists also deal with basic research in cell physiology.

In Fig. 1 we see a slightly modified version of the VA Process, first described by [7]. According to Keim, humans have to be included early in the data analysis

process. By using their background knowledge and being supported by processing, transformation and visualization tools the analysis process eventually brings up new insight. We illustrate the mapping of a VA process during one example cell physiological experiment. Life scientists especially in the domain of biomedical science may struggle with the fact, that the process starts with a first data analysis. As for the observed work process later described in the Section 2 the domain expert also started with describing the hypothesis and then choosing suitable materials and methods for data acquisition. According to [7] input for the data sets used in the VA process are of heterogeneous nature and can be results from scientific experiments. Therefore we included the prior results as input for the feedback loop. The hypothesis may be formed by a preceding exploration. The domain expert makes use of knowledge gained by preceding work.

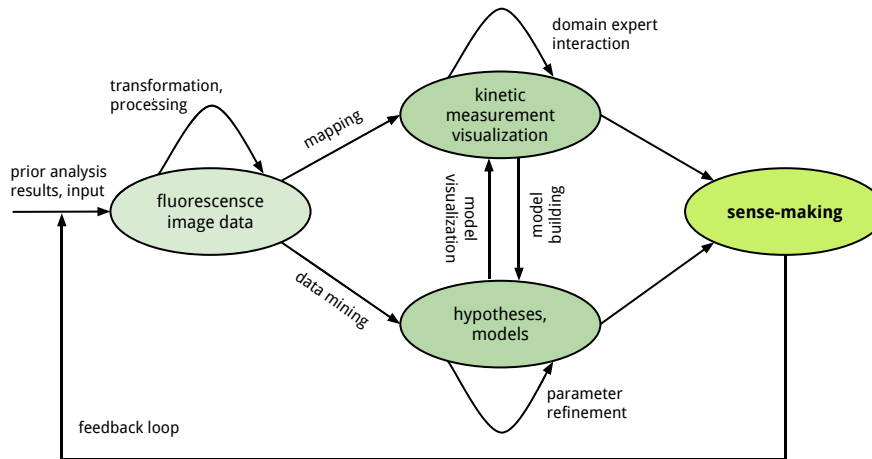


Fig. 1. Adapted Version of Keim's VA Process for the Application In Cell Physiology

Human computer interaction (HCI) and knowledge discovery (KDD) along with biomedical informatics are of increasing importance to effectively make sense out of data [8]. Biologists can benefit from a data deluge with the means of an integrated visualization approach, however, conducting evaluation and improving toolsets are still required to overcome certain hurdles on the way to new insights [9]. A domain expert as analyst often works alone while analysing data sets facing many problems, some of that have been illustrated by [10]. To foster sense-making and insights in VA systems it is essential to conduct studies and determine how people are using such systems [11, 12]. Case studies as field studies are a common approach to evaluating VA systems [13]. Qualitative evaluation such as observational studies can be conducted in a more realistic setting and allow improved understanding of existing practices for analysis and environmental constraints [14].

Consequently, we describe an observational study of a domain expert in cell physiology to present the current practice of VA in this domain.

2 Observation

The user, a domain expert within biomedical science, is part of the visual analysis and KDD process of a group of researchers dealing with cell physiology experiments. We accompanied the domain expert while investigating and analysing a set of experiments' results and observed the expert's analysis work. The analysis process includes visual analysis as part of the data processing, data analysis and KDD process as well as visual communication for dissemination.

A fluorescent biosensor [15] measures the concentration of certain molecules within cellular compartments. Fluorescent biosensors can be used for monitoring various processes and analytes such as metabolites, ions, target localization, gene expression and physiological relevant changes within subcellular regions [16]. The biosensor allows to quantify variations in concentration or localization of the specific analyte within the cell by a change in fluorescence intensity. This quantification is further visualized as intensity signal over time in terms of kinetic curves. By that method, data in hundreds of columns and rows is recorded and has to be processed further. In summary, this method provides the measurement of biological signaling dynamics *in vivo*.

Experiments start with monitoring kinetics in signal transduction. The signal represents the fluorescence intensity [17]. First of all sequences of high-resolution fluorescent imaging of cells are acquired to capture dynamic changes. This action takes place in the lab's dark room. Fluorescence images are captured by a digital camera incorporating a CCD detector, connected to the fluorescence microscope. A commercial bioimaging software is used to communicate with the hardware, translating recorded signals to raw data. The software also provides some data/image processing functionality. Once the measurements are complete, the analysis process continues with data processing and image analysis. Noise (such as background lights within the dark room) reduction of images is supported by a ratio function. The domain expert marks specific regions of interest within the cell in order to monitor biological activities in healthy and pathological cells. Image segmentation is done manually insofar as the domain expert manually selects specific regions of interest on the image data for further comparison and analysis. Hence, regions of interest as polygon shapes are placed on every raw source to display the intensity value. The evaluation of whether the data and to what extent is accurate is done by manually comparing specific regions with a background region. The software allows the scientist to explore the data only in a very limited way. For not occupying the lab's dark room workplace for the time-consuming tasks of data processing and analysis, the expert moves to another workplace outside the dark room. Consequently, when the domain expert believes, that the data is sufficient, the raw data is exported to a commercial spreadsheet computation software via CSV for further processing and analysis.

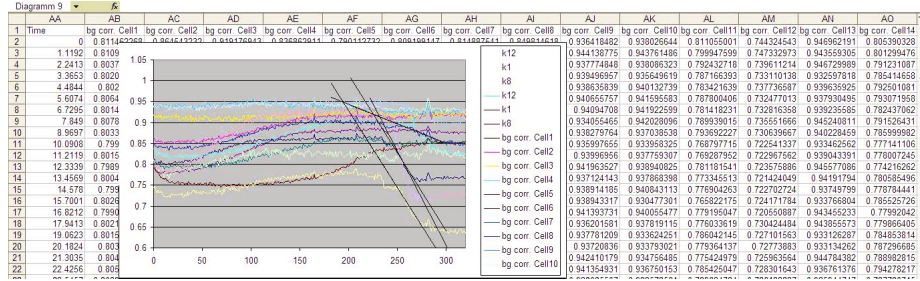


Fig. 2. First visual analysis of intensity signal over time

The domain expert creates a first visualization (compare Fig. 2) of the data, describing the kinetic changes in specific groups of healthy and pathological cells. This task is done semi-automatically by end-user development [18]. The following visual analysis shows, that the data has to be further filtered, corrected and transformed and finally improved in terms of readability and to be visualizable for the task of dissemination. It is up to the domain expert and the implicit knowledge of models, developed by the group of researchers within the lab, which transformation and manipulations are considered to be appropriate. Some of the processing tasks are automatic and some are again manual. The domain expert uses several tools for the various tasks and switches between them while advancing in the analysis process. While the process itself is occasionally being discussed in group, several smaller but complex actions are double checked by colleagues. Both the experiments and the visual analysis process are repeated many times until certain "surprising" [19] results get visible. This repetitive approach to gain new insights, also known as explorative data analysis (EDA), supports the process illustrated in Fig. 1 as it consists of a feedback loop. Finally, when the visual analysis results show surprising effects, the domain experts concludes with the dissemination (see Fig. 3) of the results, again with the means of visualization. The final visualizations are again being iteratively improved.

By further discussing the case study's process and comparing it with the VA process, we try to outline certain issues when dealing with the evaluation of scientific visualizations.

3 Discussion

The case study shows, that there are analysis processes in biomedical science which embody VA as a lived approach. At the same time, the case study also shows the need for improvements regarding HCI and end-user development. Experts in this domain are using their domain knowledge in combination with both automatic and visual analysis together, but need to be guided by computer science experts to improve the choice of tools that are used. There are certain tasks still done manually that could be automated or at least semi-automated, using

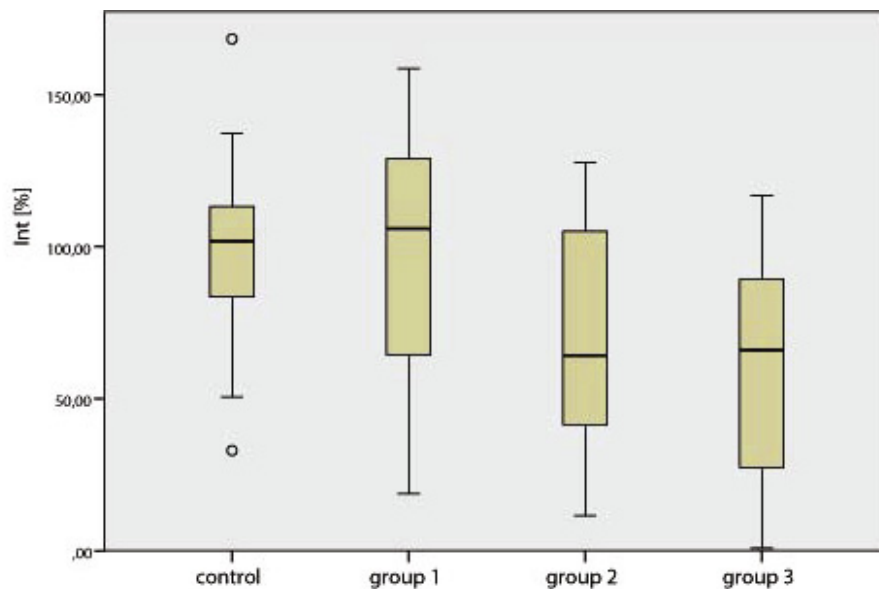


Fig. 3. Visualization of the kinetic parameters measured within the groups of healthy control to pathological cells for dissemination

the right tool, such as image segmentation, noise reduction as well as post-processing up to creating a set of fitting visualizations for dissemination. For instance, there already exist attempts to automatically estimate suitable background in fluorescence imaging [20] and automatic localization of cell nuclei [21]. The case study supports the statement, proposed by VA, that automated analysis often speeds up analysis tasks. It also shows that communication through visual representations is used for the dissemination of research results (see Fig. 3). Therefore, the case study shows, that visual analysis plays an important role for gaining new insights and dissemination. However, the case study also highlights that certain evaluation tasks are missing.

Lessons Learned include that there is a gap in free exploration of data and information due to the lacking usability and interaction possibilities of the tools used. The researchers within this field of study state, that they do not know about powerful VA solutions. At the same time, they are facing certain restrictions that hinder them to cooperate fully with computer science experts. During the observation the domain expert made complaints about shortcomings in the analysis tools' functionality and usability. Many tasks have to be repeated, not only due to data inconsistency, but also because most tools are hardly fault-tolerant and lack in supporting the user in certain data preprocessing steps as well as in the post-processing such as choosing the right visualization technique and improving the visualization's readability.

There are several possibilities to improve the end-user development and to minimize the interaction junk [22] within the observed process, such as simplifying

the creation of the effect curves for both visual analysis as well as dissemination. Furthermore, the visualizations in use are still limited to curve diagrams and bar charts. Alternative visualization metaphors such as multi-variate data visualizations [23] allow scientists to explore the data and its various dimensions in other ways and may highlight certain effects that are not visible within the current effect curves.

The discussion after the observation further included improvements and suggestions to support the whole VA process. Due to the reason that both data as well as study results are confidential we are not allowed to go into detail in this respect. However, we already communicate general aspects of HCI and KDD and present general suggestions for improving VA within this domain. The domain expert agreed that there are several possibilities how evaluation could be integrated to support VA. Lam et al. already list some fitting evaluation goals and questions within the VDAR- and the CTV scenario [13]. However, the very idea of discussing the visual analysis process with a domain expert in HCI already brought up certain shortcomings within the visual analysis work. Suggestions include: Evaluating the dataset, comparing it to datasets of other similar groups of researchers, would help validating specific models as well as techniques and speed up the analysis work. Moreover, enabling and facilitating collaboration supports scientific problem solving [24]. The researchers in the group also agree on the fact, that evaluating software in use and furthermore, having the possibility to improve and extend the tools functionality would improve their daily research tasks. Incooperating the many steps of data examination and preprocessing into a single tool would be highly appreciated. The case study highlights the need for conducting evaluation and improvements in VA in the domain of biomedical science.

4 Conclusion

Every day scientists in many sub domains of life sciences such as biomedical science are facing the challenging task of VA with the goal of reaching new insights. Life scientists may benefit from a data deluge with the means of an integrated visualization approach. However, conducting evaluation and improving certain toolsets for exploratory data analysis and end-user development are prominent challenges on the way to new insights.

We described an observational study of VA in cell physiology. We compared the process to Keim's VA process. The case study shows, that there are analysis processes in biomedical science which embody VA. Further studies may include additional practice of VA related analysis work of various other approaches in biomedical science. The observation highlights the need for conducting evaluation and improvements in VA in the domain of biomedical science. We suggested evaluation possibilities and further noted challenges regarding its' application for visualization in life sciences. Among others, suggestions include incooperation and improvement of support for developing visualization in regard to analysis.

References

- [1] Keim, D.A., Kohlhammer, J., Ellis, G., Mansmann, F.: Mastering The Information Age-Solving Problems with Visual Analytics. Florian Mansmann (2010)
- [2] Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., Andrienko, G.: Solving Problems with Visual Analytics. *Procedia Computer Science* 7, 117–120 (2011)
- [3] Gehlenborg, N., O’Donoghue, S.I., Baliga, N.S., Goesmann, A., Hibbs, M.A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D., et al.: Visualization of omics data for systems biology. *Nature Methods* 7, S56–S68 (2010)
- [4] Doncheva, N.T., Assenov, Y., Domingues, F.S., Albrecht, M.: Topological analysis and interactive visualization of biological networks and protein structures. *Nature Protocols* 7(4), 670–685 (2012)
- [5] Lex, A., Streit, M., Kruijff, E., Schmalstieg, D.: Caleydo: Design and evaluation of a visual analysis framework for gene expression data in its biological context. In: 2010 IEEE Pacific Visualization Symposium (PacificVis), pp. 57–64. IEEE (2010)
- [6] Berger, S.I., Iyengar, R., Maayan, A.: Avis: Ajax viewer of interactive signaling networks. *Bioinformatics* 23(20), 2803–2805 (2007)
- [7] Keim, D.A., Mansmann, F., Schneidewind, J., Thomas, J., Ziegler, H.: Visual analytics: Scope and challenges. In: Simoff, S.J., Böhlen, M.H., Mazeika, A. (eds.) *Visual Data Mining*. LNCS, vol. 4404, pp. 76–90. Springer, Heidelberg (2008)
- [8] Holzinger, A.: On Knowledge Discovery and interactive intelligent visualization of biomedical data: Challenges in Human–Computer Interaction & Biomedical Informatics. In: *DATA-International Conference on Data Technologies and Applications*, pp. 5–16 (2012)
- [9] Donoghue, S.I.O., Gavin, A.-c., Gehlenborg, N., Goodsell, D.S., Hériché, J.-k., Nielsen, C.B., North, C., Olson, A.J., Procter, J.B., Shattuck, D.W., Walter, T., Wong, B.: Visualizing biological data - now and in the future. *Nature Publishing Group* 7(3), S2–S4 (2010)
- [10] Wong, B.L.W., Xu, K., Holzinger, A.: Interactive Visualization for Information Analysis in Medical Diagnosis. In: Holzinger, A., Simonik, K.-M. (eds.) *USAB 2011*. LNCS, vol. 7058, pp. 109–120. Springer, Heidelberg (2011)
- [11] Kang, Y.A., Görg, C., Stasko, J.: How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 17(5), 570–583 (2010)
- [12] Wong, P.C., Shen, H.-W., Johnson, C.R., Chen, C., Ross, R.B.: The Top 10 Challenges in Extreme-Scale Visual Analytics. *IEEE Computer Graphics and Applications* 32(4), 63–67 (2012)
- [13] 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), 1–18 (2011)
- [14] Carpendale, S.: Evaluating information visualizations. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) *Information Visualization*. LNCS, vol. 4950, pp. 19–45. Springer, Heidelberg (2008)
- [15] Morris, M.C.: Fluorescent biosensors of intracellular targets from genetically encoded reporters to modular polypeptide probes. *Cell Biochemistry and Biophysics* 56(1), 19–37 (2010)
- [16] Okumoto, S., Jones, A., Frommer, W.B.: Quantitative imaging with fluorescent biosensors. *Annual Review of Plant Biology* 63, 663–706 (2012)
- [17] Mehta, S., Zhang, J.: Reporting from the field: genetically encoded fluorescent reporters uncover signaling dynamics in living biological systems.. *Annual Review of Biochemistry* 80, 375–401 (2011)

- [18] Lieberman, H., Paternò, F., Wulf, V.: End user development, vol. 9. Springer (2006)
- [19] Beale, R.: Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. *International Journal of Human-Computer Studies* 65(5), 421–433 (2007)
- [20] Chen, T.-W., Lin, B.-J., Brunner, E., Schild, D.: In situ background estimation in quantitative fluorescence imaging.. *Biophysical Journal* 90(7), 2534–2547 (2006)
- [21] Song, Y., Cai, W., Huang, H., Wang, Y., Feng, D.D., Chen, M.: Region-based progressive localization of cell nuclei in microscopic images with data adaptive modeling. *BMC Bioinformatics* 14(1), 173 (2013)
- [22] Endert, A., North, C.: Interaction junk. In: *Proceedings of the 2012 BELIV Workshop on Beyond Time and Errors - Novel Evaluation Methods for Visualization - BELIV 2012*, pp. 1–3. ACM Press, New York (2012)
- [23] Fuchs, R., Hauser, H.: Visualization of Multi-Variate Scientific Data. *Computer Graphics Forum* 28(6), 1670–1690 (2009)
- [24] Good, B.M., Su, A.I.: Games with a scientific purpose. *Genome Biology* 12(12), 135 (2011)