Visual Analytics in Environmental Research: A Survey on Challenges, Methods and Available Tools

Martin Komenda and Daniel Schwarz

Institute of Biostatistics and Analyses, Masaryk University,
Kamenice 126/3, 625 00 Brno, Czech Republic
{komenda,schwarz}@iba.muni.cz

Abstract. Visual analytics approaches bring an innovative and effective way how to deliver the knowledge from a particular domain to an individual user. With the use of visual analytics methods we can easily discover the unexpected relations and interesting patterns, which are hidden in the huge data warehouses. It builds on the human mind's ability to understand the complex visualization of information. In this paper we introduce the potential usefulness of visual analytics for researchers working in the field of environmental informatics. Current challenges beyond the survey are described here, including the summary of particular well-proven tools and scenarios, which can be applied in many various fields of environmental research.

Keywords: Visual analytics, learning analytics, human cognition, environmental data visualization.

1   Introduction

Visual analytics (VA) is a new interdisciplinary field of study that calls for a more structured scientific approach to understanding the effects of interaction with complex graphical displays on human cognitive processes. Its primary goal is to support the design and evaluation of graphical information systems that better support cognitive processes in areas as diverse as scientific research and emergency management [1]. Technologies based on VA have been moving from research to a widespread use, driven by the increased power of analytical databases and computer graphics hardware. People adopt VA to see and understand complex data, new visual paradigms and hidden insights [2]. Visual analytics techniques allow analysts to look at the data from different perspectives or in various domains and to fulfill diverse analytical tasks. Visual displays and interactive techniques are often combined with computational processing, which, in particular, enables analysis of a larger sets of data than would be possible with purely visual methods [3]. A sufficiently deep understanding of visual analytics guides developers of visualization techniques in creating more effective tools in terms of user requirements, and helps evaluators in designing future studies that are more representative of insights sought by users from their data sets [4].

That knowledge brings an interesting way how to optimize delivering knowledge from huge amount of available data in a particular domain to an individual user. The need for more effective processing of data by users has been identified by several researchers in the Technology Enhanced Learning (TEL) field [5], [6], [7]. With the establishment of TEL, a new research field, called Learning analytics (LA), has emerged in recent years. According to the 1st International Conference on Learning Analytics and Knowledge, LA is defined as measurement, collection, analysis and reporting of data about learners and their contexts   for purposes of understanding and optimizing learning processes and the environments in which it occurs [8]. Similarly to VA, LA borrows and synthesizes techniques from different related fields, such as data mining, social network analysis or business intelligence to harness them for converting data into useful information and thereon to motivate actions, like self-reflecting ones previous teaching or learning activities, to foster improved teaching and learning [9]. At present, data mining tends to concentrate on developing new tools for discovering patterns in data. LA – at least as it is currently contrasted with data mining approaches — focuses on the use of intelligent data, learner-produced data, and analytical models to discover information and social connections, and to predict and advise on learning [10], [6]. The emergence of LA as a field in its own right meant that there were now separate groupings focusing on each of the challenges driving analytics research [11]. Regardless of the great enthusiasm that is currently surrounding LA, it also raises substantial questions for research. In addition to technically focused research questions such as the compatibility of educational datasets, or the comparability and adequacy of algorithmic and technological approaches, there remain several "softer" issues and problem areas that influence the acceptance and impact of LA. Among these are questions of data ownership and openness, ethical use and dangers of abuse, and the demand for new key competences to interpret and act on LA results. It is also necessary to mention that any application that analyses student data must comply with privacy regulations, which adds cost and complexity [12]. This means that the implementation of LA in learning processes requires to be carefully crafted in order to be successful and beneficial [13].

2   Human Cognitive Abilities and Limitations

Many researchers concentrate on human cognition and user interaction issues raised by the VA community. The top ten challenges of extreme-scale VA were identified in [14], among them being the problem of the human bottleneck. It is predicted that performance of all major computing components – power, memory, storage, bandwidth, concurrence, etc. – will improve by a factor of 3 to 4,444 by 2018. Human cognitive capability will certainly remain constant. One of the challenges is, therefore, finding alternative ways to compensate for human cognitive weaknesses. Thus, it is necessary to develop new visualization approaches that facilitate the flow of human reasoning. Human cognitive models or human cognitive frameworks, including sub-processes of human thinking such as accommodation, categorization or drawing conclusions will play a key role in that development, as it is shown in [15], [16]. A novel

approach how the user interface can be adapted in relation to a current process-task and under consideration of the user's behaviour is described in [17] – a work related to FUPOL project (www.fupol.eu) aimed on social computing and crowd sourcing in e-participation. The research performed in economic decision-making [18] addresses overcoming bounded rationality issues that arise from the cognitive limitations through the application of a VA tool that can aid in complex decision making tasks in real world applications with uncertainty. A more integrative approach to the interplay of science and technology in VA would build new scientific methods for developing and empirically evaluating theories of human-computer ''mixed-initiative'' cognitive processing [19].

Users may also benefit from a personalized use of Adaptive Learning Environments with underlying context-aware recommender systems for learning [20], which provide tools such as content-based filtering, collaborative filtering, knowledge-based filtering and their hybridizations. A similar example is the content infrastructure of the EU-oriented Open Discovery Space CIP PSP initiative (http://www.opendiscoveryspace.eu) that also attempts to create a very big data infrastructure collecting learning content and usage data around it. Such global learning data infrastructures can help in scaling up recommender systems by allowing them to consume, process and use a rich variety of contextualized usage data streams, and thus enable novel forms of real time intelligence that can only become possible when processing extremely large data volumes.

3   Data Representation and Processing

Interaction with visualizations enables the dialog between user and the visual representation of the underlying data. The interactive manipulation of the data, the visual structure or the visual representation provides the ability to solve various tasks and discover insights. Classifications of interaction describe an abstract view on various visual tasks and may provide a mapping to the human information processing.
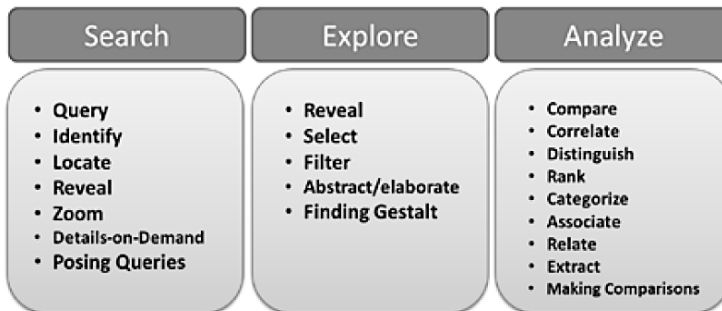

Fig. 1. High-level tasks with assigned subtasks and interactions [21]

Figure 1 illustrates three important high-level tasks and interactions assigned to them as well as subtasks derived from the existing visual task classifications. Various models for differentiating visual variables in context of human information processing are important for an accurate and beneficial adaptation of visualizations [21].

At the core of successful visual analytics systems are computational techniques that transform data into concise, human comprehensible visual representations. Computers store, move and analyze data that, on initial examination, are a simple collection of bits in various primitive data forms such as structured (for example, relational tables, geometry), semi-structured (for example, e-mail, network packets) or unstructured (for example, a collection of text). Many algorithmic approaches are organized around key data characteristics, which are applied to all primitive data forms rather than algorithms that apply to specific data types. Transforming data into an effective visual representation is fundamentally different from transforming incoming raw data. In addition to the raw performance needed to support interactive analytics, characteristics of the data itself affect the transformations and representations for both computation and visualization. Successful data transformation and representation methods combine mathematical, statistical and linguistic analysis with hardware and software techniques to handle various kind of raw data such as massive data, geospatial and temporal data, imperfect data, heterogeneous data, and users-in-the-loop [22]. For example Starlight information graphics [23] provides the classification according to the type of information they portray. The two major classes are non-spatial information graphics such as text and numeric data, and inherently spatial information graphics, which depict information containing intrinsic spatial coordinates, such as geospatial or CAD data. One of Starlight's unique capabilities is its ability to integrate these two types of information so that they may be analyzed together.

4   Current Challenges

Recently scientists have been working on creating sophisticated models of existing domains, to simulate the domain behavior and to extrapolate gained knowledge into the future. They use VA approaches as an integral part in each step of this process, especially for visualization used to verify model setup and to estimate its parameters. High-performance computing simulations of massive amounts of data are then analyzed using visualization software specifically designed for parallel processing of large amounts of structured and unstructured data [24]. Given the challenge of automatic computational analysis of complex data, much recent effort in the analytics community has focused on the design of interactive graphical representations of information that might better support the human ability to perceive and to construct meaningful patterns from data [25]. The use of various virtual environments can provide log files that can be studied with web mining techniques to understand the behaviour of the users. However, statistical metrics are not always easy to interpret for the stakeholders, which suggest the need to develop novel approaches for the easy discovery of usage patterns:

- Combining human domain knowledge with automatic data analysis techniques by providing users with interactive visual interfaces, whereby 'interactive' means that users can actively participate in the analytical process.
- Context-aware recommender systems for learning based on VA to allow users for contextualized multiscale exploratory analytical reasoning.

5   Methods

The visual analytics can be applied in many various fields of environmental research. In the following section we briefly introduce the key domains and describe the beneficial use of VA methods.

The utility of VA algorithms was reported among many fields of human scientific activity also in environmental information systems known for their large volumes of multidimensional spatial-temporal data. For illustration, POPs GMP presents analysis of available Stockholm Convention Global Monitoring Plan data on persistent organic pollutant concentrations in core matrices. It provides a set of electronic tools for storage and on-line visualization of data published therein, and proposed a comprehensive IT solution for future data collection campaigns. The visualization provides easily accessible information on performance of monitoring programs in individual countries and/or regions sorted by matrices, time, and compounds [26].
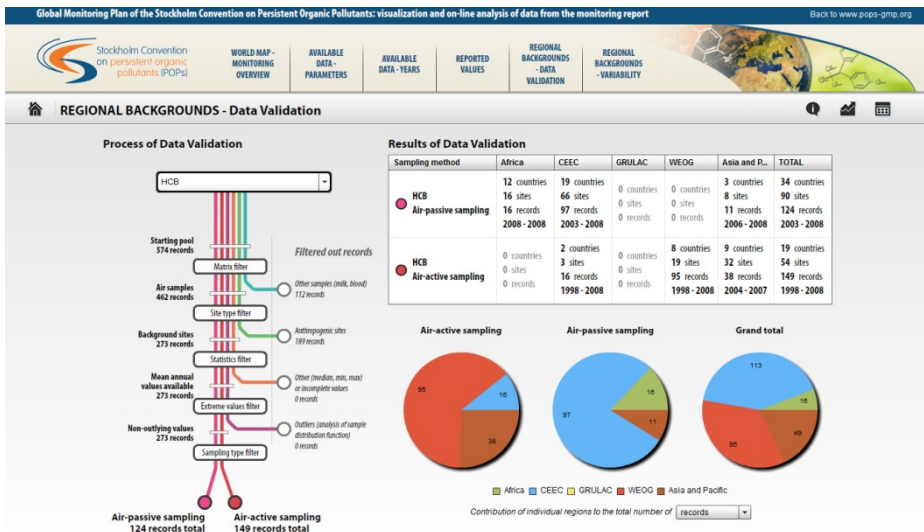

Fig. 2. A screenshot from a statistical summary of reported background atmospheric POPs concentrations [26]

Mike Sips et al. have proposed an approach for detection of interesting patterns in environmental time series: an algorithm to characterize the temporal behavior, visual detection of potentially interesting patterns in a matrix visualization, and interactive
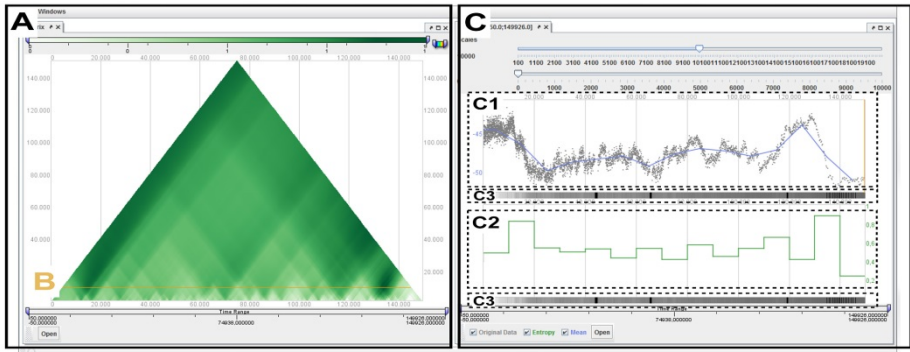
Fig. 3. The visual interface of the prototype which shows the Antarctica time series [27]

exploration of detected patterns. The system computes statistical measures for all possible time scales and starting positions. A specific feature of mentioned approach is its flexibility due to the wide range of statistical measures provided, e.g., mean, variance, entropy. This allows for capturing a large set of temporal patterns to address different analysis tasks and different types of data [27].

Natalia and Gennady Andrienko presented illustrated structured survey of the state of the art in visual analytics concerning the analysis of movement data and have demonstrated how different visual analytics techniques can support our understanding of various aspects of movement.
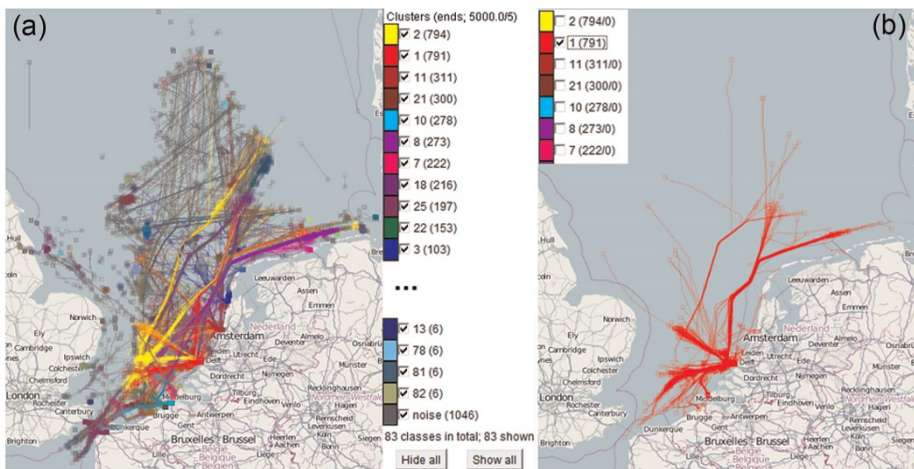

Fig. 4. Interactive progressive clustering of trajectories. (a) The ship trajectories have been clustered according to the destinations. (b) One of the clusters is selected [3].

This hot research topic in visual analytics utilize the legacy of cartography, with its established techniques for representing movements of tribes, armies, explorers, hurricanes, and so on; time geography, with its revolutionary idea of considering space and

time as dimensions of a unified continuum (space–time cube) and representation of behaviours of individuals as paths in this continuum; information visualization, with its techniques for user-display interaction supporting exploratory data analysis; and geovisualization, with its interactive maps and associated methods enabling exploration of spatial information [3].

The U.S. Department of Energy's Office of Environmental Management currently supports an effort to understand and predict the fate of nuclear contaminants and their transport in natural and engineered systems.
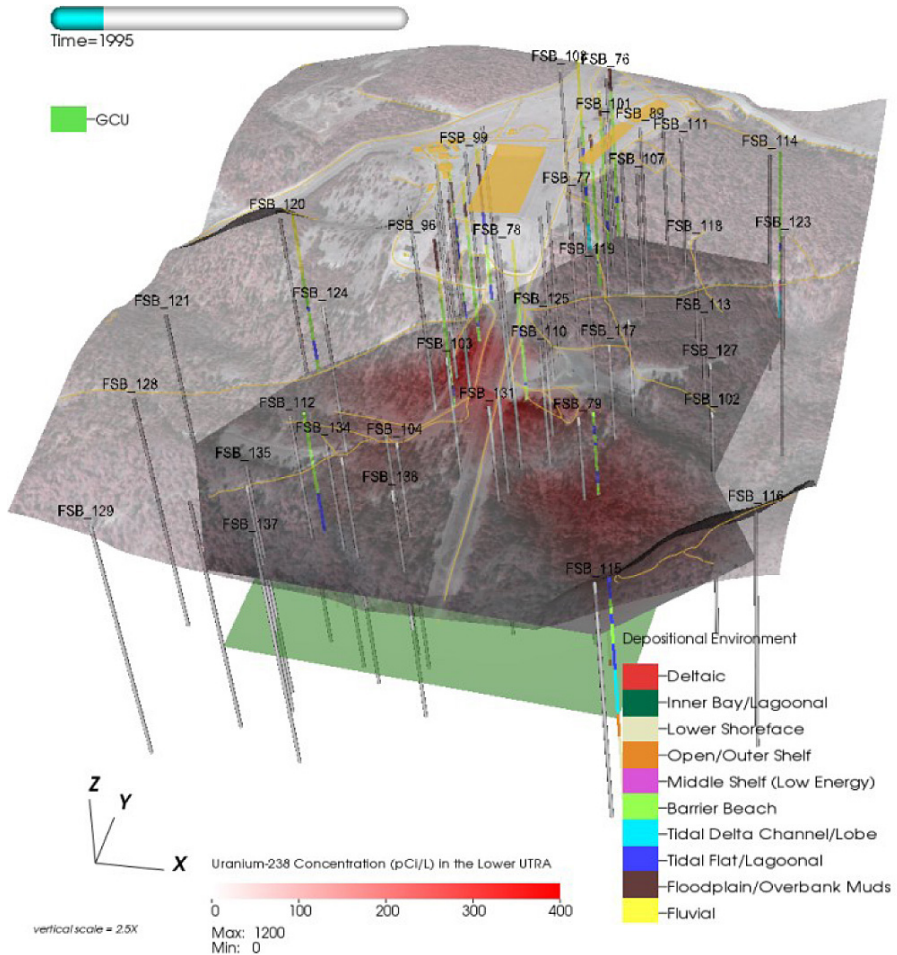

Fig. 5. Three dimensional visualization (computed in VisIt) of Uranium-238 plume migrating through the subsurface of the F-area [24]

Joerg Meyer et al. demonstrates that by bringing together experts from a variety of domains, including geology, hydrology, physics, and computer science, we can

develop an integrated system that helps scientists make well informed decisions regarding environmental management. They have developed set of tools that was designed to help determine the extent of soil and water contamination from radioactive isotopes from measured data and to develop models that will be used to predict the future progression of a contaminated site with respect to these contaminants [24].

Watershed management, in its very nature, represents a participatory process, requiring horizontal and vertical collaborations among multiple institutions and stakeholders. Alexander Sun has shown the way of the migration of an Environmental decision support systems module (EDSS) module from the traditional client-server based architecture to a client of cloud-computing services. Google Drive, which is behind the new version of the EDSS module, provides a number of basic visual analytics features that can be used to increase the collaborative decision-making experience while drastically reducing the cost of small-scale EDSS. This approach can fundamentally change the way many decisions are made by allowing collaborations to be set up quickly with less overheads than the client-server approach. The promise of using Google Fusion Table to support collaborative decision-making at only a fractional cost of a full-scale, client-server application was shown. Although only a simple analytical model is described in this study, other models can be easily incorporated into the Cloud to perform more sophisticated analyses [28].
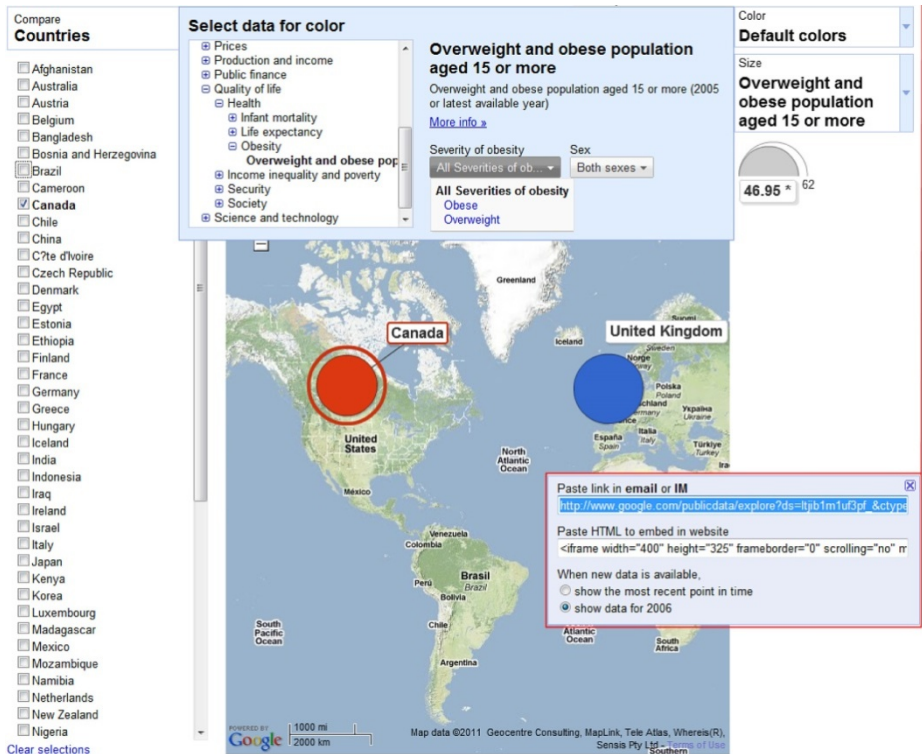

Fig. 6. Screenshot of Google Public Data Explorer showing some of its interface features and options [29]

Geospatial visual analytics, a specialized subtype of visual analytics, is an emerging multidisciplinary area which supports spatio-temporal analytical reasoning and decision-making through interactive visual interfaces (such as maps and other visual artifacts) that are linked to computational methods. Microsoft Pivot enables powerful visual zooming in and out of Web databases and the discovery of new patterns and relations in them that would otherwise be invisible in standard Web browsing of large datasets. It can also be used for geographic classification and offers a fresh way how to visually browse and arrange massive amounts of data and images online and also supports geographic and temporal classifications of datasets featuring geospatial and temporal components [29].

6   Summary on VA Tools and Their Usage

After surveying issues raised by the VA community in the paragraphs above, we present here a summary on particular VA tools which are available for various sorts of use, see Table 1.

Table 1. Selected available VA tools, their main purpose, underlying methods and scenarios of their use

VA Tool [reference]	Purpose	Methods & Techniques	Scenarios & Data sets
POPs GMP [26]	management and visualization of Global Monitoring Plan data	data pre-processing, outlier identification, LOQs replacement, data aggregation, descriptive statistics	persistent organic and pollutant data
Visualization of Multiscale Exploration [27]	detection of patterns in numerical time series from environmental sciences	time scales statistical values computing, pattern visual identification, detected pattern exploration	Earth's climate system and ocean data
Multivariate Trajectories Visualization [30]	customization and versatile exploration using multiple density fields	trajectory aggregation, density field computation	anomalous behaving ships and maritime safety data
Trajectory Contingency Table [31]	exploration of attributes of trajectories	automatic identification system, continuous density model, sampled trajectories computation	density vessel positions data
Visualization via Progressive Clustering [32]	trajectories (dis)similarity assessment	progressive clustering, OPTICS method, trajectory clustering	car GPS-tracking data

VA Tool [reference]	Purpose	Methods & Techniques	Scenarios & Data sets
Animal Ecology Explorer [33]	iterative generation of metadata based on exploratory findings	exploration through enrichment approach, segmentation, clustering	animal tracking data
Visualization Using Graph Centrality Measures [34]	places and explore city structure identification	data pre-processing, trajectory aggregation, graph centrality, visual and algorithmic methods	georeferenced photo captures data
Ship and Weather Information Monitoring [35]	weather parameters visualization along ship routes	gridded binary format processing, parallel coordinates plots, voyage time graphs, geospatial data interaction	ship voyages data
ASCEM Project [24]	decision-making process support in planning of treatment options for contaminated sites	model setup verification, critical parameters estimation, parallel processing analysis, simulation results validation	nuclear contaminant data
EDSS for Watershed Management [28]	effective watershed-scale management	total maximum daily load model, simple analytical modeling, cloud-computing approach	coastal river basin data
Geospatial Visual Analytics [29]	visual zooming and discovery of new patterns and relations	Microsoft Live Labs Pivot technology (descriptive component collection, visual data representation)	mortality dataset

7   Conclusions

This paper has discussed the potential usefulness of visual analytics methods, techniques and particular tools for researchers working in the field of environmental informatics – one of extreme-scale-data areas. Visual analytics tools are helpful in situations where a higher level of user interactivity and involvement is needed to process effectively huge amounts of multidimensional data, understand them and use the acquired knowledge to draw a decision. Challenges beyond the state-of-art in visual analytics were outlined here, including the human bottleneck lying in the limited human cognitive capability. Selected existing and well-proven VA tools were described

here together with their main purpose and scenarios of their use. Mental or human cognition modeling as well as adaptive virtual environments with underlying context-aware recommender systems - these are issues which should tease the attention of scientists and developers who need to design visual representation of extreme-scale datasets.

References

1. Fisher, B., Green, T.M., Arias-Hernández, R.: Visual Analytics as a Translational Cognitive Science. Top. Cogn. Sci. 3, 609–625 (2011)
2. Chabot, C.: Demystifying Visual Analytics. Comput. Graph. Appl. Ieee. 29, 84–87 (2009)
3. Andrienko, N., Andrienko, G.: Visual analytics of movement: An overview of methods, tools and procedures. Inf. Vis. 12, 3–24 (2013)
4. Saraiya, P., North, C., Lam, V., Duca, K.A.: An Insight-Based Longitudinal Study of Visual Analytics. Ieee Trans. Vis. Comput. Graph. 12, 1511–1522 (2006)
5. Romero, C., Ventura, S.: Educational data mining: A survey from 1995 to 2005. Expert Syst. Appl. 33, 135–146 (2007)
6. Siemens, G.: What are learning analytics (2010)
7. Duval, E.: Attention please!: learning analytics for visualization and recommendation. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge, pp. 9–17 (2011)
8. Banff, A.: 1st International Conference on Learning Analytics and Knowledge (2011), https://tekri.athabascau.ca/analytics/
9. Dyckhoff, A.L., Zielke, D., Bültmann, M., Chatti, M.A., Schroeder, U.: Design and implementation of a learning analytics toolkit for teachers. J. Educ. Technol. Soc. 15, 58–76 (2012)
10. Bienkowski, M., Feng, M., Means, B.: Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief (2012), http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf
11. Ferguson, R.: Learning analytics: drivers, developments and challenges. Int. J. Technol. Enhanc. Learn. 4, 304–317 (2012)
12. Norris, D.M.: 7 Things You Should Know About First-Generation Learning Analytics. Educ. Learn. Initiat. Eli (2011)
13. Greller, W., Drachsler, H.: Translating Learning into Numbers: A Generic Framework for Learning Analytics. Educ. Technol. Soc. 15, 42–57 (2012)
14. Wong, W., Liu, W., Bennamoun, M.: Ontology Learning from Text: A Look Back and into the Future. Acm Comput. Surv. 44 (2012)
15. Green, T.M., Ribarsky, W.: Using a human cognition model in the creation of collaborative knowledge visualizations. SPIE Defense and Security Symposium. p. 69830C–69830C (2008).
16. Green, T.M., Ribarsky, W., Fisher, B.: Visual analytics for complex concepts using a human cognition model. In: IEEE Symposium on Visual Analytics Science and Technology, VAST 2008, pp. 91–98 (2008)
17. Burkhardt, D., Nazemi, K.: Dynamic process support based on users' behavior. In: 15th Int. Conf.on Interact. Collab. Learn., ICL 2012, pp. 1–6 (2012)
18. Savikhin, A., Maciejewski, R., Ebert, D.S.: Applied visual analytics for economic decision-making. In: Ieee Symp. Vis. Anal. Sci. Technol., Vast 2008, pp. 107–114 (2008)

19. Simon, H.A.: The sciences of the artificial. MIT press (1996)
20. Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., Duval, E.: Context-Aware Recommender Systems for Learning: A Survey and Future Challenges. IEEE Trans. on Learn. Technol. 5, 318–335 (2012)
21. Nazemi, K., Kohlhammer, J.: Visual Variables in Adaptive Visualizations (2013)
22. Kasik, D.J., Ebert, D., Lebanon, G., Park, H., Pottenger, W.M.: Data transformations and representations for computation and visualization. Inf. Vis. 8, 275–285 (2009)
23. Risch, J.S., Rex, D.B., Dowson, S.T., Walters, T.B., May, R.A., Moon, B.D.: The STARLIGHT information visualization system. In: Proceedings of IEEE Conference on Information Visualization, pp. 42–49 (1997)
24. Meyer, J., Bethel, E.W., Horsman, J.L., Hubbard, S.S., Krishnan, H., Romosan, A., Keating, E.H., Monroe, L., Strelitz, R., Moore, P., Taylor, G., Torkian, B., Johnson, T.C., Gorton, I.: Visual Data Analysis as an Integral Part of Environmental Management. IEEE Trans. Vis. Comput. Graph. 18, 2088–2094 (2012)
25. Thomas, J.J., Cook, K.A.: Illuminating the path: The research and development agenda for visual analytics. IEEE Comput. Soc. (2005)
26. Hůlek, R., Jarkovský, J., Borůvková, J.: Global Monitoring Plan of the Stockholm Convention on Persistent Organic Pollutants: visualization and on-line analysis of data from the monitoring reports (2013)
27. Sips, M., Kothur, P., Unger, A., Hege, H.-C., Dransch, D.: A Visual Analytics Approach to Multiscale Exploration of Environmental Time Series. Vis. Comput. Graph. Ieee Trans. 18, 2899–2907 (2012)
28. Sun, A.: Enabling collaborative decision-making in watershed management using cloud-computing services. Environ. Model. Softw. 41, 93–97 (2013)
29. Boulos, M.K., Viangteeravat, T., Anyanwu, M.N., Nagisetty, V.R., Kuscu, E.: Web GIS in practice IX: a demonstration of geospatial visual analytics using Microsoft Live Labs Pivot technology and WHO mortality data. Int. J. Heal. Geogr. 10, 19 (2011)
30. Scheepens, R., Willems, N., van de Wetering, H., Andrienko, G., Andrienko, N., van Wijk, J.J.: Composite density maps for multivariate trajectories. Vis. Comput. Graph. IEEE Trans. 17, 2518–2527 (2011)
31. Willems, N., Van De Wetering, H., Van Wijk, J.J.: Visualization of vessel movements. Computer Graphics Forum, pp. 959–966 (2009)
32. Rinzivillo, S., Pedreschi, D., Nanni, M., Giannotti, F., Andrienko, N., Andrienko, G.: Visually driven analysis of movement data by progressive clustering. Inf. Vis. 7, 225–239 (2008)
33. Spretke, D., Bak, P., Janetzko, H., Kranstauber, B., Mansmann, F., Davidson, S.: Exploration through enrichment: a visual analytics approach for animal movement. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 421–424 (2011)
34. Vrotsou, K., Andrienko, N., Andrienko, G., Jankowski, P.: Exploring city structure from georeferenced photos using graph centrality measures. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) ECML PKDD 2011, Part III. LNCS, vol. 6913, pp. 654–657. Springer, Heidelberg (2011)
35. Lundblad, P., Eurenius, O., Heldring, T.: Interactive visualization of weather and ship data. In: 13th International Conference on Information Visualisation, pp. 379–386 (2009)