

BIG DATA VISUALIZATION AND ANALYTICS: FUTURE RESEARCH CHALLENGES AND EMERGING APPLICATIONS – PART 1

SIGMOD Blog - MARCH 20, 2020 *

* The full report is available in [54]

Data visualization and analytics are nowadays one of the cornerstones of Data Science, turning the abundance of Big Data being produced through modern systems into actionable knowledge. Indeed, the Big Data era has realized the availability of voluminous datasets that are dynamic, noisy and heterogeneous in nature. Transforming a data-curious user into someone who can access and analyze that data is even more burdensome now for a great number of users with little or no support and expertise on the data processing part. Thus, the area of data visualization and analysis has gained great attention recently, calling for joint action from different research areas from *Information Visualization*, *Human-Computer Interaction*, *Machine Learning*, to *Data management & Mining*, and *Computer Graphics*.

Several traditional problems from those communities, such as efficient data storage, querying and indexing for enabling visual analytics, ways for visual presentation of massive data, efficient interaction and personalization techniques that can fit to different user needs, are revisited with Big Data in mind. This is to enable modern visualization systems that offer scalable techniques to efficiently handle billion-object datasets, while limiting the visual response to a few milliseconds [38][7][6][35][19][8].

The *International Workshops on Big Data Visual Exploration and Analytics* (BigVis) is an annual meeting, which provides a forum for researchers and practitioners to discuss, exchange, and disseminate their work. It attracts attention from the research areas of *Information Visualization*, *Human-Computer Interaction*, *Machine Learning*, *Data Management & Mining*, and *Computer Graphics*, and highlights novel works that bring together these diverse communities.

In the context of the **BigVis2020** (<https://bigvis.imsi.athenarc.gr/bigvis2020>), the organizing committee invited *15 distinguished scientists from academia and industry, and diverse communities* to provide their insights regarding the challenges and the

applications they find more interesting in coming years, related to *Big data visualization and analytics*. Each scientist is asked to summarize his thoughts regarding the following two questions:

1. What do you consider as the *top future research challenges* in Big Data visualization and analytics?
2. What do you consider as the *top emerging applications* in the context of Big Data visualization and analytics?

The post is organized in two parts (the full report is also available in [54]). In this first part, we present the views of Gennady and Natalia Andrienko, Steven Drucker, Jean-Daniel Fekete, Danyel Fisher, Stratos Idreos, and Tim Kraska.

Visual Analytics for Data Science: A Critical View

by **Gennady & Natalia Andrienko**

Visual representations are often used in data analysis. In traditional data mining approaches, visualizations appear at the very end of analytical workflows, aiming at interpretation of identified patterns and their communication to various recipients, e.g., other analysts, decision makers or general public. According to the visual analytics philosophy [1][5], while human efforts must be reduced as much as possible by computational processing, visualization needs to be employed throughout the entire analytical workflow whenever an analyst is supposed to take informed decisions concerning further steps. The role of visualization is to convey the necessary information to the human in a form enabling effective perception and cognition. Hence, the task of visual analytics is to develop analytical workflows in which human cognition is effectively supported by visualizations and computational processing. Visual analytics should also be involved in development of new algorithms and software tools for automated analysis and modelling (e.g., machine learning methods) for checking whether and how well the methods are doing what they are intended to do.

Our major research topic is *visually-driven analysis of spatio-temporal data*. A representative example is movement data consisting of sequences of time-referenced positions. A variety of methods and tools exist already (see theoretical foundations in the book [2]); new methods are developing actively (this is confirmed by a large number of accepted papers on this topic at IEEE VIS and other highly-selective conferences) and applied successfully in such domains as transportation [3] and sport analytics [4]. While the following thoughts have arisen from our experiences with spatio-temporal data, they are applicable to other types of complex data.

Several key recent developments create great opportunities for empowering data science by visual analytics. The first one is the appearance and wide spread of data science-oriented languages such as Python and R. These languages enable step-by-step data analysis and support integration of visualizations in analytical workflows. Analytical notebooks based on these languages help to document analysis, enable

reproducibility and help to share results. These technologies are so easy to use that nowadays everyone can become an analyst. There exist textbooks explaining the basics on simple examples, and many pieces of code are available online for use and adaptation. This phenomenon has its back side: self-made analysts often lack fundamental knowledge of the overall analysis process, and miss understanding of why, when, and how visualizations need to be used in analysis. The Internet is overfilled by visualizations, often looking very impressive and fancy, that communicate spurious patterns in inadequate ways. Unfortunately, most of the available code examples and the majority of text books don't go beyond applying basic graphics to simple data and do not demonstrate the analytical value of the graphics. *Therefore, we see the major challenge in educating data scientists on how to use visualizations correctly and effectively within non-trivial analytical workflows, understanding and taking into account the data types, properties of the data, and their complexity.*

Gennady Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Gennady Andrienko was a chair of ICA Commission on GeoVisualisation, a paper chair of IEEE VAST conference (2015-2016) and associate editor of IEEE Transactions on Visualization and Computer Graphics (2012-2016). Currently, he is an associate editor of Information Visualization and International Journal of Cartography.

Natalia Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems and part-time professor at City University London. Results of her research have been published in two monographs "Exploratory Analysis of Spatial and Temporal Data: a Systematic Approach" (Springer, 2006) and "Visual Analytics of Movement" (Springer, 2013). Natalia Andrienko is an associate editor of IEEE Transactions on Visualization and Computer Graphics and editorial board member of several journals.

Systems for Machine Learning (ML) and ML for Systems

by **Steven Drucker**

While perhaps overhyped, the huge amount of attention directed towards Machine Learning is apparent throughout research and industry. Papers are appearing covering numerous aspects, from fundamental theoretical advances like causal reasoning and general intelligence to applications of machine learning in systems and knowledge work. This explosion in machine learning is enabled primarily by the vast amounts of data available and systems that allow the training for models using those large data sets. This in turn enables clever applications of these techniques above and beyond the basic frontiers of ML (classification, clustering, and regression). Given this explosion, we need far better techniques for working with the data and the models for ML. This includes helping troubleshoot models, understanding where models work and don't work, comparing models with each other, and giving understandable explanations for model behavior. Since visualization is fundamentally about helping

humans interpret and interact with data, *interactive tools and visualizations for machine learning is a one of the top challenges for visualization in the coming decade. At the same time, using the output of models to help build better visualizations (whether it's for recommending a single or sequence of visualizations) or to interact at a higher level with data by helping find optimal ways of leveraging human intuition and knowledge while exploiting more powerful computation is a key component of new applications.*

To concretize these areas, here are some of the recent research papers that presage some of the emerging applications in this area.

(a) **Methods for interpretation of ML models**, both specific types of models (such as Additive Models) which both perform well and are somewhat interpretable, and more general techniques for interacting with arbitrary models. Closely related to the above, as a requirement of the European laws for General Data Protection Regulation (GDPR), any decision made algorithmically must be explainable and whenever data needs to be explained, visualization is an important component. Recent work on this includes the research of Hohman et al. [21] from Rich Caruana's GAM models [28], Wattenberg & Viegas [53], and others on creating more interpretable models or LIME [35] and Lundberg [30].

(b) **Systems for troubleshooting and debugging models** and the spaces where they are effective as well as comparing models with each other. Recent work includes the research of Saleema Amershi in Model Tracker [1] and Besmira Nushi on **Error Terrain Analysis** (<https://slideslive.com/38915701/error-terrain-analysis-for-machine-learning-tool-and-visualizations>)

(c) **Generating recommendations for visualizations based on models of user behavior**. Work such as Moritz's Draco system [33] and Kim's GraphScape system [24].

(d) **Interpreting visualization for subsequent reuse**. Work by Poco & Heer [35] and Agrawala et al. [40 (<http://graphics.stanford.edu/projects/dataExtract>)]

(e) **Creating higher level interactions with data through NLP and other modalities** such as the VODER work of Srinivasan et al. [44].

Steven Drucker a Partner and Research Manager of the Visualization and Interactive Data Analysis (VIDA) group at Microsoft Research (MSR), and an affiliate professor at the University of Washington Computer Science and Engineering Department (CSE). In his 30+ year career, he has published over 100 academic papers, and filed over 130 patents in topics ranging from graphics and interfaces to information visualization. He was inducted into the CHI Academy in 2020.

Interactive Visual Analytics

by Jean-Daniel Fekete

To be effective, visualization and visual analytics should be interactive, meaning that computing visual representations should happen in a few seconds, interacting on them should be responsive, and analytics should also be done in accordance with the

acceptable limits of human latency as described in the literature [24]. *Building systems that remain interactive at scale and using complex analytics is a major challenge for the visualization field, which may become irrelevant if it does not address the scalability challenge properly.*

To address this scalability challenge, my new focus of research is "*Progressive Data Analysis and Visualization*", a new paradigm of computation that, instead of performing computations in one step that can take an arbitrarily long time to complete, splits them in a series of short chunks of approximate computations that improve with time. Therefore, instead of waiting for an unbounded amount of time the results of computations for visualization and analytics, analysts can see the results unfolding progressively. They can, therefore, maintain their attention and start making some decisions earlier than if they had to wait for the whole computations to finish.

Meanwhile, while the results are being computed, analysts can also interact with the ongoing computation, changing computation parameters and sometimes steering the algorithms.

This new paradigm is just starting to emerge and will require more time to become mainstream, as explained in the report we published after a Dagstuhl seminar conducted in 2018 [15]. However, *I am confident that Progressive Data Analysis will allow visualization and analytics to become more scalable while remaining interactive to facilitate the exploration of the wealth of data that the world is gathering, coupled with new powerful methods to analyze it developed in machine learning in particular [43].*

This new paradigm is not only important for visualization and visual analytics but also requires strong collaborations with researchers in Databases and Machine Learning who recognize that progressive data analysis will lead to more scalable exploratory systems.

Jean-Daniel Fekete is the Scientific Leader of the INRIA Project Team AVIZ that he founded in 2007. He received his PhD in Computer Science in 1996 from University of Paris Sud, France, joined INRIA in 2002 as a confirmed researcher, and became Senior Research Scientist in 2006. His main research areas are Visual Analytics, Information Visualization and Human Computer Interaction. He is a Senior Member of IEEE.

Understanding the User

by **Danyel Fisher**

We are learning to ask new things of our data. It's increasingly practical to interactively explore Big Data, asking novel questions to discover unexpected phenomena. The lines between different forms of analysis -from relational queries, to unstructured data, to rich media- are blurring. *I'm looking forward to our improving all steps of the*

process: to learning how to best express questions; how to get interactive-speed responses to those questions; and to iterate on those insights to ask the next round of questions.

These steps are interconnected and interdependent. To get interactive responses, for example, we might use *progressive computation*; e.g., [17][16]. That technique requires us to think about *communicating uncertainty* (e.g., [53][20]) and giving the analyst a way to record how much that uncertainty affects their analysis process.

One way to help all these stages is to focus on particular problem domains. I've spent my last few years working on analysis tools for *sampled, uncertain, high-dimensional, streaming data*. As a domain as a whole, that's huge and intimidating. Fortunately, I can target my work on **Honeycomb** (<http://honeycomb.io>) toward their real problems, and so can take advantage of the constraints of their particular context. Honeycomb is an APM (Application Performance Monitoring) tool for debugging distributed systems. Our users are DevOps, who are responsible for deploying new code — and recovering when it fails. Tools like **BubbleUp** (<https://docs.honeycomb.io/working-with-your-data/bubbleup>), a histogram comparison tool, help users isolate specific classes of failures rapidly. Our underlying data structure is similar to Facebook's SCUBA [28].

I believe that focusing relentlessly on specific use cases will make otherwise broad questions simpler. If we can really understand what the user needs to solve their problem, we can learn what sorts of data they will ingest, what queries they might want to ask, what performance characteristics they expect, and what level of precision they need.

Danyel Fisher is the Principal Design Researcher for Honeycomb.io, a data analytics startup that provides observability to engineers who maintain services in production. Before that, he was a researcher at Microsoft Research. His research focuses on ways to help users interact with their data more powerfully and easily. Danyel is an author of "Making Data Visual" (O'Reilly Press, 2018). <http://danyelfisher.info> (<http://danyelfisher.info>)

An inherently data-intensive problem

by **Stratos Idreos**

Visualizing data is one of the best ways to find patterns and information in Big Data. The reason why data visualization is interesting for the data management community is that this is an inherently data-intensive problem. In addition, data scientists may pose arbitrary queries as they create new visualizations or interact with existing ones. This means that such systems get: (1) diverse queries, (2) sequences of queries where each query may depend on the previous one, (3) queries that may be OK to abort, and (4) workloads which need rapid response times to remain interactive even if correctness is not immediately at 100%. Due to this mismatch with typical database applications, there are several long-term and exciting challenges that represent wonderful opportunities for data management researchers given the rich history of the field in data-intensive algorithms and systems. I highlight two of those opportunities as they arise from recent work in the area.

First, what is the equivalent of the relational algebra for visual analytics? It might seem daunting to condense the vast space of possible actions a data scientist may perform into a small set of operations, but this is exactly what the original relational algebra achieved. And then, more complex operations can be synthesized from primitive algebra operations. *On top of that algebra, we can then build systems that rely on a common interface and a small set of operators, allowing the community to collectively attack this problem by considering alternative designs and implementations that respect the same model and API as it happened with relational operators.* This abstraction is one of the secrets both for the adoption of the relational model across diverse applications and for the ability to relatively easily experiment with alternative implementations.

Second, what is the equivalent of the b-tree, and the row-oriented and column-oriented storage schemes? While these are by no means the only indexing and storage options, knowing the extreme designs or some of the most versatile designs, and then heavily focusing on them, allowed relational databases to mature in terms of both speed and robustness. Data visualization and visual analytics is (typically) a data-intensive as opposed to a compute-intensive problem. This means that the way we store and move data is the bottleneck. *Supporting alternative storage schemes and choosing the right one for the right queries is key.* Thus, studying the design of data structures that can absorb the new and diverse access patterns as well as the interactive response times required by visual analytics is a massive opportunity for the data management community.

Stratos Idreos is an associate professor of Computer Science at Harvard University where he leads the Data Systems Laboratory. Stratos' work focuses on discovering the fundamentals of the design of data structures and data intensive systems. Stratos was awarded the ACM SIGMOD Jim Gray Doctoral Dissertation award for his thesis on adaptive indexing. In 2015 he was awarded the IEEE TCDE Rising Star Award from the IEEE Technical Committee on Data Engineering for his work on adaptive data systems. Stratos is also a recipient of the IBM Enterprise System Recognition Award, a Facebook Faculty award, an NSF Career award, and a DOE early career award

User interactions first and then figure out the system

by Tim Kraska

Interactive data exploration for large data and more complex operations. Tools like Tableau and PowerBi praise themselves as interactive data exploration tools. Yet, for larger datasets they rely on pre-computed data cubes, materialized views, and similar techniques to stay interactive. Unfortunately, these techniques severely restrict what the user is able to do. For example, to create a data cube one needs to know upfront, what type of questions the data cube is supposed to answer. This essentially prevents interactive responses for completely new questions. *A key research challenge is, how*

we can build systems which guarantee interactive response times regardless of the question and data size. As part of Northstar, we started to explore how we can leverage progressive computation and sampling to achieve this.

Sustainable insights and insight recommendation. *We need to make finding insights easier. Thus, we need to develop tools which help non-Data Scientists to discover insights and continuously monitor the data for changes.* For example, a system should automatically recommend interesting insights and visualization about things that might interest the user. SeeDB or VizML are systems, which started to explore that. *At the same time, those insights should also be sustainable.* For example, current insight recommendation systems largely ignore the risk of finding spurious insights by testing too many hypotheses.

Novel interfaces. *We should make data analytics more accessible to a broader range of users.* This requires to fundamentally rethink the user interface. *We put everything on the table HCI has to offer; from novel visualizations, interaction patterns, touch screens, up to natural language interfaces.* Interestingly, changing the user interface often also has severe implications on how the backend has to be developed. We (the SIGMOD community) tend to first develop the system and then add the user interface as an afterthought, often leading to clunky old-style interaction. I think it should be the other way around. *Design the user interactions first and then figure out the system, which can actually support them.*

Emerging Applications. I believe, there exists not a single area which is not already impacted by analytics. Thus, it is close to impossible to find a new emerging application for analytics. However, *I am a strong believer that we need to broaden the scope of users, who are able to take advantage of the data.* Current tools are mainly designed for experts or significantly restrict what a user can do. For example, there is not a single tool out there, which makes it easy for a coffee shop owner to analyze his customer base and make predictions about future sales. At the same time, those people could also tremendously benefit from their data. This requires to rethink the way users interact with data.

Tim Kraska is an Associate Professor of Electrical Engineering and Computer Science in MIT's Computer Science and Artificial Intelligence Laboratory and co-director of the Data System and AI Lab at MIT (DSAIL@CSAIL). Before joining MIT, Tim was an Assistant Professor at Brown and spent time at Google Brain. Tim is a 2017 Alfred P. Sloan Research Fellow and received several awards including the VLDB Early Career Research Contribution Award as well as several best paper and demo awards at VLDB and ICDE.

REFERENCES

- [1] S. Amershi, M. Chickering, et al.: ModelTracker: Redesigning Performance Analysis Tools for Machine Learning. CHI 2015
- [2] G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Wrobel: Visual Analytics of Movement. Springer 2013
- [3] G. Andrienko, N. Andrienko, et al.: Visual Analytics of Mobility and Transportation: State of the Art and Further Research Directions. TITS 18(11), 2017
- [4] G. Andrienko, N. Andrienko, et al.: Constructing Spaces and Times for Tactical Analysis in

Football, TVCG, 2019

- [5] N. Andrienko, T. Lammarsch, G. Andrienko, et al.: Viewing Visual Analytics as Model Building. CGF 2018
- [6] M. Behrisch, D. Streeb, F. Stoffel, D. Seebacher, et al.: Commercial Visual Analytics Systems-advances in the Big Data Analytics Field, TVCG 25(10), 2019
- [7] N. Bikakis: Big Data Visualization Tools Survey, Encyclopedia of Big Data Technologies, Springer 2019
- [8] N. Bikakis, T. Sellis: Exploration and Visualization in the Web of Big Linked Data: A Survey of the State of the Art, LWDM Workshop 2016
- [9] E.T. Brown, A. Ottley, et al.: Finding Waldo: Learning About Users from their Interaction TVCG 20(12), 2014
- [10] D. Ceneda, T. Gschwandtner, et al.: Characterizing Guidance in Visual Analytics. TVCG 23(1), 2017
- [11] J. Choo, S. Liu: Visual Analytics for Explainable Deep Learning. IEEE CGA 38(4), 2018
- [12] J.-K. Chou, Y. Wang, K.-L. Ma: Privacy Preserving Visualization: A Study on Event Sequence Data. Comput. Graph. Forum 38(1), 2019
- [13] C. Collins, N. Andrienko, et al.: Guidance in the Human Machine Analytics Process. Visual Informatics 2(3), 2018
- [14] C. D. Correa, Y.-H. Chan, K.-L. Ma: A framework for uncertainty-aware visual analytics. VAST 2009
- [15] J.D. Fekete, D. Fisher, A. Nandi, M. Sedlmair: Progressive Data Analysis and Visualization. Dagstuhl Seminar, 2018
- [16] J.D. Fekete, R. Primet. Progressive Analytics: A Computation Paradigm for Exploratory Data Analysis. CoRR 2016
- [17] D. Fisher, et al.: Trust me, I'm Partially Right: Incremental Visualization lets Analysts Explore Large Datasets Faster CHI 2012
- [18] T. Fujiwara, O.-H. Kwon, K.-L. Ma: Supporting Analysis of Dimensionality Reduction Results with Contrastive Learning. TVCG 26(1), 2020
- [19] P. Godfrey, J. Gryz, P. Lasek: Interactive Visualization of Large Data Sets. TKDE 28(8), 2016
- [20] J. Hullman: Why Authors Don't Visualize Uncertainty. TVCG 26(1), 2019
- [21] F. Hohman, A. Head, R. Caruana, R. DeLine, S. Drucker: Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models. CHI 2019.
- [22] A. Kangasrääsio, et al.: Parameter Inference for Computational Cognitive Models with Approximate Bayesian Computation. Cognitive Science, 2019
- [23] D. Keim, J. Kohlhammer (eds.): Mastering the Information Age: Solving Problems with Visual Analytics. Eurographics 2010
- [24] Y. Kim, et al.: GraphScape: A Model for Automated Reasoning about Visualization Similarity and Sequencing. CHI 2017
- [25] N.W. Kim, L. Shao, M. El-Assady, et al.: Quality Metrics for Information Visualization. CGF 37(3), 2018
- [26] O.-H. Kwon, T. Crnovrsanin, K.-L. Ma: What Would a Graph Look Like in this Layout? A Machine Learning Approach to Large Graph Visualization. TVCG 24(1), 2018
- [27] O.-H. Kwon, K.-L. Ma: A Deep Generative Model for Graph Layout. TVCG 26(1), 2020
- [28] A. Lior, J. Allen, O. Barykin, V. Borkar, B.Chopra, et al.: SCUBA: diving into Data at Facebook. PVLDB 6(11), 2013
- [29] Y. Lou, R. Caruana, J. Gehrke, G. Hooker: Accurate Intelligible Models with Pairwise Interactions. KDD 2013
- [30] S. Lundberg, S. Lee. A Unified Approach to Interpreting Model Predictions. NIPS 2017
- [31] L. Micalef, et al.: Towards Perceptual Optimization of the Visual Design of Scatterplots. TVCG 23(6), 2017

- [32] L. Micalef, G. Palmas, et al.: Towards Perceptual Optimization of the Visual Design of Scatterplots. TVCG 23(6), 2017
- [33] D. Moritz, et al.: Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. TVCG 25(1), 2019
- [34] B. Mutlu, E.E. Veas, C. Trattner VizRec: Recommending Personalized Visualizations. TiiS 6(4), 2016
- [35] L. Po, N. Bikakis, F. Desimoni, G. Papastefanatos: Linked Data Visualization: Techniques, Tools and Big Data. Morgan & Claypool, 2020
- [36] A. Preston, M. Gomov, K.-L. Ma: Uncertainty-Aware Visualization for Analyzing Heterogeneous Wildfire Detections. IEEE CGA 39(5), 2019
- [37] J. Poco, J. Heer: Reverse-Engineering Visualizations: Recovering Visual Encodings from Chart Images. CGF 36(3), 2017
- [38] X. Qin, Y. Luo, N. Tang, G. Li: Making Data Visualization more Efficient and Effective: A survey. VLDBJ 2020
- [39] M.T. Ribeiro, S. Singh, C. Guestrin: "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD 2016
- [40] B. Saket, D. Moritz, H. Lin, V. Dibia, C. Demiralp, J. Heer: Beyond Heuristics: Learning Visualization Design. CoRR 2018
- [41] M. Savva, N. Kong, A. Chhajta, L. Fei-Fei, M. Agrawala, J. Heer: ReVision: Automated Classification, Analysis and Redesign of Chart Images. UIST 2011
- [42] B. Shneiderman: Response Time and Display Rate in Human Performance with Computers. ACM Comput. Surv. 16(3), 1984
- [43] N. Silva, et al.: Eye Tracking Support for Visual Analytics Systems: Foundations, Current Applications and Research Challenges. ETRA 2019
- [44] A. Srinivasan, S.M. Drucker, A. Endert, J. Stasko: VODER: Augmenting Visualizations with Interactive Data Facts to Facilitate Interpretation and Communication. TVCG 25(1), 2019
- [45] S. Thalmann., J. Mangler, et al.: Data Analytics for Industrial Process Improvement. CBI 2018
- [46] C. Turkay, N. Pezzotti, et al.: Progressive Data Science: Potential and Challenges. CoRR 2019
- [47] Y. Wang, K.-L. Ma: Revealing the fog-of-war: A visualization-directed, uncertainty-aware approach for exploring high-dimensional data. IEEE BigData 2015
- [48] X.-M. Wang, W. Chen, J.-K. Chou, C. Bryan, H. Guan, W. Chen, R. Pan, K.-L. Ma: GraphProtector: A Visual Interface for Employing and Assessing Multiple Privacy Preserving Graph Algorithms. TVCG 25(1), 2019
- [49] X.-M. Wang, J.-K. Chou, W. Chen, H. Guan, W. Chen, T. Lao, K.-L. Ma: A Utility-Aware Visual Approach for Anonymizing Multi-Attribute Tabular Data. TVCG 24(1), 2018
- [50] M. Wattenberg, F. Viegas: Visualization: The Secret Weapon of Machine Learning. EuroVis 2017 (Keynote)
- [51] Y. Wu, G.-X. Yuan, K.-L. Ma: Visualizing Flow of Uncertainty through Analytical Processes. TVCG 18(12), 2012
- [52] F. Zhou., X. Lin, et al.: A Survey of Visualization for Smart Manufacturing. Journal of Visualization 22 (2), 2019
- [53] T. Zuk, S. Carpendale: Theoretical Analysis of Uncertainty Visualizations. Visualization and Data Analysis, 2006
- [54] Gennady Andrienko, Natalia Andrienko, Steven Drucker, Jean-Daniel Fekete, Danyel Fisher, Stratos Idreos, Tim Kraska, Guoliang Li, Kwan-Liu Ma, Jock D. Mackinlay, Antti Oulasvirta, Tobias Schreck, Heidrun Schmann, Michael Stonebraker, David Auber, Nikos

Bikakis, Panos K. Chrysanthis, George Papastefanatos, Mohamed Sharaf: "Big Data Visualization and Analytics: Future Research Challenges and Emerging Applications". 3rd Intl. Workshop on Big Data Visual Exploration & Analytics (BigVis 2020) .

BigVis2020 Profiles

David Auber is a professor of computer science at LaBRI (University of Bordeaux). His main expertise is in information visualization and, more particularly, in visualization of large graphs. He has developed the project Tulip, an information visualization framework dedicated to the analysis and visualization of relational data. His recent research focus on leveraging big data infrastructure and deep neural network to level up information visualization.

Dr. Nikos Bikakis is a data engineer at Atypon Inc. and a postdoctoral researcher at the ATHENA Research Center in Greece. He received his PhD in Computer Science in 2016 from the NTU of Athens. Nikos is a co-author of "Linked Data Visualization: Techniques, Tools and Big Data", Morgan & Claypool 2020. He is co-organizing the annual international workshop "Big Data Visual Exploration & Analytics" (BigVis 2020, 2019 & 2018). Also, he has served as Guest Editor of the special issues "Interactive Big Data Visualization & Analytics" and "Big Data Visualization, Exploration & Analytics" of the Big Data Research Journal. In 2018, his research in the field of Big Data visual analytics was supported by a young researcher Nat/EU grand. Nikos has been awarded an ADBIS 2019 best paper, ESWC 2015 best poster and an honorary scholarship for his PhD studies.

Panos K. Chrysanthis is a professor of computer science and the founding director of the Advanced Data Management Technologies Laboratory at the University of Pittsburgh. His research interests lie within the areas of data management (big data, databases, data streams & sensor networks, data analytics and visualization). He received the US National Science Foundation CAREER Award (1995), Pitt's Provost Award in Excellence in Mentoring (2015), and UMass Outstanding Award in Education (2019). He is an ACM distinguished scientist and a senior member of the IEEE. He received the BS degree from the University of Athens, Greece, and the MS and PhD degrees from the University of Massachusetts at Amherst.

Dr. George Papastefanatos obtained his PhD from National Technical University of Athens and since 2009, he is a research associate at Athena Research Center, Greece. George has more 60 publications in international conferences and journals and has co-authored 3 chapters in books in areas related to indexing and query optimization, data integration, data visualization and visual analytics. Three of his articles has been selected as Best Papers in International conferences, he has supervised and contributed to the development of more than 5 prototype tools related to Data Visualization and he served as a keynote speaker in DESWeb2017, (workshop in ICDE 2017) for linked data visualization. He is co-organizing the International Workshop on Big Data Visual Exploration and Analytics, an annual event co-occurring with EDBT and served as Guest Editor of the special issues "Interactive Big Data Visualization & Analytics" and "Big Data Visualization, Exploration & Analytics" of the Big Data Research Journal. In 2018, George was one of the receivers of a postdoc research grant in the area of data visualization and visual analytics.

Mohamed Sharaf is an Associate Professor in Computer Science at the United Arab Emirates University (UAEU), which he joined in 2019. Prior to that, he held positions as a Senior Lecturer at the University of Queensland, and a Research Fellow at the University of Toronto.

He received his Ph.D. in Computer Science from the University of Pittsburgh in 2007. His research interest lies in the general area of Data Science, with a special emphasize on large-scale big data analytics, interactive human-in-the-loop data exploration, and scalable data visualization.

Copyright @ 2020, Gennady and Natalia Andrienko, Steven Drucker, Jean-Daniel Fekete, Danyel Fisher, Stratos Idreos, and Tim Kraska , David Auber, Nikos Bikakis, Panos K. Chrysanthis, George Papastefanatos, and Mohamed Sharaf, All rights reserved.