BIG DATA VISUALIZATION AND ANALYTICS: FUTURE RESEARCH CHALLENGES AND EMERGING APPLICATIONS - PART 2

SIGMOD Blog - JUNE 26, 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 second part, we present the views of Guoliang Li, Kwan-Liu Ma, Jock D. Mackinlay, Antti Oulasvirta, Tobias Schreck, Heidrun Schumann, Michael Stonebraker.

Visualization Processing

by **Guoliang Li**

Future Challenges.

– ***Automatic visualization.*** Existing data visualization methods require uses to write visualization queries. However, in many cases, users cannot precisely write queries, because it is hard to understand all the underlying data, e.g., in data lake. Thus, *it is important to automatically visualize the data, including visualization-aware data discovery, data integration, and data cleaning*.

– ***Federate visualization.*** Most of data visualization systems do not consider data privacy. *It is challenging to support privacy-preserving visualization. Furthermore, it is more promising to support federate visualizations on data across different sources.* Considering two companies A and B, A requires to use the data of both A and B to do visualizations, but B cannot release the data to A and can only release some privacy-preserved data.

– ***Visualization benchmark.*** Like ImageNet or the classic TPC benchmarks, *it is important to develop benchmarks for performance and recommendation*. *The benchmarks should be faithful to the visual analysis tasks*, *provide reusable traces and data, and in the case of recommendation, have high coverage and quality of its labels*.

– ***Visualization databases.*** Databases are widely used and deployed in many real applications. It calls for a visualization database that (1) *designs special operators for visualizations and supports optimizations for multiple operators*; (2) *efficiently visualizes a large-scale data*; (3) *supports interactive visualizations with users*; (4) *supports collaborative visualizations by multiple users on smart devices*; (5) *accelerates the performance with new hardware*.

Emerging Applications.

– ***Visualization on clouds.*** Most of existing visualization systems are on premise or on smart devices. However, they cannot handle Big Data and thus it is promising to utilize client-cloud collaboration techniques to support data visualizations, where the data and visualization results are automatically transmitted between clients and clouds. *It requires to address the challenges of data consistency, data security and data sampling*.

– ***Visualize the time-series data for 5G and IOT.*** The 5G age is coming and there are generating more and more high-*volume and high-speed time-series data*. *It is promising to visualize the time-series data and help users find insights from them instantly*.

– ***Visualization for debugging.*** *It is promising to utilize visualizations for debugging, including both bug debugging and data debugging*. The former utilizes visualizations to find bugs in a workflow, e.g., finding the root causes why a SQL query is slow. The latter aims to find the reasons why data visualization result is wrong, e.g., wrong data sources, wrong parameters, wrong visualization queries. *There are various applications in healthcare and education*.

Guoliang Li is a full professor in the Department of Computer Science, Tsinghua University, Beijing, China. His research interests mainly include data cleaning and integration, crowdsourcing, data visualization, and database systems for machine learning. He got VLDB 2017 early career research contribution award, TCDE 2014 early career award, CIKM 2017 best paper award, and best paper nominations of ICDE 2018 and KDD 2018.

The interplay of ML and Visualization

by Kwan-Liu Ma

Future Challenges. Many Big Data problems find machine learning a promising solution. *One existing challenge is how to effectively assist machine learning with visualization, which includes both interpreting and optimizing machine learning* [11][18]. The inverse problem probably interests visualization researchers more. It has been shown machine learning can assist visualization design and generation [27][40][26]. *Usability of visualization must be validated and enhanced* before it can make a true impact to Big Data applications. A promising approach to enhancing the usability of visualization is exactly to add intelligence to the overall process of selecting data, feature extraction, visual encoding, layout generation, annotation, rendering, and response to interaction, allowing the user to focus on perceiving and interpreting information. This marriage makes computer and human each does its best. Finally, e*thic including privacy* is another area of increasing importance that must be addressed when designing visualization solutions for exploring Big Data [12][48][49].

Emerging Applications. When involving Big Data, there are tremendous opportunities to employ visualization and analytics in a wide range of applications from *health care, manufacturing, IoT, cybersecurity, to learning and design*. Progress in each area is being made but the deployment, adoption, and thus benefits of visualization remain to be seen. *One area of my personal interest is the analysis of agricultural and* *environmental data collected with all means of monitoring*. The high dimensional and heterogeneous nature of the collected data presents tremendous opportunities for visualization research and innovations, from designing data aggregation and visual composition strategies, resolving visual scalability for adapting to different display and interaction types, enhancing computing scalability to meet the required level of interactivity, and addressing uncertainty due to data generation as well as data and visual transformations [14][51] [36][47].

Kwan-Liu Ma, an IEEE Fellow, is a distinguished professor of computer science at the University of California, Davis. He leads VIDI Labs and the UC Davis Center of Excellence for Visualization. Professor Ma received his PhD degree in computer science from the University of Utah in 1993. His research interests include visualization, computer graphics, high-performance computing, and human-computer interaction. Professor Ma was a recipient of the NSF PECASE award in 2000 and the IEEE VGTC 2013 Visualization Technical Achievement Award. He has served as a papers co-chair for SciVis, InfoVis, EuroVis, PacificVis, and Graph Drawing, and as an associate editor of IEEE TVCG (2007-2011) and the Journal of Computational Science and Discoveries (2009-2014). Professor Ma presently serves as the AEIC of IEEE CG&A, and on the editorial boards of the Journal of Visualization, the Journal of Visual Informatics, and the Journal of Computational Visual Media.

Humans and computers: a monitoring partnership

by **Jock D. Mackinlay**

Emerging Applications. *One of the biggest opportunities for Big Data innovation is a more natural, more intuitive integration of the complementary skills of humans and computers*.

Humans have incredibly rich understandings of the world that empowers their work with data that describes the world. Computers have superficial understandings of the world but can process Big Data 24´7´365 with sophisticated computations based on statistics and machine learning.

Imagine the potential impact of a better partnership between humans and computers, one in which computers monitor the data coming from the world and advise humans where to visually explore and analyze data.

User interfaces that support effective visual analysis will allow humans to quickly determine when computers have generated false positives. Humans will also be able to dive in deeply, especially with their colleagues, to explore something that deserves human attention and decision-making. Human deep dives can lead to significant value to organizations. Additionally, false positives can be reduced over time by using telemetry from human visual analysis to update the machine learning that decides when to advise humans.

Future Challenges. This vision of a *monitoring partnership* involves a broad range of interesting research challenges, including the following:

– ***Scale*** is a traditional research challenge for the database community, including the processing of federated queries that combine long and/or wide data tables. *Today, the scale of monitoring is bottlenecked by data scientists and analysts who are using their rich understanding of the world to tell computers where to monitor. However, important changes in the world can often be in data not seen by humans*. Imagine scaling our computer monitoring to ALL the data *streaming* from the world using machine-learning algorithms trained on human data work, including the directives from data scientists and analysts, to identify important changes in the world.

– ***Speed*** to process streaming data is important in scenarios that lead to human analysis like intrusion detection and when machine-learning modules are autonomously deployed. *Computer architectures built for speed also need to support the partnership with humans, including humans using their rich understanding of the world to supervise the autonomous machine-learning modules*.

– ***Automatic data stories*** are key to fostering the monitoring partnership. Computers need to explain why they are advising humans to engage in visual analysis. *A key research challenge is explaining the relevant statistics and machine learning to support the human deep-dives when the computer's advice is not a false positive, particularly explanations for humans that are not data scientists or analysts*.

This small list is intended to suggest the range of interesting research challenges associated with the monitoring partnership vision. Toward the human end of the research range, my area of expertise, it is important to embrace the richness of the world. Even unicorn data scientists typically need to collaborate with other people in their organizations to be effective. Each organization (and often parts of organizations) have unique data they need to monitor. Most people are engaged in the mission of their organization, which means they have latent interest in the data describing their organizations. Therefore, *almost everyone in an organization will benefit from partnering with computers monitoring changes in the world*.

Jock D. Mackinlay is the first Technical Fellow at Tableau Software, an expert in visual analytics and human-computer interaction. He believes that well-designed software can help a wide range of individuals and organizations to work effectively with data, which will improve the world. Jock joined Tableau in 2004 after being on the PhD dissertation committee of Chris Stolte, one of the cofounders of Tableau. Jock got a computer science PhD at Stanford University in 1986 and joined the research team at Xerox PARC that coined the phrase "Information Visualization". In 2009, Jock received the IEEE Visualization Technical Achievement Award for his seminal technical work on automatic presentation tools and new visual metaphors for information visualization.

The greatest challenge is, still, the human

by **Antti Oulasvirta**

Future Challenges. The greatest challenge is, still, the human. You cannot do a visualization without understanding human perception. But when it comes to that, visualization as a field relies excessively on trial and error and empirical testing. The ideal should be a field that has human factors integrated directly throughout the visualization pipeline, akin to mature engineering disciplines where theories and models (e.g., from physics) are used to derive optimal solutions. In the case of visualization, that necessarily means models of human perception, attention, and cognition -instead of physics- as these determine success/failure across visual analytics tasks. What does this mean in practice? One recent example is our work with Luana Micallef, who sadly passed away recently, and Tino Weinkauf, on perceptual optimization of scatterplots [30]. We used models from human vision research as objective functions and generated visualizations computationally, such that the underlying structure of the data could be better perceived by human observers [30]. This becomes more and more important as the complexity of the dataset grows. But we can go further than generative design! We can better ***explain*** data. Using methods like Bayesian optimization, we can now much better fit models of human behavior directly to observational data, such as clickstream data, in order to make better sense of them [21]. We can go way beyond variables and actually start explaining the processes that generated the observed data! We can also use models from psychology to ***adapt*** visualizations to individuals and their tasks and preferences. Models can be used to compute visualizations that are strike optimal trade-offs among tasks and user groups. I firmly believe that the combination of human vision research and modern computational sciences offers a new foundation to visualization research that is less reliant on heuristics, manual tuning, and empirical testing, and takes a step closer to mature engineering disciplines.

Emerging Applications. I don't think there is a need to open new applications for Big Data visualization and analytics, rather to solve the core problem really well, and in a principled manner. ***Why*** *should a particular visualization favored over another one in some context*? ***Why*** *should one choose particular design parameters over other ones*? *What are the* ***limits*** *of a particular type of visualization, what can it do and -more importantly- what can it* ***not*** *do*? Core problems like these need principled answers, and the answers will not come from computer science only but will need to seriously engage with vision science. *The main application of Big Data visualization and analytics will be just doing it 10´ better than now: better efficacy and higher efficiency*.

Antti Oulasvirta leads the User Interfaces research group at Aalto University and the Interactive AI research program at FCAI (Finnish Center for AI). Prior to joining Aalto, he was a Senior Researcher at the Max Planck Institute for Informatics and the Cluster of Excellence on Multimodal Computing and Interaction at Saarland university. He received his doctorate in Cognitive Science from the University of Helsinki in 2006, after which he was a Fulbright Scholar at the School of Information in University of California-Berkeley in 2007-2008 and a Senior Researcher at Helsinki Institute for Information Technology HIIT in 2008-2011. During his postgraduate studies in 2002-2003, he was an exchange student at UC Berkeley's Neuropsychology Lab. He was awarded the ERC Starting Grant (2015-2020) for research on computational design of user interfaces. Dr. Oulasvirta serves as an associate editor for ACM TOCHI and has previously served International Journal of Human-Computer Studies, as well as served as a column editor for IEEE Computer. He frequently participates in the paper committees of HCI conferences, including the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI). His work has been awarded the Best Paper Award and Best Paper

Honorable Mention at CHI twelve times between 2008 and 2019. He has held keynote talks on computational user interface design at NordiCHI'14, CoDIT'14, EICS'16, IHCI'17, ICWE'19, and Chinese CHI '19. He is a member of ELLIS (European Laboratory for Learning and Intelligent Systems). In 2019, he was invited to the Finnish Academy of Science and Letters.

Guidance-based visual analytics systems

by **Tobias Schreck**

Future Challenges. Visual analytics approaches integrate interactive data visualization with automatic data analysis methods, supporting expert-in-the-loop exploration of large data sets for potentially interesting and actionable endings. *As data sets grow larger, and become more heterogeneous, there is increased risk that users may have difficulties to effectively navigate the data and take longer to arrive at insights.* For example, as dimensionality and size of data sets grow, there is an exploding amount of possible data subsets to select and visualize, in which users may get lost navigating.

Approaches to user guidance, adaption and personalization in the visual data exploration process may help this situation. Recent works [9][10] have identified requirements and desiderata for guidance-based visual analytics systems. For example, different types of guidance for orienting, directing or even prescribing visual exploration paths can be distinguished. *While the need and potential of guidance in the visual analytics process is eminent, how to design, integrate and evaluate actual guidance approaches in the visual analytics process, supporting specific tasks of specific users on specific data, is less clear to date*. There exists a large and exciting theoretical and applied research space for guidance approaches in many areas, and we can build on a number of promising starting points in this direction. For example, mining user interaction patterns can help to classify and predict user interest and search strategy, and help to guide the exploration. For example, recent research has mined user interaction input during search tasks [9], or proposed to derive user interest models from eye-tracking the user [43]. Principles from information retrieval like relevance feed-back and recommending [34] can be adapted for the visual exploration process, promising to arrive more efficiently at relevant findings*. Besides mining user interaction, also approaches for estimation the visual quality and information content in visualizations have been proposed, for different types of visualizations* [25]*, and can be used to rank and suggest candidate views for inspection by users. As a third line of research, approaches for description of design rules and learning-based approaches may allow to automatize the otherwise interactive and open-ended search for appropriate visual mappings of data, and support visualization automation* [33][40].

To summarize, *I see promising research directions in implementing guidance for the visual analytics process, supporting more effective and efficient, adaptive and personalized visual analysis systems*.

Emerging Applications. There are certainly very many promising application areas for Big Data visualization and analytics. Progress has been made already in a number of domains including biomedical applications, financial data visualization, social media and text visualization, just to name a few. *We recently observe strongly growing interest in visual analysis of data in industrial contexts, driven by efforts to collect data from industrial production processes by sensor equipment becoming available, and introduction of data-driven approaches*. In [50], the authors give a survey of relevant use cases and first solutions for industrial data visualization, including internal and external equipment environment visualization, and for purpose of creation, including design, production, testing, and service.

Visual analysis of data in industrial contexts presents challenging problems, due to large amounts of data produced by increasing numbers of sensors applied to the whole production pipeline, and delivering data at high frequency. *The data is often of heterogeneous nature, comprising different units of measurement which interact with each other. The temporal alignment (normalization) of data along possibly long-running production phases poses a challenge with respect to data accuracy and alignment*. *The filtering and selection of relevant data is an-other challenge, which often needs to be found by trial-and-error in lack of best practices and experience*. There exist bodies of data analysis and visualization methods applicable in principle for industrial data, including real-time monitoring, prediction, identification of influence factors (correlations), and anomaly detection. Each of these methods can be very useful to solve problems of resource planning, quality control, or equipment condition monitoring. *How to adequately select, filter, integrate and transform data, and which analysis and visualization techniques to combine, poses an interesting and rewarding application challenge*. *In addition, a research problem we are currently pursuing involves the unification of domain process knowledge with process data patterns in an integrated production information system, supporting planning and monitoring* [44]. While first concepts exist, much needs to be done in the future.

Tobias Schreck is a Professor and head of the Institute of Computer Graphics and Knowledge Visualization at Graz University of Technology. He previously held positions as Assistant Professor with University of Konstanz, Germany, and as Postdoc Fellow with Technische Universität Darmstadt. He obtained a PhD in Computer Science in 2006 from the University of Konstanz. Tobias Schreck works in the areas of Visual Analytics, Information Visualization, and 3D Object Retrieval. He has previously served as a program co-chair for the IEEE Conference on Visual Analytics Science and Technology. He currently serves as an associate editor for IEEE Transactions on Visualization and Computer Graphics.

Not only Big Data is a challenge, but also the large volume of methods and tools

by **Heidrun Schumann**

Future Challenges. Big Data visualization and analytics require a combination of visual, interactive and automatic analysis methods. However, each of these aspects covers quite a lot challenging topics, which are communicated in separate journals, and discussed at separate conferences and workshops. My question is: *How can this diversity be used in order to provide effective and efficient analysis tools that offer this wide range of functions*?

For specific applications, methods from different fields can certainly be inspected and selected on demand. Then, tailored analysis systems can be developed. But, given the large number of applications: *Will we manage to develop independent analysis systems for each application*?

I think, we should also pursue a more generic approach. However, if we take a generic approach, then the question arises: *How to combine different methods from different fields in order to create a unified application*? I don't believe that it makes sense (or is even possible) to integrate all required or desired functionalities into one and the same easy-to-use analysis tool*. So how can different tools and methods be combined to visualize and analyze large data in a uniform way*?

My key message is therefore: *Not only the large volume of Big Data is a challenge, but also the large volume of methods and tools needed to process it.*

Emerging Applications. I identified *climate change and health care as emerging application examples because of their high social relevance*. Although visual analysis methods are already used in both applications, various challenges require the development of new strategies and solutions. These challenges are for example:

– ***Related to the data.*** The data are heterogeneous and complex, come from different sources, are calculated or measured, are subject to uncertainties, and particularly, they are available on different scales.

– ***Related to the users.*** Scientists from several domains need to discuss and analyze the data. For example, in the case of studying the impact of climate change, scientists from ecology, physics, biology, geology, and mathematics compare their data and projections based on different models and parameterizations to predict the impact of climate change on different aspects.

Heidrun Schumann is a professor at the University of Rostock, Germany, where she is heading the Chair of Computer Graphics at the Institute for Visual & Analytic Computing. Her research and teaching activities cover a variety of topics related to computer graphics, including information visualization, visual analytics, and rendering. Heidrun co-authored the first German textbook on data visualization in 2000, a textbook specifically on the visualization of time-oriented data in 2011, and a textbook about interactive visual data analysis that will appear in March 2020. In 2014, Heidrun was elected as a Fellow of the Eurographics Association.

Scalability, scalability

by **Michael Stonebraker**

Future Challenges & Emerging Applications. One of the consequences of having "big data" is that there is a lot of it. The issue of scale breaks many of our cherished assumptions. First the data does not fit in main memory, and any platform that assumes main memory will fail. Also, the data is inevitably skewed – i.e., 90% of the data is in 10% of the available real estate. Hence, the issue of clutter is omnipresent. At scale, it is very costly to prebuild visualization structures, and most data sets are multi-user, leading to different access patterns and use cases. Multi-user data sets are often updated, and assuming the data is static is dangerous. Lastly, terascale and petascale data will always come out of a backing database system, and our community should always deal with server-side DBMSs. As a result, *I think the biggest challenge our community needs to solve is scalability*.

The days of rendering teapots are long gone. Data scientists in industry have to support user communities of business analysts, physical scientists and decision makers. Their objective is to gain actionable insights from very large data sets. If the domain is simple, they will go straight to analytics. In complex domains (where it is not clear what question to even ask), scalable visualization will be the "go to" technology. *Hence, supporting data scientists is likely to be the "sweet spot" for Big Data visualization technology*.

Dr. Stonebraker has been a pioneer of data base research and technology for more than forty years. He was the main architect of the INGRES relational DBMS, the object-relational DBMS, POSTGRES, the column store C-Store, the OLTP engine H-store, the array store SciDB, and the integration engine Data Tamr. Over his career, Stonebraker has started 10 venture capital backed software companies, and presently serves as Chief Technology Officer of Paradigm4 and Tamr, Inc. He was awarded the IEEE John Von Neumann award in 2005 and the 2014 Turing Award, and is presently an Adjunct Professor of Computer Science at M.I.T.

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. Kangasrääsiö, 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. Micallef, et al.: Towards Perceptual Optimization of the Visual Design of Scatterplots. TVCG 23(6), 2017
[32] L. Micallef, 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) .