SIGHT: System for Interaction Gathering using HTML event Tracing

APOORVA NORI, New York University
COLE POLYCHRONIS, Westminster College
LANE HARRISON, Worcester Polytechnic Institute

Abstract–Creators of web-based interactive news visualizations currently receive little to no feedback concerning how their audience uses the visualizations that they deploy. Previous research has demonstrated the possibility to collect this interaction data; however, it often requires significant expertise, time and effort on the part of the developer and the journalist themselves. Interaction data is crucial to revealing how a particular audience is navigating any given visualization. Collected information can be used to improve news story designs and help justify the time and effort creators spend in the development of these visualizations. There is currently no widely adopted mechanism for collecting and analyzing visualization interaction data; therefore, any improvement can have significant impact when distributed across these existing visualizations. In this paper, we propose a model for researchers to follow to develop meaningful data collection platforms for journalists that would enable them to convert their existing visualizations from artifacts to assets that can help guide future visualization design. We also formalize three major challenges that future researchers face on the path to developing platforms that are useful to journalists: minimizing interactions latency, ensuring ease of deployment for the developer, and determining what data is most important to collect.

ACM Reference format:
Apoorva Nori, Cole Polychronis, and Lane Harrison. 2018. SIGHT: System for Interaction Gathering using HTML event Tracing. 1, 1, Article 1 (January 2018), 5 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the advent of an increasingly "big-data" driven world, the way people consume news has drastically changed over the past decade. Though visualization techniques have long been integrated into various news articles, recent advancements in data visualization techniques have integrated big-data findings into an easily comprehensible platform for viewers. This significant transformation of journalism in the digital age can be attributed to an increasingly visually literate consumer. Previous research has even supported this transformation, suggesting that there exists a positive correlation between the integration of visualization techniques and an audience's degree of political awareness [Boy et al. 2015].

The next notable phase in this shift is the introduction of interactive elements to these already-visual stories. The most recent Global Data Visualization Market Report has valued the visualization industry at over 4 billion dollars [Costello Costello]. The creators of these expensive visualizations currently receive little or no feedback about how people actually use the visualizations they deploy. Any previous research that has considered a similar problem has only addressed the interaction between a single user and an interactive display, rather than an aggregate collection of the success or failure of an instrumentation based upon overall interaction. For example, leading news organizations such as The New York Times have invested significant amounts of money in engaging readers through interactive visualizations; however, because much of the interaction takes place solely on the reader's device, the creators are provided with little no no feedback on their valuable visualizations.

Current methods for collecting interaction data are demanding in terms of both time and expertise on the creator's end, often more convoluted and complicated as the complexity of the interactive increases. Furthermore, there is no baseline of understanding for whether or not certain feedback elements would prove to be worthwhile in providing meaningful feedback. The goal at news organizations is often to turn extensive data sets into compelling and engaging stories for the general public; however, there exists no way of determining whether or not this goal is truly being met. An uncovering of this interaction data may reveal how readers are navigating a visualization and could potentially lead to A/B testing for competing designs and potentially increase and enhance national data literacy.

Recent advancements and innovations in the sphere of data analytics, particularly visual interaction analysis, may allow us to transform these interactive storytelling methodologies from artifacts into design assets as they may help us to answer how users navigate these visualizations. This paper focuses on the current gap between the developers with journalistic agendas and the interactions of their audiences. Subsequently, we propose a research agenda to bridge the aforementioned gap, based upon the capabilities of cutting-edge data analytics technology and the challenges that the developers of such a tool may face.

2 BACKGROUND

There has been a considerable amount of research on developing platforms to collect user interaction with websites, such as Hong et. al's work with WebQuilt [Hong and Landay 2001] and Atterer et. al's work with UsaProxy [Atterer et al. 2006]. WebQuilt and UsaProxy both redirect web traffic through proxies in order to log information about how the user is interacting with a given website, but operate at different granularities. WebQuilt constructs diagrams of the common paths that users take between pages of a website (navigating from the home page to the about page, etc.), while UsaProxy tracks lower-level interactions such as clicking on a link, moving the mouse, or pressing a key.

However, neither of these programs are well-suited to track information about data visualizations that reside on a web page, whether they are static or interactive. WebQuilt operates at too high of a level

This work is supported by the National Science Foundation Research Experiences for Undergraduates (REU) program, Grant 1560229.
Author's addresses: A. Nori, Business and Computer Science Departments, New York University; C. Polychronis, Math and Computer Science Departments, Westminister College; L. Harrison, Computer Science Department, Worcester Polytechnic Institute;.

to track nuanced user exploration of a visualization. Conversely, the data that UsaProxy collects is too low level, providing general information about user input while not tying it to specific sections, objects, or data points in a visualization. A host of other proposed platforms that are well-suited for collecting interaction data concerning websites ([Heer 2002], [Beymer and Russell 2005], [Brown et al. 2014], [Bavoil et al. 2005], [Cowley et al. 2005]) are not as suitable for collecting visualization-centric interaction data.

Even some of the most popular commercially-available software, such as Google Analytics, are not well-suited for collecting interaction data about visualizations. Google Analytics is appealing to many developers due to its ease of use; one simply has to register an account and then copy and paste tracking code into the web page's source code. This enables the client to track user demographic data, such as the average number of visits a web page receives per week, how long the average user stays on the page, and what kinds of devices and browsers are being used to visit the website. Adobe's competing service, the Adobe Activity Map, offers slightly more specific data such as the raw number of clicks on a given button or link, but still does not offer the capability to capture user interaction with data visualizations. What's more, Adobe's Activity Map requires developers to manually add the tracking code to every item that they are interested in tracking, making it far more time intensive for developers to deploy.

3 RELATED WORK

Prior to developing our proposed model, it is important to first describe some of the other significant work surrounding visualization interaction analysis and usability instrumentation that informs our approach. Figure 1 shows a provenance chart that compares the most relevant work in this sphere. Each of the studies is distinguished by these four user-study focused factors: *population*, *study*, *number of participants* and *latency*. Additionally, the following four instrumentation factors are considered: *area*, *scope*, *automation* and *data type*. The given columns are defined by the metrics in Figure 1.

User Study	
population-intended	The group considered when writing the paper (designer vs. developer vs. user). Dark Green cells denote that both designer/developer and user were addressed.
user-study-conducted	Whether or not experimentation included a user study. Dark Green cells denote a True value, meaning that a study was carried out.
user-study-#ofparticipants	The quantity of subjects *not trials* included in a single study of the paper. Dark Green cells denote a quantity of subjects >25.
latency-mentioned	Whether or not the notion of latency/graphical slowdown is acknowledged. Dark Green cells denote that a paper explored the idea of latency. Light Green cells denote that the concept was mentioned but not explored.
Instrumentation	
instrumentation-area	The platform on which instrumentation was explored or carried out. Dark Blue cells denote that a web-based visualization was studied. Light Blue cells denote that website to application visualizations were explored.
instrumentation-scope	The breadth of visualization explored. Dark Blue cells denote that interactive elements both inside and between web pages were explored.
instrumentation-automatic	Whether or not the instrumentation and data capturing aspects of the project were automatically logged. Dark Blue cells denote that the entire process was automatic. Light Blue cells denote a semi-automatic process.
instrumentation-data	The format of the resulting captured interaction data. Dark Blue cells denote a resultant sophisticated dashboard. Light Blue cells denote a platform more sophisticated than a simplistic log file.

Fig. 1. Summary of Provenance Chart Factors

The columns outlined in red in Figure 2 are demonstrative of the "hole" in existing research, substantiating and justifying a need for further research to meet the advancements that have taken place in the field of interactive visualizations. The first column denoted in red, *population-intended*, shows that only 5 papers of the collected sample consider both the developer/designer of a given visual and the ultimate user. While many other papers address one or the other, it is the communication of the two together that will close the feedback loop between creator and user, which in this case can be considered the journalist and the audience.

Next, the column *latency-mentioned* reveals that only 9 papers in the studied sample addressed the issue of latency within these visualizations, 3 of those 9 having only minimally explored the concept. The lack of exploration in terms of latency, or interaction speed in context of an implemented visualization is detrimental to the conceptualization of a tool that closes the aforementioned feedback loop. If the cost of implementing a tool that can collect sophisticated user interaction data is that the visualization itself becomes slower and more difficult to navigate, it can altogether compromise the usability of an implemented visualization, causing a drastic decline in audience and user participation.

In terms of instrumentation, the column *instrumentation-scope* is particularly noteworthy in that it exhibits a need for further research exploration into dynamic visualizations as opposed to static ones. The previously mentioned state-of-the-art tools such as Google Analytics provide basic feedback for static web page elements. Only 10 of the sampled 36 research papers consider user interaction both inside and between web page elements, meaning that any dynamic-specific traits that a news story developer would implement such as hovering data revelation properties would go undetected. Any worthwhile analysis of user-interaction data would have to account for the large quantity of complex functionalities implemented by a developer.

Finally, the last column, *instrumentation-data*, is also noteworthy as it corroborates that much of the research is one-sided and that most conducted studies yielded log reports of collected interaction data. This is particularly infeasible for web-based data visualizations instrumented by news organizations because the writer of the story is often times not trained in data analysis techniques. Therefore, when presented with a log file of immensely convoluted interaction data, it is essentially as useful as no data collection at all. A more sophisticated dashboard or graphical representation however, as implemented by a minimal 6 papers and merely explored by another 4 papers, would be much more useful to a given journalist. ie. Collected data must be in a comprehensible and easily digestible format in order to close the loop entirely between creator and user.

4 PROPOSED MODEL

The proposed new model in Figure 3 outlines visualization deployment on the web. Red lines indicate the current model of interactive visualization deployment, where creators, such as those from web-based news organizations, deploy a visualization, but receive little or no feedback about how audiences actually interact with it. Our new model implements an instrumentation library (green lines) that minimizes the time, cost, and effort on the part of creators while maximizing visualization interaction data throughput to a proposed analytics platform (yellow lines). The impact of instrumentation on user experience and performance (purple line) is also assessed. This new deployment model closes the feedback loop in interactive

	User Study				Instrumentation			
Citations Abbreviated	population-intended	user-study-conducted	user-study-#ofparticipants	latency-mentioned	instrumentation-area	instrumentation-scope	instrumentation-automatic	instrumentation-data
amar2005analytic	designers	TRUE	50	TRUE	website,app	inside	FALSE	semi-structured, varied
andreasen2007happened	developers	TRUE	24	FALSE	website	inside, between	FALSE	video, voice recording
atterer2006logging	designers	FALSE		FALSE	website, app	inside, between	TRUE	mouse,key movements
atterer2006tracking	developers, designers	TRUE	n/a	FALSE	website, app	inside, between	TRUE	click, scroll, key presses, resize
barik2016bones	developers, designers	FALSE		TRUE	app	inside	FALSE	semi-structured, varied
bavoil2005vistrails	user	FALSE		FALSE	website	inside	TRUE	low-level event log
beauvisage2009computer	developers	TRUE	1,434	FALSE	app	inside	TRUE	low-level event log
beymer2005webgazeanalyzer	user	TRUE	n/a	TRUE	website	inside	FALSE	recorded eyegaze video
blascheck2016va	developers	TRUE	n/a	FALSE	app	inside, between	semi-automatic	low-level interaction logs
boy2015storytelling	designers	TRUE	n/a	FALSE	website	inside	TRUE	low-level event log
brown2014finding	developers	TRUE	118	FALSE	website	inside	TRUE	mouse key events, paths
callahan2006vistrails	users	FALSE		FALSE	website	inside	TRUE	low-level event log
cowley2005glass	analysts	TRUE	n/a	FALSE	website	inhside, between	TRUE	low-level event log
dabek2017grammar	developers, users	TRUE	300	mentioned not explored	website	inside	FALSE	survey results, tracking log
davidson2008provenance	analysts	FALSE		TRUE	website	inside, between	semi-automatic	tracking log
dou2009recovering	developers, designers	TRUE	10	FALSE	website	inside	TRUE	low-level event log
eltayeby2016survey	developer	FALSE		FALSE	website	inside	TRUE	low-level event log
endert2012semantic	developer, designer	FALSE		mentioned not explored	website	inside	TRUE	Spatial Metaphor
gotz2008characterizing	developers, users	TRUE		FALSE	website	inside	FALSE	mouse and key movements
gotz2008empirical	designers, users	TRUE		FALSE	website	inside, between	FALSE	survey, video recordings
gutwin2006improving	designers	TRUE	12	FALSE	app	inside	TRUE	listbox data set
hanel2016towards	developer	FALSE		TRUE	website	inside	TRUE	low-level event log
heer2002capturing	designers	FALSE		FALSE	website	inside, between	TRUE	overall page interactions
heer2008design	designers	FALSE		FALSE	website	inside	FALSE	video recordings
hilbert2000extracting	developer	FALSE		FALSE	app	inside	TRUE	survey responses
islam2016towards	analyst	FALSE		FALSE	websites	inside, between	TRUE	provenence graphics
kandel2012enterprise	developers	TRUE	35	FALSE	n/a	inside	FALSE	semi-structured, iterative coding
lipford2010helping	analysts	TRUE	10	FALSE	website	inside	FALSE	screen recording
ottley2015personality	developers, users	TRUE	54	mentioned not explored	website	inside	semi-automatic	dendrogram visualization
plumbaum2009semantic	designers	TRUE	n/a	FALSE	website	inside	FALSE	low-level event log
pohl2010exploring	designers	TRUE	16	FALSE	website	inside	FALSE	survey
pohl2016using	designers	TRUE	300	FALSE	website	inside	semi-automatic	log files, thinking aloud data
ragan2015characterizing	designers, developers	FALSE		FALSE	n/a	n/a	FALSE	provenance chart
santana2010summarizing	developers, designers	TRUE	84	FALSE	website	inside	TRUE	Usage Graph
stitz2016avocado	designers, developers	FALSE		FALSE	website	inside, between	TRUE	Provenance Graph
thompson2004here	developer	TRUE	10	TRUE	website, app	inside	TRUE	low-level event log, survey
yi2007toward	developer, designer	FALSE		FALSE	website, app	inside	FALSE	taxonomy

Fig. 2. PROVENANCE CHART. Overview of user study and instrumentation types and purposes in a sample of visualization projects. User Study factors are shown in green and Instrumentation factors are shown in blue. Darker colored cells indicate heavier emphasis, lighter colored cells indicate a lesser emphasis, and white cells indicate an absence of the particular factor within the project. Citations: [Amar et al. 2005], [Andreasen et al. 2007], [Atterer 2006], [Atterer and Schmidt 2007], [Barik et al. 2016], [Bavoil et al. 2005], [Beauvisage 2009], [Beymer and Russell 2005], [Blascheck et al. 2016], [Boy et al. 2015], [Brown et al. 2014], [Callahan et al. 2006], [Cowley et al. 2005], [Dabek and Caban 2017], [Davidson and Freire 2008], [Dou et al. 2009], [ElTayeby and Dou 2016], [Endert et al. 2012],[Gotz and Zhou 2009], [Gotz and Zhou 2008] [Gutwin and Cockburn 2006], [Hänel et al. 2016], [Heer 2002], [Heer and Agrawala 2008] [Hilbert and Redmiles 2000], [Islam et al. 2016], [Kandel et al. 2012], [Lipford et al. 2010], [Ottley et al. 2015], [Plumbaum et al. 2009], [Pohl et al. 2010], [Pohl et al. 2016], [Ragan et al. 2016], [De Santana and Baranauskas 2010], [Stitz et al. 2016], [Thompson et al. 2004], [Yi et al. 2007]

visualization deployment, transforming visualizations from artifacts to design assets.

This model will provide a low-effort means for collecting interaction data in the context of news organizations, leading to potentially richer findings. In some cases, there are as many as hundreds of thousands of visits to a single dynamic news visualization on the web. With a model such as the one outlined above, blank can speak directly to these news organizations with very little effort. The circulation in the model shows a closed feedback loop, directly addressing the hole in communication shown by Column 1 in the Provenance Chart mentioned earlier (Figure 2). The purple line in the model leads directly to a sophisticated dashboard interface for the creator, addressing the "hole" pointed out by the last red column in Figure 2.

5 CHALLENGES TO FUTURE RESEARCH

Through analyzing a swath of the existing literature on interactive data visualizations and proposed platforms for analysis of such visualizations, we have developed the deployment model illustrated in Figure 3 as a framework for future researchers and platform developers to use to help close the "feedback loop" between developers and end-users. In dividing our model into these distinctive aspects (an

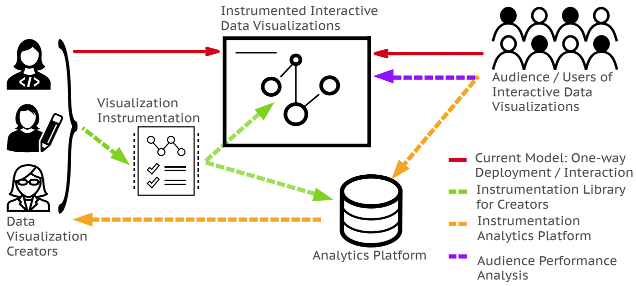

Fig. 3. CLOSING THE FEEDBACK LOOP

Instrumentation Library, an Analytics Platform, and a Performance Platform), several key challenges to future researchers emerge.

Minimizing interaction latency. While the throughput of interaction data to the analytics platform is the driving force behind data visualization analytics research, steps must be taken to ensure that future analytics platforms do not cause noticeable slowdown or latency for the user when interacting with the instrumented visualization. Research from Liu et. al suggest that consistent latency of only about 500 ms is all that it takes to adversely affect the user experience, limiting their ability to draw meaningful conclusions from the data they're presented [Liu and Heer 2014]. Given that maintaining user attention and intrigue is a journalist's top priority when telling stories with data visualizations, this point is crucial.

Ease of deployment for the developer. One of the key factors that makes Google Analytics so popular among web-developers is the ease at which it can be implemented; simply dropping code into the source files. The easier that the deployment of the platform is for the developer, the more impactful it will be for journalists, as it would lower the barrier of entry and allow those with less technical programming knowledge to still make use of robust analytics platforms.

Determine what needs to be collected. Data visualization paradigms like Scalable Vector Graphics (SVG) mean that a broad range of data can possibly be collected to inform how users are interacting with data visualizations. Understanding the scope and granularity of data that will be most useful to collect to present to developers will likely require surveying journalists and visualization professionals alike to find common values, understanding of meaningful interaction, and goals of visualization development.

6 CONCLUSIONS

The case presented here illustrates that visualization analytics holds a lot of potential for journalists, but only if researchers are cognizant of the problems that journalists truly need solved. In surveying the existing literature on analyzing user interaction with data visualizations, several major holes appear that demonstrate that little work is being done to close the feedback loop and allow content creators to gain insight about their audience and how they are interacting with the visualizations they create. To address this current lack of ability to close the feedback loop, we present our model as a framework for future researchers to follow such that developers are given the means to convert their visualizations from stagnant artifacts to valuable assets, providing information that can guide future visualization projects to more fully engage the user. Additionally, we present three major challenges that must be considered by future researchers if they wish to develop a platform that is useful to journalists, given the limited amount of time and resources that they may be faced with: minimizing interaction latency for the user, ensuring the deployment for visualization developers is as easy as possible, and determining what kinds of data should be collected so as to maximize usefulness for journalists.

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on.* IEEE, 111–117.

Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. What happened to remote usability testing?: an empirical study of three methods. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, 1405–1414.

Richard Atterer. 2006. Logging usage of AJAX applications with the" UsaProxy" HTTP proxy. In *Proceedings of the WWW 2006 Workshop on Logging Traces of Web Activity: The Mechanics of Data Collection.*

Richard Atterer and Albrecht Schmidt. 2007. Tracking the interaction of users with AJAX applications for usability testing. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM, 1347–1350.

Richard Atterer, Monika Wnuk, and Albrecht Schmidt. 2006. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web.* ACM, 203–212.

Titus Barik, Robert DeLine, Steven Drucker, and Danyel Fisher. 2016. The bones of the system: a case study of logging and telemetry at Microsoft. In *Proceedings of the 38th International Conference on Software Engineering Companion.* ACM, 92–101.

Louis Bavoil, Steven P Callahan, Patricia J Crossno, Juliana Freire, Carlos E Scheidegger, Cláudio T Silva, and Huy T Vo. 2005. Vistrails: Enabling interactive multiple-view visualizations. In *Visualization, 2005. VIS 05. IEEE.* IEEE, 135–142.

Thomas Beauvisage. 2009. Computer usage in daily life. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* ACM, 575–584.

David Beymer and Daniel M Russell. 2005. WebGazeAnalyzer: a system for capturing and analyzing web reading behavior using eye gaze. In *CHI'05 extended abstracts on Human factors in computing systems.* ACM, 1913–1916. Says Not to Cite.

Tanja Blascheck, Markus John, Kuno Kurzhals, Steffen Koch, and Thomas Ertl. 2016. Va 2: A visual analytics approach for evaluating visual analytics applications. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 61–70.

Jeremy Boy, Francoise Detienne, and Jean-Daniel Fekete. 2015. Storytelling in information visualizations: Does it engage users to explore data?. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 1449–1458.

Eli T Brown, Alvitta Ottley, Helen Zhao, Quan Lin, Richard Souvenir, Alex Endert, and Remco Chang. 2014. Finding waldo: Learning about users from their interactions. *IEEE Transactions on visualization and computer graphics* 20, 12 (2014), 1663–1672.

Steven P Callahan, Juliana Freire, Emanuele Santos, Carlos E Scheidegger, Cláudio T Silva, and Huy T Vo. 2006. VisTrails: visualization meets data management. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data.* ACM, 745–747.

Hector Costello. 2017 Global Data Visualization Market is expected to grow. (????). http://www.reuters.com/brandfeatures/venture-capital/article?id=5123

Paula Cowley, Lucy Nowell, and Jean Scholtz. 2005. Glass box: An instrumented infrastructure for supporting human interaction with information. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on.* IEEE, 296c–296c.

Filip Dabek and Jesus J Caban. 2017. A Grammar-based Approach for Modeling User Interactions and Generating Suggestions During the Data Exploration Process. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 41–50.

Susan B Davidson and Juliana Freire. 2008. Provenance and scientific workflows: challenges and opportunities. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data.* ACM, 1345–1350.

Vagner Figuerêdo De Santana and M Cecília C Baranauskas. 2010. Summarizing observational client-side data to reveal web usage patterns. In *Proceedings of the 2010 ACM Symposium on Applied Computing.* ACM, 1219–1223.

Wenwen Dou, Dong Hyun Jeong, Felesia Stukes, William Ribarsky, Heather Richter Lipford, and Remco Chang. 2009. Recovering reasoning processes from user interactions. *IEEE Computer Graphics and Applications* 29, 3 (2009).

Omar ElTayeby and Wenwen Dou. 2016. A Survey on Interaction Log Analysis for Evaluating Exploratory Visualizations. In *Proceedings of the Beyond Time and Errors on Novel Evaluation Methods for Visualization*. ACM, 62–69.

Alex Endert, Patrick Fiaux, and Chris North. 2012. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 473–482.

David Gotz and Michelle X Zhou. 2008. An empirical study of user interaction behavior during visual analysis. *Technical Report RC24525* (2008).

David Gotz and Michelle X Zhou. 2009. Characterizing users' visual analytic activity for insight provenance. *Information Visualization* 8, 1 (2009), 42–55.

Carl Gutwin and Andy Cockburn. 2006. Improving list revisitation with ListMaps. In *Proceedings of the working conference on Advanced visual interfaces*. ACM, 396–403.

Claudia Hänel, Mohammad Khatami, Torsten W Kuhlen, and Benjamin Weyers. 2016. Towards Multi-user Provenance Tracking of Visual Analysis Workflows over Multiple Applications. (2016).

Jeffrey Heer. 2002. Capturing and Analyzing the Web Experience. In *Proc. CHI 2002*.

Jeffrey Heer and Maneesh Agrawala. 2008. Design considerations for collaborative visual analytics. *Information visualization* 7, 1 (2008), 49–62.

David M Hilbert and David F Redmiles. 2000. Extracting usability information from user interface events. *ACM Computing Surveys (CSUR)* 32, 4 (2000), 384–421.

Jason I Hong and James A Landay. 2001. WebQuilt: a framework for capturing and visualizing the web experience. In *Proceedings of the 10th international conference on World Wide Web*. ACM, 717–724.

Junayed Islam, Craig Anslow, Kai Xu, William Wong, and Leishi Zhang. 2016. Towards analytical provenance visualization for criminal intelligence analysis. In *Proceedings of the conferece on Computer Graphics & Visual Computing*. Eurographics Association, 17–24.

Sean Kandel, Andreas Paepcke, Joseph M Hellerstein, and Jeffrey Heer. 2012. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2917–2926.

Heather Richter Lipford, Felesia Stukes, Wenwen Dou, Matthew E Hawkins, and Remco Chang. 2010. Helping users recall their reasoning process. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*. IEEE, 187–194.

Zhicheng Liu and Jeffrey Heer. 2014. The effects of interactive latency on exploratory visual analysis. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2122–2131.

Alvitta Ottley, Huahai Yang, and Remco Chang. 2015. Personality as a predictor of user strategy: How locus of control affects search strategies on tree visualizations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3251–3254.

Till Plumbaum, Tino Stelter, and Alexander Korth. 2009. Semantic web usage mining: Using semantics to understand user intentions. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 391–396.

Margit Pohl, Günter Wallner, and Simone Kriglstein. 2016. Using lag-sequential analysis for understanding interaction sequences in visualizations. *International Journal of Human-Computer Studies* 96 (2016), 54–66.

Margit Pohl, Sylvia Wiltner, and Silvia Miksch. 2010. Exploring ' visualization: describing different interaction patterns. In *Proceedings of the 3rd BELIV'10 Workshop: Beyond time and errors: novel evaluation methods for information visualization*. ACM, 16–23.

Eric D Ragan, Alex Endert, Jibonananda Sanyal, and Jian Chen. 2016. Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 31–40.

Holger Stitz, Stefan Luger, Marc Streit, and Nils Gehlenborg. 2016. AVOCADO: Visualization of Workflow–Derived Data Provenance for Reproducible Biomedical Research. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 481–490.

Katherine E Thompson, Evelyn P Rozanski, and Anne R Haake. 2004. Here, there, anywhere: remote usability testing that works. In *Proceedings of the 5th conference on Information technology education*. ACM, 132–137.

Ji Soo Yi, Youn ah Kang, and John Stasko. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1224–1231.