Trrack: A Library for Provenance Tracking in Web-Based Visualizations

Zach Cutler* University of Utah
Kiran Gadhave† University of Utah
Alexander Lex‡ University of Utah

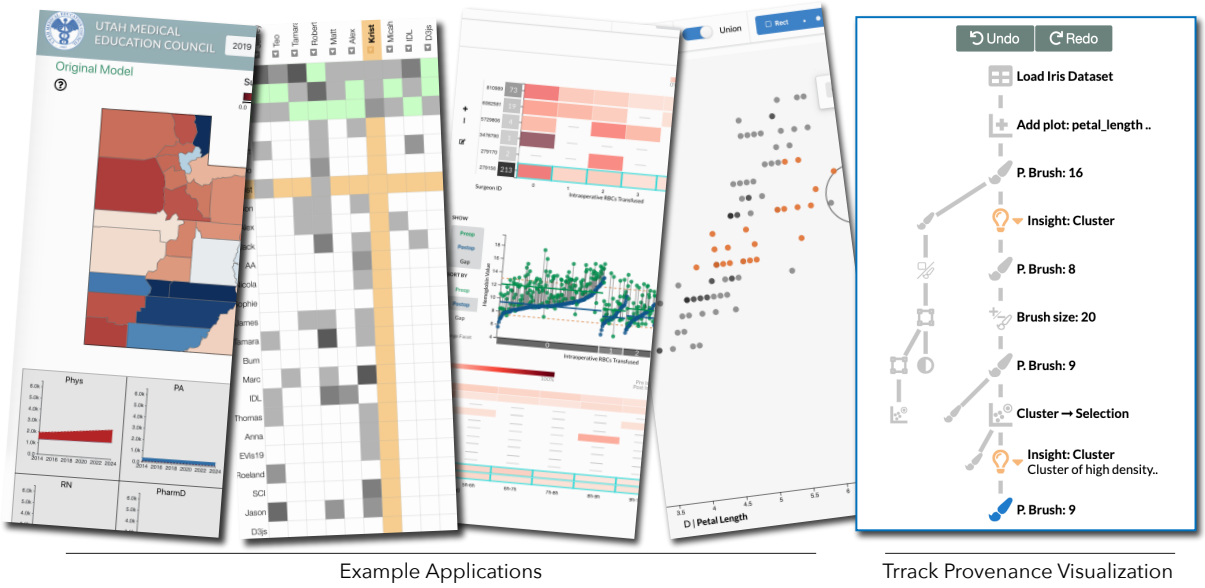

Figure 1: Four example applications using Trrack, our provenance tracking library and TrrackVis, the associated provenance visualization library for different purposes ranging from action recovery to logging for user studies. TrrackVis, shown on the right, utilizes custom icons, annotations, and grouping of nodes.

Abstract

Provenance tracking is widely acknowledged as an important component of visualization systems. By tracking provenance data, visualization designers can achieve a wide variety of important functionality, ranging from action recovery (undo/redo), reproducibility, collaboration and sharing, to logging in support of quantitative and longitudinal evaluation. Yet, for web-based visualizations, there are currently no libraries that make provenance tracking easy to implement in visualization systems. The result of this is that visualization designers either develop ad-hoc solutions that are rarely comprehensive, or don't track provenance at all. In this paper, we introduce a web-based software library — Trrack — that is designed for easy integration in existing or future visualization systems. Trrack supports a wide range of use cases, from simple action recovery, to capturing intent and reasoning, and can be used to share states with collaborators and store provenance on a server. Trrack also includes an optional provenance visualization component that supports annotation of states and aggregation of events.

Index Terms: Human-centered computing—Visualization—Visualization systems and tools

1 Introduction

At least since Ben Shneiderman argued in his "The Eyes Have it" paper that we can't expect users to get it right immediately, and visualization systems need the ability to correct a sequence of actions and replay it, the value of undo/redo and replay [14] (or action recovery) has been undisputed. While professional desktop applications commonly support action recovery, many web-applications, and especially academic visualization prototypes do not. For example, the authors of both Lyra [13] and Data Illustrator [8], two popular web-based data visualization authoring tools, mentioned the lack of undo in their original designs as a cause for concern for users which resulted in less exploration.

Action recovery is one of the purposes of collecting provenance data, but provenance data has many other uses, ranging from recalling an analysis process, to reproducing it, to collaborating, and to logging for evaluation or meta-analysis [12].

Up to this point, designers and developers of web-based visualization tools and prototypes had to develop custom software to capture, store, and utilize provenance information. Doing so, however, can be tedious and is often not in immediate service of the goals of a visualization prototype. Home-grown solutions are also unlikely to leverage the full potential of provenance, such as history visualization, or easy state sharing between remote collaborators.

To remedy this, we developed the Trrack library (the name derives from **r**eproducible **track**ing), our primary contribution, which provides provenance tracking for the purpose of action recovery, reproducibility, collaboration, and logging. Trrack can widely benefit existing and future systems, and support data collection in quantitative and longitudinal evaluations. Trrack is designed to support a wide variety of types of provenance, including provenance of interactions, insights, and rational. Trrack is accompanied by TrrackVis, which provides optional provenance UI elements, such as simple undo/redo buttons, or a sophisticated provenance visualization that can be customized and supports annotation and aggregation of states. Finally, Trrack enables sharing states through a

*e-mail: zcutler@sci.utah.edu
†e-mail: kiran@sci.utah.edu
‡e-mail: alex@sci.utah.edu

URL, as well as provenance data management on a server. Trrack comes with well-documented examples of how to integrate it in web-based visualization applications, and uses tests to ensure code quality. Trrack is available under a permissive open source license at `https://github.com/visdesignlab/trrack`.

Our contribution is in the spirit of an **applications note** —– a short paper with the goal of informing the visualization research community of a robust, well-tested, and easy to integrate technology of broad interest. We believe this is consistent with the renewed calls for applications and systems oriented research that is necessary to tackle the ever-increasing complexity of datasets and challenges.

To show the utility of Trrack we have integrated it in four different visualization tools, including a technique developed in support of a research paper [3], a visualization systems for blood product management, a visualization system for modeling workforce needs in the medical sector, and two network visualization tools used in a large empirical study [9].

2 Design Goals

We designed the library based on our experience with developing a provenance graph for a storytelling application [4]. The overarching design goals are (a) versatility of use — the library should support all different purposes of provenance tracking, and (b) ease of use — developers should be able to track and visualize provenance with minimal effort. Here we list our specific design goals.

Allow Developer Agency. Application developers should have the flexibility to decide which actions to track, to suit their needs. For example, what is sensible to track might be vastly different between a production visual analytics system, and a prototype used in a crowdsourced user-study.

Support Action Recovery. Undo/redo are important in all user interfaces, so that mistakes can be quickly recovered from. In addition to undo/redo, we want to make it possible to quickly browse to any prior state. Knowing they can always return to a previous state no matter how far away they are, analysts might be more willing to try certain actions or investigate paths they otherwise wouldn't have.

Support Reproducibility. Reproducibility of analysis processes is critical in data analysis. However, unlike analysis done in computational notebooks, interactive visualizations are difficult to make reproducible. We strive to not only capture interactions, but also annotations so that user intent and reasoning can be captured as well.

Support Collaboration. It is rare that analysts work in a vacuum: they either need to share and communicate their results, or collaborate with other analysts on the same project. A provenance tracking library needs to support both. To give developers flexibility, we envision two ways of collaborating: A light-weight approach of sharing the state of an application by copying the browser URL, and a more sophisticated approach to share the full analysis provenance.

Support Meta-Analysis. We want to design our library to also support collecting information about usage, either for analysis of long-term use in the field, or for controlled studies. The requirements for action-recovery/reproducibility are different from meta-analysis; usually the latter requires more fine-grained logs, that can be a hindrance for the former. It is our goal to design our system such that both can be supported simultaneously. Meta-analysis also requires that the provenance data can be exported in a format that is amenable to further processing. Finally, unlike traditional logging, we want to be able to jump into any recorded session, so that the context of, e.g., a performance problem can be investigated.

Support Annotation, Highlighting, Bookmarking. As already alluded to when discussing reproducibility, it is important to be able to bookmark states, so that they can be quickly found and retrieved, and annotate states, so that users or systems can provide context, including insights or rationals for a specific action. More generally, a developer might want to store a variety of meta-information with individual states, which should be supported by a provenance tracking library.

Provide UI elements and Provenance Visualization. In addition to provenance tracking, we also intend to provide user interface elements that can be used to navigate provenance data, if desired by the developer. A provenance visualization should be able to manage the whole feature set of the library, including branching states, annotations, bookmarks, etc. Also, as provenance data can quickly grow, it needs to be able to manage large interaction graphs.

Technical Goals Large visualization with complicated states can lead to a very large provenance graph if every interaction is tracked or the tracking continues for long periods of time. It is therefore necessary that a library is designed with efficient storage in mind. A provenance library also should support the quick recovery of a particular state.

3 Related Work

There is a long history of research on provenance capture and visualization. Here we present only a small subset of related papers that are particularly relevant to our work. We refer to Xu et al.'s recent survey [18] and Ragan et al.'s analysis for a more complete picture [12].

Provenance can either be captured through explicit workflow modeling systems [2], or by capturing the analysis process in an interactive system [10]. We are interested in the latter approach, as it can be easily integrated in arbitrary systems. This tracking based approach requires that the system either stores each state of a system, a sequence of actions, or a combination thereof [5] with various implications for the abilities of such systems. Examples include the graphical histories by Heer et al. [5], and CzSaw [6]. Multiple systems present provenance data as node-link diagrams [7, 15, 17] including some of our own work [4, 16].

As far as logging for evaluation is concerned, tools such as EvalBench [1] or GraphUnit [11] are examples of frameworks that provide logging capability, but neither captures a complete provenance graph that can then be re-played.

Even though undo/redo is not widely used in web applications, there are examples of systems and non-academic software libraries that provide related functionality. Many document editors, such as the Google suite of productivity tools, have some version of undo/redo. It is however unclear how these tools are implementing action recovery, and most are likely not using publicly available libraries to do so. There are a few front-end libraries which offer basic undo/redo capabilities. Redux [1] is a popular library used with React which allows users to store past and future states for undo/redo. NgRx [2] is a similar library for use with Angular. While allowing for basic undo/redo, these libraries do not allow for complete action recovery for every previous state. Branching off from the original path will result in future states being permanently lost, and most implementations will limit how many previous states get stored. None of these libraries visualize provenance.

4 Trrack Design

Here we describe the design decisions that went into the development of the Trrack library, as motivated by our design goals. We also briefly give an architectural overview, and describe how Trrack interacts with a visualization application (see Figure 2).

Trrack uses a provenance graph approach where each action that is recorded results in a new node in the graph. Nodes can be attached at any point in the graph, following the branching model of provenance.

[1] `https://redux.js.org/`
[2] `https://ngrx.io/`

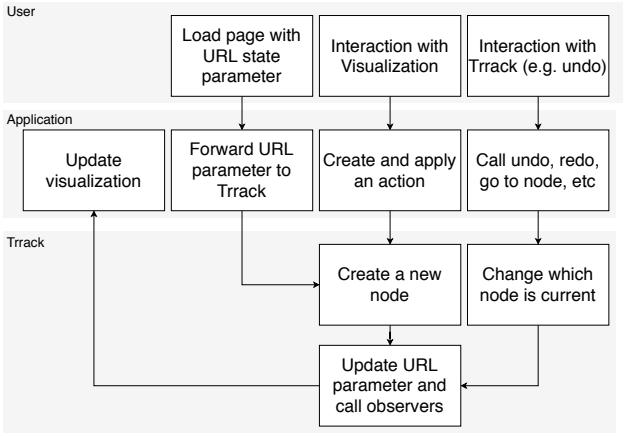

Figure 2: The relationships between the user of a visualization application, the application itself, and the Trrack library for storing provenance.

To utilize Trrack, developers must define a state which fully describes their application. Developers may then add observer functions to individual keys in the state. These observers will be triggered if that particular part of the state changes, either from the addition of a new state or from changing the current node in Trrack with functionality such as undo. We recommend only updating the frontend application within these observers. The only other direct interaction that developers must implement is to create and apply an action when a user interacts with the visualization (see Figure 2). This action is what will then create a new node with a modified state in the graph. We provide a repository with examples of different complexity to illustrate this approach[3].

Trrack Architecture The common storage types for action recovery systems are state based and action based [5]. **State-based** systems store the user defined state of their application at every node, allowing for instant jumps to any node in the history, no matter how large it grows or how far apart two nodes are. The down-side of this approach is inefficient use of memory or disk space, as storing the state frequently can result in significant storage requirements, especially in long-running applications with a complex state. **Action-based** systems store the action required to get from one node to the next. This can lead to slow performance when jumping between states since actions must be applied sequentially to maintain proper recovery. The advantage of this approach is that it minimizes the storage/memory overhead. Heer et al. [5] also discuss **hybrid approaches**, where action-based systems periodically store a state to increase performance when loading a previous state.

In Trrack, we use a different approach: **differential states**. To ensure quick loading of arbitrary nodes of the provenance graph, we store states, and require developers to define a state to be used in Trrack. However, to address the size issues common with such an approach, we do not store the state at every node. Instead, we store a difference between the current node's state and the last node which stored the entire state (as opposed to the immediately preceding node). We use a constant percentage value as a heuristic to determine how many keys in a state object have changed, and if that threshold is surpassed, we store a full state. This is to ensure that the stored difference doesn't grow larger than the original state, and to minimize the size of differences of ensuing nodes. Developers may also specify that a particular action should always store the entire state, if desired.

It is important to note that this mechanism is completely abstracted from developers using the library. A developer only has to request a particular state, e.g, through the undo function, and receives a corresponding state, as illustrated in Figure 2. In addition to the state information, we also store meta-information about the actions, which are useful for visualization and to maintain a user's mental map.

Ease of Integration Trrack is a JavaScript/TypeScript library. Our goal while developing was to make the integration of the library as seamless as possible. There are two ways to use Trrack: it can take over the state management in the application. This is most useful when designing a new web-based visualization. Alternatively, Trrack can interact with any existing state management solution to track changes to state without affecting the UI. Trrack is designed to be framework agnostic. It can work with vanilla JavaScript, UI frameworks like React and state management libraries like Redux.

Sharing State The ability to rapidly share the current state of an application for the purpose of collaboration is rare among visualization applications. Trrack allows for easy, instant sharing by sending the current URL to a collaborator. The current state of an application will be encoded and added to the URL via a URL parameter. Whenever the state of an application changes, the URL will be updated. When a page with a matching URL parameter is loaded, the Trrack library parses it and returns the desired state to the application.

Persistence By default, the provenance graph maintained by Trrack is stored only in memory. To ensure reproducibility and also collaboration beyond just sharing states, we need to make the provenance graph persistent. To achieve this, Trrack provides interfaces to store the graph. To ensure developer agency, Trrack has functionality for exporting provenance graphs, so that they can be managed the way a developer desires. We also provide a default implementation based on Google Firebase. Users can connect a Firebase database to Trrack during setup, and have every node in the graph stored in Firebase automatically, every time a change is made.

Reproducibility and Capturing Intent An important aspect of reproducibility is understanding the reasoning behind steps taken by the user. For this purpose, each node in the provenance graph stores multiple types of metadata. Every node has an annotation property, and a "type", a user-defined string meant to identify the purpose of the state change at that node. Additionally, a generic object may be defined by the developer and stored with the node. This metadata can be used to capture and later interpret actions taken at each node. This can range from simple bookmarks, to written annotations, to complex cases involving algorithmically predicted intents [3]. In the latter case, we used Trrack to capture higher level semantics, such as when a user is trying to select specific patterns in a scatterplot.

Logging for User Studies Due to Trrack's unique ability to reproduce entire sessions and store generic metadata, we believe it is especially useful for user studies. Typically, analysis of individual actions in a user study would require manual annotation of every user, a time intensive process. By storing labels, event types, time stamps, and generic metadata on every action by the user, most meta-analysis of a study using Trrack may be captured automatically. This can save time, reduce human error, and allow for more diverse analysis. Trrack provides dedicated export functionality tailored to the needs of post-hoc data analysis. Specifically, it can export details about each action which are not stored directly on the graph and not required for action recovery. Finally, Trrack makes it straightforward to jump into a specific analysis session of a user, allowing analysts to understand their context. [Here] is an example of a link to a specific state for one of our application examples.

Logging vs. Action Recovery There may be some actions developers want to track (e.g., for the purpose of logging), but that are either too frequent or inconsequential to include in the undo/redo chain. A hover action that highlights an item when the mouse rests

[3] https://github.com/visdesignlab/trrack-examples

on it is an example of this. Users would not expect to undo/redo a hover, but when analyzing logs, it can be important to see when hover was used. For these cases, Trrack we allows developers to label actions as ephemeral. By default, the undo/redo functions in the library will skip over ephemeral nodes, and ephemeral nodes are also treated specially in the provenance visualization.

5 TrrackVis

TrrackVis is a separate library complementing Trrack to visualize the provenance graph, as shown in Figure 2. The primary purpose of this library is to allow for easy navigation in the provenance graph, as well as provide an interactive way to create and view annotations and metadata. The graph is shown as an interactive node link tree, for which clicking on any node will change the applications state to the one associated with that node. TrrackVis is designed to be highly customizable, allowing the user to choose to integrate features like tooltips on hover or annotations. Custom icons can be added to nodes to match user defined event types.

To address scalability of the visualization, consecutive nodes can be defined as a group, which can collapse the constituting elements. For example, groups can be used to identify specific sections of the analysis process as related, and organize them as such for additional annotation. The "Insight" nodes in Figure 1 contain nested actions that are associated with a particular insight captured in this application. By default, we also use groups to collapse nodes that are labeled as ephemeral.

6 Implementation, Testing, and Documentation

We developed Trrack over the course of 18 months, constantly refining and adapting the library to a variety of changing needs. The library has roots in an integrated application we designed for visual storytelling [4], and in a prototype library based on the concepts from the aforementioned paper[4].

To encourage adoption of the library, we created a series of basic examples which demonstrate the core features of the library[3]. There is additional documentation and other examples in the repository.

Trrack also includes a comprehensive suite of unit tests to ensure previous versions of the library do not break with future additions.

7 Usage Examples

We have used the Trrack library in multiple projects, spanning prototypes developed for a technical visualization paper, applications used by medical professionals[5], and in user studies. Figure 1 shows some of the examples. Here we provide details on two of them.

We used Trrack to capture data on participants in a crowd-sourced study [9] for evaluating multivariate network visualization techniques. The study used two levels of provenance, a study-level provenance to track the progression of the study and a task-level provenance to track the interactions in a particular task. A combination of these two provenance graphs allowed us to collect data about interaction patterns in both conditions while participants completed the tasks. To explore the logged data we developed a custom visualization, which also allowed us to jump into individual analysis sessions. The study stimulus is available at `https://vdl.sci.utah.edu/mvnv-study/`.

We also used Trrack in our prototype system designed for capturing analysis intents when brushing in scatterplots [3]. We made heavy use of Trracks's ability to store metadata to help calculate intents and make retrieving past intents simple. We also used TrrackVis extensively to capture annotations and intents, and customized the view using icons, groups, and custom stylings for annotations. The screenshot for TrrackVis shown in Figure 1 is taken from this project. Trrack was also used as our means to capture data when we ran a crowd-sourced evaluation of this project. A demo is available at `http://vdl.sci.utah.edu/predicting-intent/`.

[4] `https://github.com/VisualStorytelling/provenance-core/`

[5] `https://github.com/visdesignlab/workforce-frontend` `https://github.com/visdesignlab/bloodvis`

8 Limitations

While Trrack is well suited to track interaction data, it is not designed to store the provenance of datasets. For example, interactively running a normalizing procedure would create a new dataset. This dataset could either be stored directly, as an attribute of a node, or could be written to a separate file, and that file could be referenced as the output. Both solutions have disadvantages: the former mixes actions and data, and creates a large provenace graph; the latter makes the association between data and an application state brittle. We plan on investigating the integration between data and action provenance in the future.

Sharing state through URLs may become a problem if the state is large and hence exceeds URL size limits. However, we have successfully tested this approach with over 150 keys describing a state. We believe most applications will be able to store a state much smaller than this, as long as data is stored separately.

9 Future Work

In the future, we plan to strengthen the collaborative aspects of our method. While we currently enable asynchronous collaboration, we do not track contributions by separate users. We also would like to allow synchronous collaboration, so multiple users may view and interact with the same visualization session, similar to the functionality available in Google Docs.

As an immediate next step, we want to add story-editing and story-viewing capabilities to our library, similar to the approach demonstrated by Gratzl et al. [4].

With our hybrid strategy for storing states and diffs, the primary question to answer is how frequently states should be stored. We hope to do more investigation into ideal times to store states. We also hope to add functionality that optimizes the size of the graph when it is being exported by iterating over the graph and re-storing nodes as appropriate.

10 Conclusion

In this paper we introduced Trrack, a library to track provenance in web-based visualizations. Our goal is to provide an option to easily track provenance in existing or future visualization systems. The companion library TrrackVis allows for visualization of and interaction with the provenance graph. Tracking provenance with Trrack allows action recovery, collaboration, reproducibility, annotation of analysis steps, post-hoc meta analysis of the interaction sequences, and provides persistent storage with Firebase, or other custom storage solutions. Trrack is the first web-based library to enable this functionality.

Both libraries are open source, and published to npm. We provide extensive usage examples and documentation and hope that our library will contribute to increased provenance tracking in more visualization tools. Also, while Trrack was designed primarily with visualization tools in mind, it can be equally used with general purpose web applications.

Acknowledgments

We want to thank Sai Varun Addanki for his help with the implementation, as well as Samuel Gratzl, Marc Streit, Nils Gehlenborg, and Holger Stitz for making the precursor library available. We also want to thank past and present members of the VDL team who integrated our library in their projects and provided valuable feedback. We gratefully acknowledge funding by the National Science Foundation (IIS 1751238).

REFERENCES

[1] W. Aigner, S. Hoffmann, and A. Rind. EvalBench: A Software Library for Visualization Evaluation. *Computer Graphics Forum*, 32(3pt1):41–50, 2013. doi: 10.1111/cgf.12091

[2] L. Bavoil, S. P. Callahan, C. Scheidegger, H. T. Vo, P. Crossno, C. T. Silva, and J. Freire. VisTrails: Enabling Interactive Multiple-View Visualizations. In *Proceedings of the IEEE Conference on Visualization (VIS '05)*, pp. 135–142. IEEE, 2005. doi: 10.1109/VISUAL.2005.1532788

[3] K. Gadhave, J. Görtler, O. Deussen, M. Meyer, J. Phillips, and A. Lex. Capturing user intent when brushing in scatterplots. Jan. 2020. doi: 10.31219/osf.io/mq2rk

[4] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit. From Visual Exploration to Storytelling and Back Again. *Computer Graphics Forum*, 35(3):491–500, 2016. doi: 10.1111/cgf.12925

[5] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '08)*, 14(6):1189–1196, 2008. doi: 10.1109/TVCG.2008.137

[6] N. Kadivar, V. Chen, D. Dunsmuir, E. Lee, C. Qian, J. Dill, C. Shaw, and R. Woodbury. Capturing and supporting the analysis process. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pp. 131–138, Oct. 2009. doi: 10.1109/VAST.2009.5333020

[7] M. Kreuseler, T. Nocke, and H. Schumann. A History Mechanism for Visual Data Mining. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '04)*, pp. 49–56. IEEE, 2004. doi: 10.1109/INFVIS.2004.2

[8] Z. Liu, J. Thompson, A. Wilson, M. Dontcheva, J. Delorey, S. Grigg, B. Kerr, and J. Stasko. Data Illustrator: Augmenting Vector Design Tools with Lazy Data Binding for Expressive Visualization Authoring. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–13. Association for Computing Machinery, Montreal QC, Canada, Apr. 2018. doi: 10.1145/3173574.3173697

[9] C. Nobre, D. Wootton, L. Harrison, and A. Lex. Evaluating Multivariate Network Visualization Techniques Using a Validated Design and Crowdsourcing Approach. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–12. Association for Computing Machinery, Honolulu, HI, USA, Apr. 2020. doi: 10.1145/3313831.3376381

[10] C. North, R. Chang, A. Endert, W. Dou, R. May, B. Pike, and G. Fink. Analytic Provenance: Process+Interaction+Insight. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pp. 33–36. ACM, New York, NY, USA, 2011. doi: 10.1145/1979742.1979570

[11] M. Okoe and R. Jianu. GraphUnit: Evaluating Interactive Graph Visualizations Using Crowdsourcing. *Computer Graphics Forum*, 34(3):451–460, 2015. doi: 10.1111/cgf.12657

[12] E. Ragan, A. Endert, J. Sanyal, and J. Chen. Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Transactions on Visualization and Computer Graphics (VAST '15)*, 22(1):31–40, 2016. doi: 10.1109/TVCG.2015.2467551

[13] A. Satyanarayan and J. Heer. Lyra: An Interactive Visualization Design Environment. *Computer Graphics Forum*, 33(3):351–360, 2014. doi: 10.1111/cgf.12391

[14] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the IEEE Symposium on Visual Languages (VL '96)*, pp. 336–343, 1996. doi: 10.1109/VL.1996.545307

[15] Y. B. Shrinivasan and J. J. van Wijk. Supporting the analytical reasoning process in information visualization. In *Proc. CHI 2008*, pp. 1237–1246. ACM, 2008. doi: 10.1145/1357054.1357247

[16] M. Streit, H.-J. Schulz, A. Lex, D. Schmalstieg, and H. Schumann. Model-Driven Design for the Visual Analysis of Heterogeneous Data. *IEEE Transactions on Visualization and Computer Graphics*, 18(6):998–1010, 2012. doi: 10.1109/TVCG.2011.108

[17] S. van den Elzen and J. J. van Wijk. Small Multiples, Large Singles: A New Approach for Visual Data Exploration. *Computer Graphics Forum (EuroVis '13)*, 32(3pt2):191–200, 2013. doi: 10.1111/cgf.12106

[18] K. Xu, A. Ottley, C. Walchshofer, M. Streit, R. Chang, and J. Wenskovitch. Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum*, 2020. doi: 10.1111/cgf.14035