Indy Survey Tool: A Framework to Unearth Correlations in Survey Data

Tarik Crnovrsanin * Sara Di Bartolomeo Connor Wilson Cody Dunne

Northeastern University

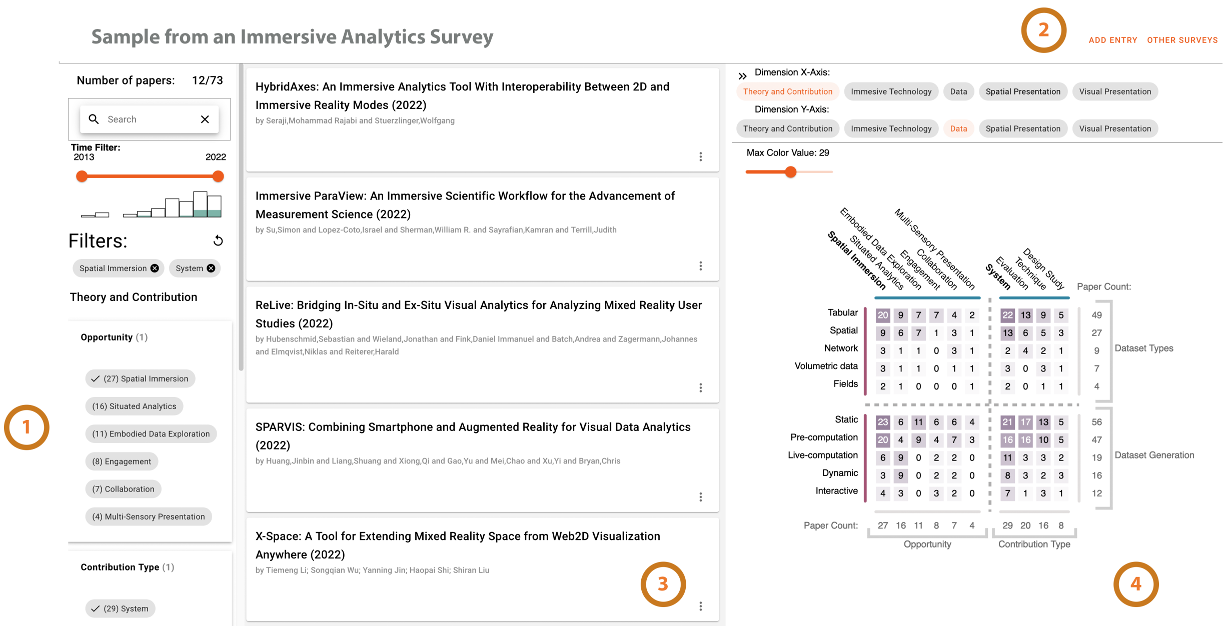

Figure 1: An example of how the Indy Survey Tool we present was used in recent survey on Immersive Analytics [16]. The left panel (1) lets users filter using a search bar, timeline, and topic selector. The top bar (2) provides information about the survey and how to add new entries. The center (3) shows a short summary of each included paper. The collapsible visualization panel (4) on the right shows a correlation matrix for two selected dimensions. Interacting with the left and right panels filters the papers displayed in the center. Upon selection of a paper, a detail view pops up with all of its information (not shown).

Abstract

Survey companion websites allow users to explore collected survey information more deeply, as well as update or add entries for papers. These sites can help information stay relevant past the original release date of the survey paper. However, creating and maintaining a website can be laborious and difficult, especially when authors might not be experienced with programming. We introduce Indy Survey Tool to help authors develop companion websites for survey papers across diverse fields of study. The tool's core aim is to identify correlations between categorizations of papers. To accomplish this, the tool offers multiple combined filters and correlation matrix visualizations that enable users to explore the data from diverse perspectives. The tool's visualizations, list of papers, and filters are harmoniously integrated and highly responsive, providing users with feedback based on their selections. Identifying correlations in survey papers is a pivotal aspect of research, as it can enable the recognition of common combinations of categorizations within the papers—as well as highlight any omissions. The versatility of Indy Survey Tool enables researchers to delve into the correlations between categorizations in survey data, an essential aspect of research that can reveal gaps in the literature and highlight promising areas for future exploration. A preprint and supplemental material for the paper can be found at osf.io/tdhqn.

*{crovrsanin.t, dibartolomeo.s, wilson.conn, c.dunne}@northeastern.edu

Keywords: Software tool, systematic review, literature review, survey, co-occurrence visualization.

Index Terms: Human-centered computing—Visualization—Visualization systems and tools—Visualization toolkits; General and reference—Document types—Surveys and overviews

1 Introduction

A survey paper is a literature review that provides a comprehensive overview of a particular research area. It typically covers a broad range of studies, summarizing their key findings, methodologies, and contributions to the field. The goal of a survey paper is to provide readers with a comprehensive understanding of the current state of research in a particular area and identify gaps in the literature and areas for future research.

Surveys collect and classify vast amounts of data and have extensive metadata associated with the collected papers. To ensure the knowledge contained within survey papers remains easily accessible to other people interested in exploring the field, a large number of surveys on visualization topics offer companion websites. They allow users to peruse the collection of papers, often aided by a number of interactive functions that allow for filtering, linking, and possibly a few visualizations. These companion websites range from simple lists of papers [2, 22] to more complicated presentations, such as Friendly et al.'s [9] site that displays the temporal aspect of the data by placing element on a large timeline in the center of the screen.

Building a companion website is laborious and prohibitive as it requires extensive programming knowledge to build, host, and

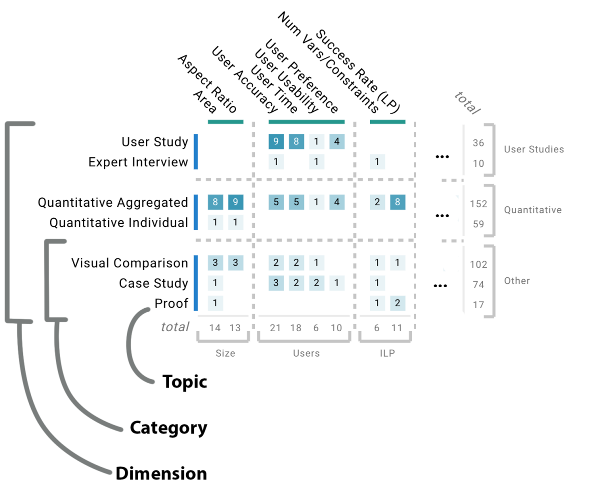

Figure 2: In Indy Survey Tool, every paper can be assigned multiple topics, which can be further organized and nested—dimensions contain categories; categories in turn contain topics. The image presents the correlation matrix in a graph drawing survey [7] between the evaluation method used (left dimension) and what metrics are reported (top dimension).

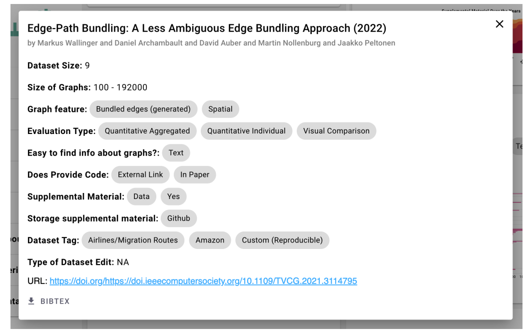

Figure 3: Detail View of the metadata associated with a selected paper in the graph drawing survey [7]. The survey authors decide which information to present for each paper and the presentation order.

maintain. If researchers want to incorporate visualization as well, this requires another set of skills. These requirements limit access to particular fields and sets of expertise. Creating a companion website is further complicated by the volume of data and information contained in a survey. Many survey papers split the research they present into several categories to have a more focused discussion about each, as well as spot similarities and differences between them. Because of this, many companion websites focus on presenting, filtering, and organizing information into categories, tags, and topics.

In this paper, we introduce Indy Survey Tool, a framework for making companion websites for surveys with little coding effort. Our main goal is to enable researchers from varying fields and coding experience to build their survey companion websites with as little effort as possible. Our second aim is to enable the users of the survey and the researchers themselves to have access to automatically-generated visualizations that compare classifications to find correlations and gaps in surveyed literature.

Indy Survey Tool came about from conducting our own systematization of knowledge tasks and survey research. We needed an improved system to organize such categories and, more importantly, we needed to see correlations and gaps within them. Seeing correlations between classifications (e.g. how many times a topic is found in a curated collection of papers associated with another topic) can aid researchers in spotting patterns and common combinations which might indicate juxtapositions of topics that work really well together, or that have been favored in previous research. In contrast, being able to spot gaps and overlooked combinations can help identify possible future research or mixes of topics that do not work well together. In the two case studies we conducted—a survey on immersive analytics visualizations [16] (Section 5.1) and another on the computational evaluation of graph layout algorithms [7] (Section 5.2)—seeing such correlations was of fundamental importance.

Specifically, we contribute:

1. Indy Survey Tool, an open-source automatic web-based framework to create, explore and distribute, surveys online. The tool focuses on using co-occurrence visualization to show correlations between topics and gaps in existing research.
2. A demonstration of the utility and versatility of the tool via two surveys in different subjects that the tool has been used for.

The tool can be found at `github.com/VisDunneRight/Indy-Survey-Tool` and archived at `osf.io/tdhqn`.

2 Related Work and Background

As a central point in surveys is categorizing different papers, many companion websites offer filtering functions based on such categories. The filters can either be simply different navigable pages [3, 8, 20], or, in case of more complex data being presented, offer combined filtering options that allow for multiple simultaneous selections, so that the user can explore intersections between different categories [15, 21]. An example of such a system that offers insight into intersections of categories is H. -J. Schulz's Treevis [17], which is also used by Kehrer et al. for a different subject [11]. Schöttler et al. [18] also implement a similar multi-filtering system.

The importance of combining multiple filters is underlined by the prevalence of the feature in other similar systems. Expanding on the previously explored multi-filtering, Kucher and Kerren [13] compliment the filters with a timeline where the occurrences are the year of the included work—which can be used to further filter by time. The same system is reprised by Kerren et al. in BioVis [12]—where they also include a spatial placement of the papers meant to aid in recognizing clusters of topics—and by Aigner et al. for their TimeViz system [1], proving the versatility of the tool over different subjects. A similar combination of mix-and-match filters and timeline filtering can be found in Fabian Beck's SurVis system [5] which had been used in a multitude of contexts, including dynamic graph visualization [4], sparklines visualization [6], and more [10, 14, 19].

Ultimately, filtering functions are widespread in these types of survey tools due to (a) the fundamental practice of using categories to perform the systematization of knowledge tasks required for surveys—often organized in hierarchies—and (b) the need to explore the intersections between these categories to find papers with specific features. However, our requirement analysis for our use cases showed that the functions could be further improved—for instance, by offering a holistic overview through a correlation matrix of the most common occurrences and the most overlooked ones. Indy Survey Tool is focused not only on offering such insights but also on leveraging those to improve navigation of the collection. Additionally, one of the most important objectives of Indy Survey Tool is to be able to be highly versatile, enabling other researchers to easily use it without needing to write custom code. Among the companion websites mentioned above, the only one designed to be used by others is SurVis [5]. While they provide tagging capabilities and a timeline bar chart, Indy Survey Tool also provides a correlation matrix.

3 Design Requirements

Our target users for the system are researchers conducting a survey that either do not want to go through the hassle of building a website from scratch or lack the technical experience to create one. As this needed to be a general tool for any kind of survey, we worked in close contact with researchers developing surveys on different subjects to identify a set of requirements the tool needed to have.

DR 1: Plug and Play — The tool ought to be user-friendly and easy to set up and deploy, ensuring accessibility for users with varying levels of programming experience. It should not require monetary investment so as to be cost-effective for a broader audience. Likewise, being open-source, maximizes customizability and maintainability, while promoting continued support and community contributions.

DR 2: Focus on correlations of dimensions, categories, topics — We place significant emphasis on identifying and analyzing the correlations between various features in the data. Thus we should present data on individual categorizations and consider how the categorizations relate. A tool for exploring these relationships can provide a more comprehensive understanding of the data and uncover valuable insights that may not be otherwise apparent. Furthermore, by recognizing the interconnections between different features, the tool can help identify potential causal relationships and guide more informed decision-making. Therefore, analyzing correlations between different features is an essential aspect of the survey tool and a crucial step in extracting meaningful insights from data.

DR 3: Nested categorization — As the classification systems used in surveys are often extensive and complex, the categories used to classify the papers can be organized hierarchically. For instance, the category of "temporal data" can include sub-categories "continuous temporal data" and "discrete temporal data". The tool should thus support the capability to maintain a navigable hierarchy.

DR 4: Mix-and-match filters — Extensive filtering would allow users to navigate the categories, ideally with multi-category selection so users can see the intersection of the selections. As with nested categorization, extensive filtering functions enable users to see and navigate correlations and categories.

DR 5: Strong links between all components of the system — As a companion site can consist of several components, it is important that changes in one be reflected in all the other associated components. For instance, selecting a topic of interest should select the appropriate papers to display and update the visualizations accordingly.

DR 6: Easy data updates — Democratizing the process of updating data is vital for ensuring the sustainability and accuracy of any system that relies on it. Democratizing data updates spreads the responsibility of maintaining the system across a wider group of individuals, reducing the workload for maintainers and helping keep the system relevant and up-to-date. Easy updates are key to making the site sustainable in the long term, without requiring extensive manual intervention every time new information needs to be added.

DR 7: Customization — The tool should be able to adapt to different needs while maintaining ease of use. A layer of customization should be easily accessible—in the case of Indy Survey Tool, it can be done by editing a JSON file—but further customization should also be available in case niche features are required (see Section 5.2).

4 Methodology and Implementation

In the following section, we discuss the overview and the components that make up the tool, implementation details, and usage.

4.1 Overview

Indy Survey Tool is comprised of a top bar, filter panel, paper view, visualization, and a detail view (see Figure 1). When discussing these components, we will refer to those that created the survey website as *researchers* and those that use the site to access information as *users*.

1 The **Filter Panel** allows users to filter the relevant papers through search, topic selection, and time range. Starting from the top, we display the current number of papers selected versus the total number. The search bar looks through the titles of the paper for matching words. Next is the time filter, where the users can select the time range they are interested in. The timeline visualization indicates the number of papers for each year. It updates when a time range is selected and when papers are filtered through any other means. Lastly, the *Filters section* allows users to filter papers by topic. Information is displayed in accordion style for more straightforward navigation and compactness. A specific topic also displays the number of papers that contain that topic. The tool displays selected topics at the top of the filter list for easy reference and removal.

The researcher provides the structure and hierarchy for the filters in the configuration file. The researcher can decide which dimensions or categories to provide for filtering through the *filterby* variable. These variables allow for a finer level of control, such as deciding which topics to display and their order and the color used for the visualization. The full extent of configuration and use of the tool can be found on GitHub: `github.com/VisDunneRight/Indy-Survey-Tool`.

2 The **Top Menu** allows users to add entries or locate other surveys of interest. Information about the survey, description, icon, and authors can be found in the top left. A JSON file contains a list of other surveys to reference and is modifiable by the researcher. When the user clicks on *Add Entry*, a popup window appears with a description from the researcher and fields to populate a new entry. The tool automatically generates fields from the *filterby* configuration. For fields with multi-select, auto-compilation and adding a new topic are provided. At the bottom of the window is the reset feature, copy to clipboard, and the option to either open a GitHub issue or send an email, depending on which configuration the researcher decided on. The *Open Issue* button opens a new window with a link provided by the researcher to a new issue in GitHub, where the user can paste information copied to the clipboard and provide any other information requested. Using issues opens up the possibility for any maintainer of GitHub to add a new entry to the website. We also provide an email option as an alternative means of updating the site.

3 The **Paper View**, in the center, displays a list of papers where each paper is either a text block or image based on researcher configuration. The researcher configures the information displayed either in a text block or as a hover-over on the image. The researcher decides which categories to display and in what order. Clicking on the three dots opens up a menu with *update entry* and *copy BibTex* options. This functions similarly to *Add Entry* with all the information per-filled, allowing a user to make edits quickly. Clicking on the block or image opens up a **Detail View**, providing more information on the paper, including BibTex and URL.

4 The **Visualization** panel, on the right, shows a correlation metric. The matrix and dimension are generated automatically from the *filterby* configuration and the paper data. Each row and column represents a specific topic, while the cross-section contains the number of papers with both topics. The end of each row and column indicates the number of papers that include that topic and which category the topics belong to. The user can select which two dimensions of interest they wish to explore. The tool also allows the user to filter the visualization through a minimum number of papers cutoff. Selecting a row or column name will filter papers with that topic, while selecting the cross section will select both topics. The visualization panel can support other types of visualization, as seen in Figure 4 and discussed in Section 5.2, with more planned.

4.2 Using the Tool

The tool leverages a single JSON file as its primary source of information. Researchers can easily organize their paper metadata in various formats, such as a Notion table, Excel sheet, or Airtable, to create their own survey companion website. No extensive programming

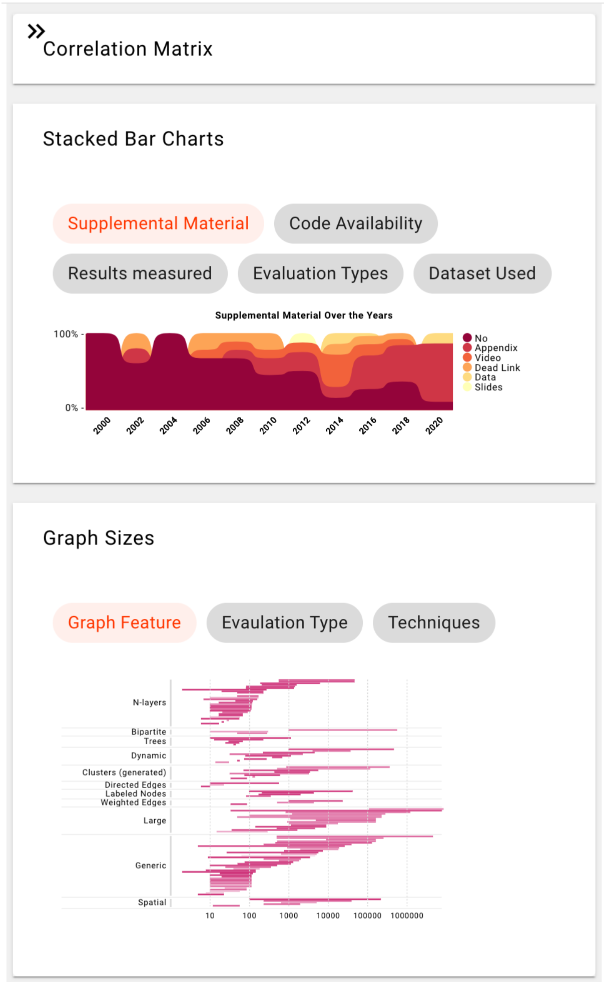

Figure 4: When dealing with metadata that is not categorical (e.g. integers or dates) more visualizations can be integrated in the website. The two figures above (from the survey website discussed in Section 5.2) show the evolution of a feature over time (top figure—in this case, the ratio of types of supplemental material through the years) and an integer value split by different categories (bottom figure—in this case, graph sizes in number of nodes by graph feature).

skills are required, as we provide comprehensive instructions on GitHub that guide users on how to utilize the tool effectively.

Acquiring the metadata is a task undertaken by researchers as part of their survey process, allowing them to use their preferred method of data collection. Once researchers have gathered the necessary information, they can fork our template survey, which comes with pre-set functionality but no content. To simplify deployment, researchers can configure the repository for GitHub Pages.

After exporting their table to a CSV file, researchers can employ our provided Python script to help convert CSV into the appropriate JSON file format. This JSON file can be further customized based on their specific needs. Researchers also have the flexibility to set up different panels and determine how information is presented to users through the JSON file. A more detailed explanation with figures is provided on the repository.

Once the website is live, users of the survey can contribute additional entries to the collection through the *add entry* or *update entry* buttons (refer to Section 4.1) making maintaining the website easier on the researchers.

5 Case Studies and Discussion

Here we describe two case studies that used Indy Survey Tool for creating survey companion websites. The descriptions of the case studies make several references to the design requirements, linked through "DR" references.

5.1 Design space of immersive analytics

Our tool was employed in crafting a survey on the usage of visualization in Mixed Reality (XR) by Saffo et al. [16], which encompasses both Virtual and Augmented Reality. The companion site for the survey is shown in Figure 1 and is available at `iadesign.space` The survey aimed to identify common combinations of features used in Immersive Analytics for various purposes and use cases (DR 2). For example, small-situated visualizations are often combined with augmented reality applications. As XR is a relatively new space, it is crucial to identify the absence of such correlations, as they might be unfeasible or illogical. The survey focused on numerous topics and categories (DR 3). It was essential to highlight frequently-occurring correlations and gaps, which were organized into semantically meaningful groups using a nested categorization system. Our tool incorporated mix-and-match filters (DR 4) and strong linkage (DR 5) between all components, facilitating navigation through the intricate categorization system. Furthermore, the ability for users to quickly and easily update and add to the system was an indispensable feature, given the rapid evolution of the XR field (DR 6).

5.2 Computational evaluation of graph layout algorithms

Our second case study is a survey by Di Bartolomeo et al. [7], which encompassed 161 papers that examined the evaluation of graph layout algorithms. The survey classified the findings in layout algorithm research across multiple dimensions, such as the dataset utilized for algorithm evaluation, graph features, and reported metrics. Identifying correlations between these categories was particularly crucial for this project (DR 2). For example, the authors aimed to uncover any connections between graph features and metrics reported in computational evaluations. In the survey, both overlooked and rare combinations of features, which may suggest implausible or infeasible pairings, remained relevant discussion points, marking this as an area that begs further research.

Figure 2 comes from this survey and illustrates the correlations between the evaluation type used and the quality metrics reported in the results of various graph layout algorithms. The full figure can be found in the supplemental material and the live interactive page at `visdunneright.github.io/gd-comp-eval/`.

The survey's metadata included numerical and temporal aspects. To display this information, the authors integrated supplementary visualizations into the tool, as depicted in Figure 4. Serving as a foundation for crafting new visualizations as needed (DR 7), the tool demonstrated its versatility and applicability for this type of research. The researchers used Notion to collect metadata, then plugged the table into the tool to report on their findings (DR 1).

6 Conclusion and Future Work

We introduce Indy Survey Tool, a framework for constructing companion websites for surveys that emphasize the promotion of survey research by allowing researchers to easily create interactive and navigable websites that organize their papers cohesively. Our tool specifically addresses the relationships and gaps in paper categories. We achieve this through correlation matrix visualizations displayed throughout the tool, facilitating extensive interconnections that enhance accessibility to the research content. In the future, our goal is to continue to update the tool by working with individuals constructing new surveys and extending feature sets to meet their demands. Specifically, we hope to support more visualization options like those shown in the second case study.

Acknowledgments

This work was supported in part by the U.S. National Science Foundation (NSF) under award number CAREER IIS-2145382 and by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at Northeastern University administered by Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence (ODNI).

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, second edition ed., 2023.

[2] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. The state-of-the-art of set visualization. *Computer Graphics Forum*, 35(1):234–260, 2016. Companion website: https://www.cvast.tuwien.ac.at/SetViz. doi: 10.1111/cgf.12722

[3] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A descriptive framework for temporal data visualizations based on generalized space-time cubes. *Computer Graphics Forum*, 36(6):36–61, 2017. doi: 10.1111/cgf.12804

[4] F. Beck, M. Burch, S. Diehl, and D. Weiskopf. The state of the art in visualizing dynamic graphs. In R. Borgo, R. Maciejewski, and I. Viola, eds., *EuroVis - STARs*. The Eurographics Association, 2014. doi: 10.2312/eurovisstar.20141174

[5] F. Beck, S. Koch, and D. Weiskopf. Visual analysis and dissemination of scientific literature collections with survis. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):180–189, 2016. doi: 10.1109/TVCG.2015.2467757

[6] F. Beck and D. Weiskopf. Word-sized graphics for scientific texts. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1576–1587, 2017. doi: 10.1109/TVCG.2017.2674958

[7] S. Di Bartolomeo, T. Crnovrsanin, D. Saffo, and C. Dunne. Designing computational evaluations for graph layout algorithms: the state of the art, Mar 2023. Companion website: visdunneright.github.io/gd-comp-eval/. doi: 10.31219/osf.io/ms27r

[8] P. Dragicevic and Y. Jansen. List of physical visualizations. www.dataphys.org/list, 2012. Last accessed 2023-04-21.

[9] M. Friendly, D. Denis, and H. Truman. Milestones in the history of thematic cartography, statistical graphics, and data visualization, 01 2001. Companion website: https://www.datavis.ca/milestones/.

[10] K. E. Isaacs, A. Giménez, I. Jusufi, T. Gamblin, A. Bhatele, M. Schulz, B. Hamann, and P.-T. Bremer. State of the art of performance visualization. In *Eurographics Conference on Visualization*, 2014.

[11] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, 2013. doi: 10.1109/TVCG.2012.110

[12] A. Kerren, K. Kucher, Y.-F. Li, and F. Schreiber. Biovis explorer: A visual guide for biological data visualization techniques. *PLOS ONE*, 12(11):1–14, 11 2017. doi: 10.1371/journal.pone.0187341

[13] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 117–121, 2015. doi: 10.1109/PACIFICVIS.2015.7156366

[14] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci. Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1249–1268, 2017. doi: 10.1109/TVCG.2016.2640960

[15] C. Nobre, M. Streit, M. Meyer, and A. Lex. The state of the art in visualizing multivariate networks. *Computer Graphics Forum (EuroVis)*, 38:807–832, 2019. doi: 10.1111/cgf.13728

[16] D. Saffo, S. Di Bartolomeo, T. Crnovrsanin, L. South, J. Raynor, C. Yildirim, and C. Dunne. Unraveling the design space of immersive analytics: A systematic review. *IEEE Transactions on Visualization and Computer Graphics*, 2023. VIS '23. Preprint & supplemental material: https://osf.io/2e9x4. Companion website: https://iadesign.space/.

[17] H.-J. Schulz. Treevis.net: A tree visualization reference. *IEEE Computer Graphics and Applications*, 31(6):11–15, 2011. doi: 10.1109/MCG.2011.103

[18] S. Schöttler, Y. Yang, H. Pfister, and B. Bach. Visualizing and interacting with geospatial networks: A survey and design space. *Computer Graphics Forum*, 40(6):5–33, 2021. doi: 10.1111/cgf.14198

[19] F. Sperrle, M. El-Assady, G. Guo, R. Borgo, D. H. Chau, A. Endert, and D. Keim. A survey of human-centered evaluations in human-centered machine learning. *Computer Graphics Forum*, 2021.

[20] 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

[21] 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

[22] L. Zhang, A. Stoffel, M. Behrisch, S. Mittelstadt, T. Schreck, R. Pompl, S. Weber, H. Last, and D. Keim. Visual analytics for the big data era — a comparative review of state-of-the-art commercial systems. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 173–182, 2012. Updated companion website: https://commercialtools.dbvis.de/home. doi: 10.1109/VAST.2012.6400554

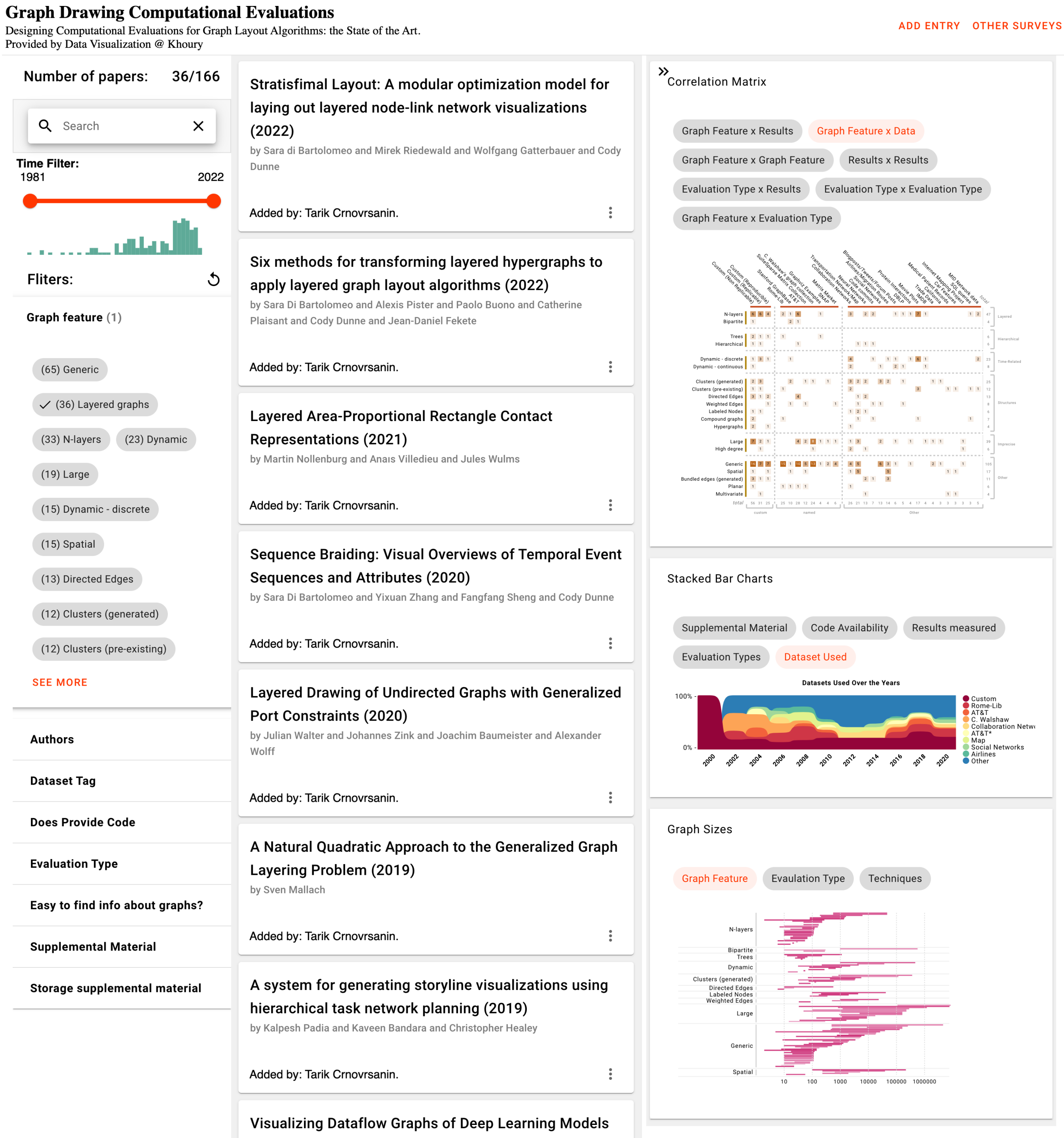

Figure 5: A screenshot of the entire interface used in the computational evaluation survey [7].