Department of Computer Science
Series of Publications A
Report A-2022-1

Entity-Based Insight Discovery in Visual Data Exploration

Chen He

Doctoral dissertation, to be presented for public examination with the permission of the Faculty of Science of the University of Helsinki in Hall PIII, Porthania, on January 27th, 2022 at 16 o'clock.

University of Helsinki
Finland

Supervisor
Giulio Jacucci, University of Helsinki, Finland

Pre-examiners
Alex Endert, The Georgia Institute of Technology, United States
Paolo Buono, University of Bari Aldo Moro, Italy

Opponent
T.J. Jankun-Kelly, Mississippi State University, United States

Custos
Giulio Jacucci, University of Helsinki, Finland

Contact information

Department of Computer Science
P.O. Box 68 (Pietari Kalmin katu 5)
FI-00014 University of Helsinki
Finland

Email address: info@cs.helsinki.fi
URL: http://cs.helsinki.fi/
Telephone: +358 2941 911

ISSN 1238-8645
ISBN 978-951-51-7837-4 (paperback)
ISBN 978-951-51-7838-1 (PDF)
Helsinki 2022
Unigrafia

Entity-Based Insight Discovery in Visual Data Exploration

Chen He

Department of Computer Science
P.O. Box 68, FI-00014 University of Helsinki, Finland
chen.he@helsinki.fi
https://researchportal.helsinki.fi/en/persons/chen-he

PhD Thesis, Series of Publications A, Report A-2022-1
Helsinki, January 2022, 63 + 60 pages
ISSN 1238-8645
ISBN 978-951-51-7837-4 (paperback)
ISBN 978-951-51-7838-1 (PDF)

Abstract

Visual data exploration (VDE) allows the human to get insight into the data via interaction with visual depictions of that data. Despite the state-of-the-art visualization design models and evaluation methods proposed to support VDE, the community still lacks an understanding of interaction design in visualization and how users extract insight through interacting with the data. This research aims to address these two challenges.

For interaction design, a literature review reveals that a lack of actionability hinders the application of existing visualization design methods. To address this challenge, this research proposes an approach abstracting data to entities and designing entity-based interactions to achieve the higher-level interaction goals. Three case studies, i.e., interacting with information facets to support fluid exploratory search, interacting with drug-target relations for insight discovery and sharing, and supporting insight externalization through references to visualization components, demonstrate the applicability of this approach in practice. The three cases detail how the approach could address the design requirements derived from related work to fulfill the various task goals following the nested model of visualization design and the resulting designs' transferability to other datasets. Reflecting on the case studies, we provide design guidelines to help improve the entity-based interaction design.

To understand the insight generation process of VDE, we present two user studies asking users to explore a visualization tool and externalize insights by inputting notes. We logged user interactions and characterized collected insights for correlation and prediction analysis. Correlation analysis of the first study showed that exploration actions tended to relate to unexpected insights; the drill-down interaction pattern could lead to insights with higher domain values. Besides asking users to input notes as insights, the second study enabled users to refer to relevant entities (visualization components and prior notes) to assist their narration. Results showed evidence that entity references provided better predictions than interactions on insight characteristics (category, overview versus detail, and using prior knowledge). We discuss study limitations and results' implications on knowledge-assisted visualization, such as supporting insight recommendations.

As future work, structuring user notes by entities could make the insight machine-readable to stimulate mixed-initiative exploration, e.g., machines help to collect evidence to validate the insight. Creating a platform that supports uncertainty-aware insight and insight provenance across tools could facilitate practical analysis which usually involves multiple analysis tools.

Computing Reviews (2012) Categories and Subject Descriptors:

Human-centered computing → Interaction design → Interaction design process and methods

Human-centered computing → Visualization → Empirical studies in visualization

General Terms:
information visualization, interaction, visualization exploration, insight

Additional Key Words and Phrases:
interaction design, entity, insight-based evaluation

Acknowledgements

First and foremost, my sincere gratitude goes to my supervisor, Professor Giulio Jacucci. Having admired your work as a master's student, I was honored to join your group and work on your various exciting projects under your guidance. I am grateful for your flexibility and openness in research and your ability to maintain the perfect balance between trusting my free exploration and providing critical course corrections, which greatly improved the efficiency of this research.

I extend my gratitude to my mentor, Dr. Luana Micallef. Your passion and knowledge are always an inspiration to me both energetically and practically in moving this work forward.

This research would also not be possible without the generous financial support from the Helsinki Doctoral Education Network in Information and Communications Technology, the Doctoral Program in Computer Science, University of Helsinki, and the Strategic Research Council at the Academy of Finland through the DataLit project.

I would also like to thank friends and researchers from the department, Tung Vuong, Barış Serim, Imtiaj Ahmed, Mikko Kytö, Yuxing Chen, Pengfei Xu, Chao Zhang, Qingsong Guo, Gongsheng Yuan, Sara Ramezanian, Krista Longi, and Aditya Jitta, for the company on this journey and encouraging conversations. My thanks extend to all the staff at the department for creating a comfortable working environment and accessible resources, especially our research coordinator, Pirjo Moen, for helping with my doctoral study process; IT specialists, Pekka Niklander and Sami Niemimäki, for solving my computer and server issues; HR coordinator, Roosa Sillanpää, for helping with my contract; and translator, Marina Kurtén, for proofreading this dissertation.

My gratitude also goes to my pre-examiners, Professor Alex Endert and Professor Paolo Buono, for your acknowledgment of the work and valuable suggestions to help improve the dissertation. My gratitude extends to my opponent, Professor T.J. Jankun-Kelly, for your time and energy to inspire a critical rethinking of this research and Professor Indrė Žliobaitė to be the faculty representative for the examination process.

Special thanks go to Professor Wei Chen for hosting my research visit during my challenging time. Your compassion and inclusiveness set an example for me to learn from. I would also like to thank Wei Zhang, Yingchaojie Feng, Xumeng Wang, Jiehui Zhou, Rusheng Pan, and other group members for the insightful talks and discussions.

On my research path, I am grateful to my master's supervisors, Professor Yoshifumi Kitamura and Professor Kazuki Takashima, who inspired me to take on the path of research. My gratitude extends to my former supervisor, Professor Katrien Verbert, who helped me embark on this journey, and other friends from KU Leuven, Yucheng Jin, Robin De Croon, Francisco Gutiérrez, and Sven Charleer, for the exciting discussions. Many other passionate researchers from the research community showed me to follow one's passion and curiosity and think critically, which I appreciate very much.

I am deeply grateful to my family for their unconditional love and support. I am filled with admiration for your courage and resilience in facing life's challenges and your love and devotion toward each other under critical conditions.

Last but not least, my heartfelt appreciation is toward S. N. Goenka, Eckhart Tolle, Bashar, Marina Jacobi, the Arcturian Council, Ryok, Anthony William, and many others for your dedication to humanity and for telling me the truth about this reality which completely changed my perception for the better. We develop technologies to preserve humanity and expand our consciousness. This is what the dissertation is dedicated to.

Helsinki, November 2021
Chen He

Contents

List of Publications

This dissertation is based on the following original publications, which are referred to throughout the dissertation as Articles I–IV. Authors' contributions to the publications are detailed as follows. The publications are reprinted at the end of this dissertation.

Article I: Chen He, Luana Micallef, Barış Serim, Tung Vuong, Tuukka Ruotsalo, and Giulio Jacucci. Interactive visual facets to support fluid exploratory search. In *the International Symposium on Visual Information Communication and Interaction*. ACM, 2021

Contribution: Giulio Jacucci conceived the idea behind this work. Barış Serim and Tung Vuong designed and implemented the system and conducted the user study. The author created the use cases and drafted the article. All of the authors participated in the revisions.

Article II: Chen He, Luana Micallef, Ziaurrehman Tanoli, Samuel Kaski, Tero Aittokallio, and Giulio Jacucci. MediSyn: Uncertainty-aware visualization of multiple biomedical datasets to support drug treatment selection. *BMC Bioinformatics*, 18(S-10):393:1–393:12, 2017

Contribution: Samuel Kaski, Tero Aittokallio, and Giulio Jacucci conceived the idea behind this work. Luana Micallef and the author designed the visualization, MediSyn, and the user study. Ziaurrehman Tanoli provided one of the datasets used in MediSyn. The author implemented MediSyn, conducted the user study, and drafted the article. Tero Aittokallio proposed the representative use case of MediSyn. Giulio Jacucci and Luana Micallef provided critical revisions to the writing.

Article III: Chen He, Luana Micallef, Liye He, Gopal Peddinti, Tero Aittokallio, and Giulio Jacucci. Characterizing the quality of insight by interactions:

A case study. *IEEE Transactions on Visualization and Computer Graphics*, 27(8):3410–3424, 2021

Contribution: The author conceived the idea behind this work, developed the prototype, conducted the user study, and drafted the article. Liye He and Gopal Peddinti evaluated the collected insights. Giulio Jacucci provided critical revisions to the writing. Luana Micallef, Giulio Jacucci, and Tero Anttikallio supervised this work.

Article IV: Chen He, Tung Vuong, and Giulio Jacucci. Characterizing visualization insights through entity-based interaction: An exploratory study. Submitted

Contribution: Giulio Jacucci and the author conceived the idea behind this work. The author developed the prototype, conducted the user study, and drafted the article. Tung Vuong evaluated the collected insights. All of the authors participated in the revisions of the article.

Chapter 1

Introduction

With the advent of the world wide web, smartphones, and all kinds of sensors, our everyday lives yield a considerable amount of data including online surfing records, activity tracking data, etc. According to the sources, as of 2020, humans created 2.5 Quintillion bytes of data daily (= 2.5 billion Gigabytes per day).[1] Without extracting useful information and knowledge from the raw data, the value of the collected data is not fulfilled [23].

To help derive knowledge from data, various visualization and automation techniques are developed, as can be seen in the rapid growth of artificial intelligence (AI) and data science. As an interface for people to access data, visualization plays a critical role in enhancing humans' analytical capability and helping people make sense of data [24]. Numerous visualization design books and tools published in recent years [122] speak to this point.

Conventionally, even in the visualization community, researchers have a linear view of the relations between humans and automation from low automation/high human operation to full automation/zero human operation. However, Shneiderman [136] rewrote the concept by proposing a two-dimensional chart relating human control and computer automation with the ultimate goal of creating AI with both high human control and high automation, i.e., human-centered AI, which is leading the evolution of AI and human-computer interaction (HCI) to a bright future. As an example, self-driving cars accommodate high automation; meanwhile, high human control is also needed to ensure that passengers can travel along preferred routes, such as the fastest route or the route with better scenery, and reach the desired destinations. From the human-centered view, visualiza-

[1] https://techjury.net/blog/how-much-data-is-created-every-day

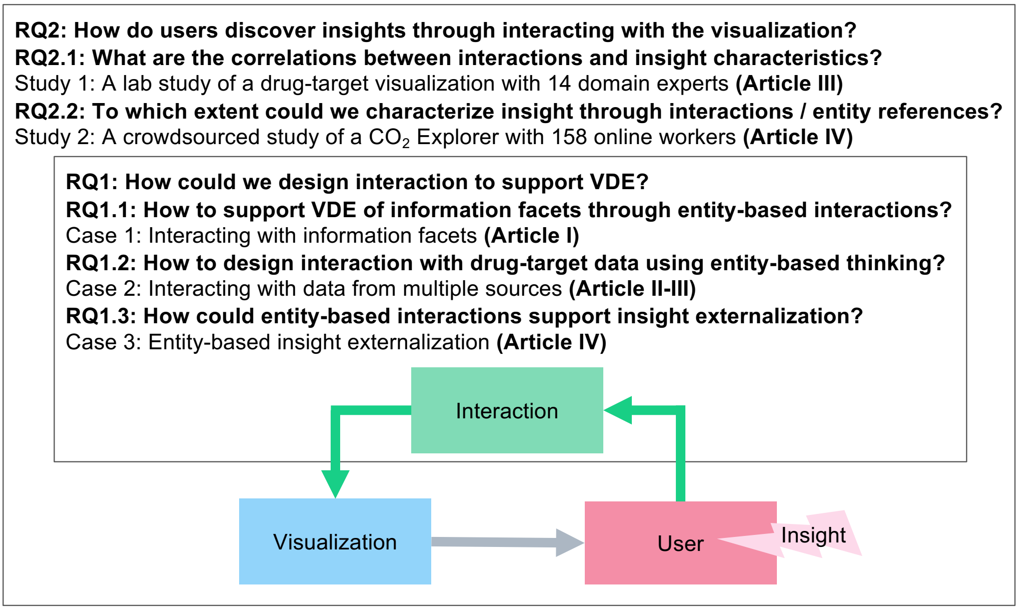

Figure 1.1: Overview of the RQs and published articles of this dissertation concerning the visualization process.

tion is and will always be indispensable to support users in comprehending and manipulating the automation and data to acquire insights.

1.1 Research Questions (RQs) and Motivations

As supported by many researchers (e.g., [29, 163]), "the purpose of visualization is insight, not pictures [19]." To support visualization insight, we can not overlook the aspect of visual data exploration (VDE). It is through interacting with the visual representation, the human-computer discourse, that insights are derived.

However, the visualization community lacks an understanding of VDE and how users generate insights through interacting with the visualization [34]. This research aims to support interaction in visualization by proposing an interaction design approach and understand the user insight generation process through empirical studies. Specifically, this research explores the following two RQs, which have been broken down into five sub-RQs and studied through four research publications (Articles I–IV). Figure 1.1 provides an overview of the RQs and

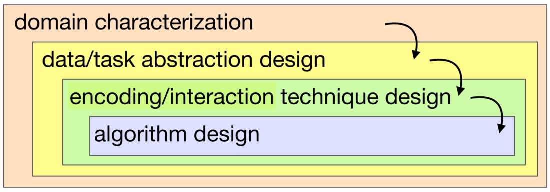

Figure 1.2: The nested model for visualization design by Munzner [102].

articles in relation to a typical visualization process (simplified based on Chen et al. [23]).

RQ1: *How could we design interaction to support VDE?* VDE denotes the process of getting insight into the data via interaction with visual depictions of that data [78]. To design VDE, Munzner [102, 103] proposed a nested model for visualization design and validation, which has been widely adopted and successfully guided the design of many visualization tools (e.g., [53, 70]). The model consists of four nested layers (Figure 1.2). To design a visualization, one characterizes the domain problems, abstracts the domain-specific data and problems into generic descriptions, designs visual encodings and interaction techniques, and implements the algorithms to realize the design. The process is always iterative and involves rapid prototyping. Moreover, Meyer et al. [99] extended the nested model by proposing a nested blocks and guidelines model where blocks capture design decisions in each layer and guidelines relate decisions within or between layers. A research paper could contribute new blocks and/or new guidelines.

From data and task abstraction to visual encoding and interaction technique design, principles are well established on how to assign visual encodings to various types of data to facilitate human perception (e.g., [67, 94]). However, research on designing interaction from abstracted data and problems is less developed [34, 146]. Although the well-known information-seeking mantra [135], "overview first, zoom and filter, then details on demand," works well in many cases (e.g., [36]), when the data get huge, and the domain situation and task characterization become complex, interactions need to be carefully devised to accomplish the task goals [131].

Researchers suggested adopting methods from HCI and social science to design interaction, which emphasizes the in-situ design collaboration between users and designers (e.g., [55, 72]). However, there is a lack of an anchor point to ground the design thinking and communication. To address this limitation, this research proposes to abstract data to entities for interaction design.

Entities are widely used in text analysis to represent any real-world objects and concepts. People naturally perceive things as entities and create mental models by relating entities to understand external information [9, 111]. Thus entity-based design thinking tends to be user-centric and could provide actionability in interaction design by using the aforementioned models and frameworks. Following the nested visualization design model (Figure 1.2), this research demonstrates the applicability of the entity-based interaction design approach in practical cases.

RQ2: *How do users discover insights through interacting with the visualization?* With discovering insight as a primary purpose of VDE, understanding the user insight generation process becomes critical in visualization design. The traditional task-based evaluation limits in understanding users' open-ended exploration beyond task time and error, such as evaluating how well a visualization supports insight [110]. To assess the ability of VDE in supporting insight, Saraiya et al. [129] proposed an insight-based evaluation, which measures the characteristics of insights users derive from exploring the visualization tools, such as breadth versus depth and domain values.

However, without looking into the insight generation process, results can be limited in informing the visualization design. Mayr et al. [97] compared the evaluations on task performance (time and error), insight characteristics, and problem-solving strategies. Evaluating problem-solving strategies involved analyzing think-aloud data, interaction logs, and viewing behaviors. They found that compared with the other two methods, analyzing problem-solving strategies shed more light on how to improve the visualization.

Insight results from user interaction with the visualization tools. To explore how VDE supports insight, this research takes a holistic approach, investigating the user insight generation process by linking interaction types/patterns to insight characteristics, and provides implications on designing knowledge-assisted visualization.

1.2 Research Methods

RQ1 has three sub-RQs which were explored through three case studies (Figure 1.1). First, to answer RQ1, this research proposes to abstract data to entities and design entity-based interaction to support user task goals. Three design case studies, i.e., interacting with information facets, interacting with data from multiple sources, and entity-based insight externalization, demonstrate the applicability of this approach. Each case presents design requirements (DRs) derived from prior work to fulfill the task goals, how the entity-based interaction design could address the DRs, and the resulting designs' transferability to other types of data in responding to the statement by researchers that the goal of visualization design is "transferability, not reproducibility" [60, 134].

RQ2 was studied in two phases through two sub-RQs (Figure 1.1). To answer RQ2.1, we conducted a lab study with a domain-specific visualization depicting drug-target relations by asking domain experts to freely explore the visualization and generate insights by writing notes. Besides logging user interactions and insights, we extracted interaction patterns, characterized insights, and analyzed the correlations between interaction types/patterns and insight characteristics.

Building on the promising results from RQ2.1, we raised RQ2.2. We studied a generic visualization—CO_2 Explorer—through crowdsourcing to answer RQ2.2. Besides note taking, the CO_2 Explorer enables users to cite relevant entities (visualization components and prior notes) to assist their narration. We then used interactions and entity references to predict insight characteristics through advanced machine learning models; to explain prediction performance, we calculated feature importance on individual cases and performed a similar correlation analysis as we did in the first study.

1.3 Contribution

The contribution of this dissertation is three-fold:

- We provide an in-depth analysis of VDE through reviewing related literature, looking into the holistic process from interaction to visualization insight, and raise open research questions that require research attention from multiple perspectives (Chapter 2).
- We propose an entity-based interaction design approach to provide an anchor point and actionability in interaction design thinking, and demonstrate the applicability of this approach through three case studies (Chapter 3).
- To understand the holistic insight generation process, we present results from two user studies that linked interactions and insight characteristics and provide implications on knowledge-assisted visualization (Chapter 4).

1.4 Outline

The remainder of this dissertation is organized as follows: Chapter 2 reviews related work on VDE, interaction design and analysis, and visualization insight, and discusses open research challenges and how this research contributes to the community. Chapter 3 exemplifies the entity-based interaction design through existing work and three case studies and concludes with design guidelines to answer RQ1. Chapter 4 presents two user studies to explore RQ2 and provides design implications to support visualization insight. Chapter 5 concludes this dissertation by answering the two RQs and discussing the limitations and future directions of this research.

Chapter 2

Background

Researchers studied *exploratory data analysis* in practice through interviewing professional data analysts (e.g. [2, 6, 75, 157]). They provided several common suggestions to help improve visualization tools, which include integrating tools to support the existing analysis ecosystem [2, 6, 157], such as combining visual interactions with command-line tools [2, 6], integrating data from multiple sources [2, 75, 157], using automation to save time for repetitive tasks [2, 157], recording and exporting analysis provenance [2, 6, 75, 157], and supporting insight [2, 6, 75, 157], such as insight automation [2] and insight export [6, 157]. This chapter analyzes these aspects in research by first characterizing VDE in the scope of data analysis (Section 2.1) and then reviewing related publications on interaction design and analysis as well as on visualization insight (Sections 2.2-2.4). Section 2.5 summarizes open research questions and positions this research in the relevant fields.

2.1 Characterizing VDE

Tukey [147] introduced the concept of *exploratory data analysis* back in 1977 to differentiate exploratory analysis from confirmatory analysis. Exploratory analysis supports hypothesis formulation, whereas confirmatory analysis helps to test the hypothesis. Battle and Heer [7] defined *exploratory visual analysis*[1] as a subset of

[1]VDE [78] has been studied under various terminologies including *visualization exploration* [73] and *exploratory visual analysis* [7]. From a user-centered perspective, all of these terms convey a concept of getting insight into the data via interaction with visual depictions of that data. Thus we treat these terms as equal when reviewing related literature.

exploratory data analysis to emphasize the use of visualization in assisting users to explore the data as opposed to *automatic data analysis*. As discussed at the beginning of Chapter 1, we need both types of analysis, combining the strengths of humans and machines to create a synergistic way forward. This is where the term *visual analytics* [79] comes from.

Keim [78] considered *the involvement of humans* critical in the data exploration process with their creativity, knowledge, etc. in getting insight into the data, though a different view exists that most data processes can be automated in the "big data" era [24]. For instance, researchers proposed techniques to automatically generate insights from data/visualizations (e.g., [31, 140, 152]). However, automation can only generate insights about the data while losing the context of the domain [77, 128]. Karer et al. [77] criticized the data-centric view on analysis and argued to involve various levels of contexts to acquire domain-related insight. Sacha et al. [128] asserted that the process of collecting versatile evidence to generate knowledge could not be automated.

To make the role of users concrete, through literature review, Battle and Heer [7] characterized exploration in visual analysis as often involving *browsing* and *search*, and alternating between *open-ended* and *focused* exploration and between *top-down* and *bottom-up* exploration. Focused (top-down) exploration contrasts the popular view that exploration is opportunistic and does not have a clear goal [2]. When VDE is guided by a focused goal, the findings are not necessarily relevant to the goal but can open new analysis directions [128]. Sacha et al. [128] identified three inter-linked human cognitive processes during visual analysis, namely exploration / verification / knowledge generation loops. Linked to the verification loop, "the exploration loop is steered to reveal findings that verify or falsify the hypothesis." Keim [78] and Battle and Heer [7] also identified verifying hypotheses as a common task in VDE. These views blurred the boundary between exploratory and confirmatory analysis with visualization.

However, both types of analysis are necessary to get insight into the data [128, 148]. Insights generated from VDE should be considered as hypotheses that need to be validated [77, 128]. Alspaugh et al. [2] interviewed data analysts about the reasons behind not practicing exploratory analysis in their daily work and received answers on avoiding spurious findings or multiple comparisons resulting from exploration. With the multiple comparisons problem, Zgraggen et al. [164] found over 60% of the findings from visual exploration were false. Thus exploratory findings of the data are preliminary, "requiring confirmation with an independent data source" [83, 147, 164].

Existing visualizations widely support exploration but provide limited support for confirmatory analysis [30, 83]. To tackle this issue, studies tried to nudge users toward confirmatory analysis by eliciting user expectations about the data before users view the data [30,83], which present an opportunity for users to reach a balanced analysis on the exploratory-confirmatory spectrum to make diverse, sound discoveries [83, 164].

Activities of *browsing* and *search* are another characteristic of VDE [7]. Chen et al. [23] argued that a visualization process is a search process, but not like the traditional query search interfaces. With VDE, users usually search in a high-dimensional space for insights. Another view on VDE as a hypothesis generation process [78] focuses more on the exploration results, whereas this view emphasizes the process itself. As search is common in VDE, Green et al. [54] proposed search by example/pattern to support intuitive and fluid exploration and suggested, "search by example should be part of any visual analytics interface involving analysis or reasoning tasks for large amounts of information."

2.2 Interaction Design for VDE

To guide the design of effective and efficient visualization, researchers proposed various design methods including the well-known nested model for visualization design and evaluation [102] (Figure 1.2) and the nine-stage framework for design study [134]. Based on the nested model, McKenna et al. [98] proposed a design activity framework to provide actionable guidance on the design process, and Meyer et al. [99] proposed a nested blocks and guidelines model to help capture design decisions and the rationale behind. Chen and Ebert [22] proposed using entity graphs to capture design problems, causes, and solutions and expose causal relations of the design workflow to support the recording, sharing, and reproduction of design knowledge. To support collaborative design among visualization designers, developers, and domain experts, approaches are proposed based on practices [8, 59, 151].

However, despite the efforts on building visualization design disciplines, the community still lacks an understanding of interaction or interaction design in visualization [34]. The reason may be that the aforementioned mainstream design approaches are task-oriented, rather than user-oriented [34, 86]. Dimara and Perin [34] attempted to address this issue by characterizing interaction in visualization through literature review. They suggested that interaction design for visualization needs to 1) consider "broader spectra of user profiling" and 2) enrich interactions to flexibly support diverse data-related intents.

To practice user profiling, visualization researchers borrowed methodologies from HCI (e.g., user-centered design [72]) and social science (e.g., action research [60]). Green et al. [55] discussed applying *participatory design* and *activity theory* for visualization design emphasizing an in-situ design collaboration between users and designers through an iterative design process. In this way, designers could gain a holistic understanding of users and their situations including their domain knowledge, problems, individual and group environment. From the human cognition perspective, Liu and Stasko [90] identified a user-centered design process as a convergence of mental models of users and designers, with external visualization as an integral part of the human cognitive system and interaction as the focus to understand reasoning using visualization.

The other concern raised by Dimara and Perinn [34], interaction flexibility within the visualization, intends to address the gulf of execution—the gap between user intention and the interaction possibilities of the tool [109]. To assist user intent, we need to know what the possible interactions with visualization tools are. Interactions are usually characterized at multiple levels of granularity [34]. A popular characterization is tasks, sub-tasks, actions, and events by Gotz and Zhou [50]. They identified the importance of actions, as actions, indicating distinctive user intents, are generic (different from tasks and sub-tasks) and semantically meaningful (different from events, such as mouse clicks). Action taxonomies based on user intent could support a wide range of tasks (e.g., [34, 48, 50, 90]).

ElTayeby and Dou [41] suggested an extra level between sub-tasks and actions as patterns composed of multiple actions to support analysis reuse. Sedig and Parsons [113, 131] proposed an interaction design space with 32 action patterns and 10 adjustable properties of visual representation to support complex cognitive activities, e.g., analytical reasoning and knowledge discovery. They defined *inter-activity* as the quality of interaction and provided interactivity characterization to support the design and evaluation of interaction in human-centered visualization tools [132, 133]. Case studies demonstrated how the concepts could be applied in practice [5, 114].

To improve the effectiveness and intuitiveness of interaction, Pike et al. [116] raised several interaction challenges in visual analysis including ubiquitous, embodied interaction, capturing higher-level thought processes, supporting collaboration, and others. Similarly, Lee et al. [86] suggested interaction in visualization to go beyond mouse and keyboard to support freedom of expression and collaboration.

Though frameworks and methodologies are proposed / borrowed from other research fields to stimulate interaction design in visualization, a lack of actionability hinders the application. To provide an anchor point for the interaction design thinking using the aforementioned frameworks and methods, this research proposes to abstract data to entities for interaction design, building upon the existing visualization design model [102]. Chapter 3 demonstrates that the entity-based interaction could flexibly support various task goals through existing work and case studies, and the resulting design could be transferred to other types of data through the abstraction.

2.3 Analysis of User Interactions

Interaction is critical in complex cognitive tasks, including analytical reasoning, decision making, etc. Researchers provided comprehensive surveys on the analysis of interaction data (e.g., [41, 57, 159]). This section reviews interaction analysis based on the four major analysis goals: provenance analysis, visualization evaluation / behavioral analysis, reasoning & sensemaking, and prediction & recommendation, and discusses the contribution of this research in relation to existing work.

Provenance analysis. Provenance records the analysis history, such as interaction logs and analytical thoughts, to support analysis reuse, result dissemination, collaboration, etc. Through interviewing data analysts, Madanagopal et al. [95] revealed that provenance data are critical to support practical analysis tasks. However, existing visualization tools provide poor support for provenance [95]. Ragan et al. [119] and Xu et al. [159] provided comprehensive surveys on various types of provenance data and their purposes and analysis techniques, whereas Hall et al. [58] specifically reviewed the work of insight provenance and provided guidelines on supporting such provenance.

Visualizing provenance features automatically capturing interaction and visualization states, which are then displayed as timelines (e.g., [92, 166]) or trees (e.g., [12, 16, 17, 52, 107, 138, 142]), as well as manual creation of analysis trails (e.g., [40, 74, 96]). For instance, KnowledgePearls visualizes automatically captured interaction and visualization states in trees and support flexible search techniques, such as weighing multiple search terms and query by example, to retrieve analysis states [142]; ExPlates enables users to spatialize visualization workflows by creating data or visualization plates and connecting the plates in terms of the data flow. To capture users' thought processes during VDE requires externalization, which we will elaborate on in Section 2.4.

Building a community standard to transfer provenance among diverse analysis tools is beneficial as analysts seldom complete an analysis within a single tool [46, 116, 159]. As a step forward, Cutler et al. [32] built a web-based library—Trrack—to be integrated into visualization systems for provenance tracking and management. The library would be more powerful in history management if it could incorporate the conceptual model of interaction history proposed by Nancel and Cockburn [104].

Researchers also suggested supporting hierarchical provenance data from low-level interactions to high-level tasks [13, 119, 158] to guide/prompt users through the analysis tasks [13]. Although higher-level tasks and user intents are difficult to capture automatically, research exists aiming to categorize actions using topic modeling [26] and segment interaction logs into higher-level activities [161].

Visualization evaluation / behavioral analysis. To support fraudulent behavior detection, Nguyen et al. [105, 106] proposed two visual analytics approaches enabling analysts to explore hierarchical user profiles including overview-, group-, and individual-level user activities. To support analysis of user strategies and the cognitive processes of using visualizations, Blascheck et al. [10, 11] proposed two visual analytics systems integrating interaction, eye tracking, and think-aloud data. Automatic pattern detection and search-by-pattern interactions are supported for analysis. Liu et al. [91] visualized web clickstream data in multiple levels of granularity (patterns and sequences) for analysis. Additionally, researchers also proposed other novel interaction metrics (e.g., [47]) and analysis methods (e.g., [56, 121]) to support visualization evaluation.

Reasoning & Sensemaking. Research shows that interaction logs could help users recover their own as well as others' reasoning processes [38, 89]. SensePath depicts web browsing actions in a timeline coupled with video recordings for analysts to understand the user sensemaking process [108]. Dou et al. [39] proposed a framework of capturing user interaction and thought processes to construct one's reasoning process and introduced three criteria to disambiguate the meaning behind interactions. Several systems support the user reasoning and sensemaking process by visualizing interaction histories and enabling users to create a knowledge graph to externalize their discoveries [107, 138, 139]. Moreover, Pohl et al. [117] analyzed theories from psychology and HCI to help explain the exploratory reasoning process of visual analytics systems.

Prediction & Recommendation. Interaction could be used to predict the next actions [100, 112, 144], personality traits [15], and user tasks [49]. *Semantic interaction* implies that systems could learn from user interactions and make adaptations [42]. ForceSPIRE infers a set of relevant entities based on user interaction and co-creates with users a spacialization of a collection of documents for sensemaking [43, 44]. Through modeling user interaction, visualization could recommend relevant resources including external articles [167], appropriate visualizations [49, 137], and next actions [33, 144] to assist VDE. For instance, Zhou et al. [167] proposed a model to contextualize visualization by surfacing relevant articles based on interaction history; Dabek and Caban [33] proposed an approach building a set of rules from user interactions to guide new users along the analytic process.

As the primary goal of VDE is to support insight, this research explores the relations between interactions and insights, relating to the analysis purposes of visualization evaluation and prediction & recommendation. The work most relevant to ours is the one by Guo et al. [56], which correlated types of interactions to three types of insights, i.e., facts, hypotheses, and generalizations. In difference from Guo et al. [56], we provide an in-depth analysis of insights by quantifying one insight from multiple perspectives following Saraiya et al. [129], such as its domain value and breadth versus depth, and correlating the characteristics with interaction types/patterns. Further, we use interactions to predict insight characteristics and provide implications on knowledge-assisted visualization based on user interaction.

2.4 Visualization Insight

The cognitive science community defined an insight as an “Aha!” moment, a sudden breakthrough that evokes a unique neural activity pattern [14, 21], whereas the visualization community assigned a broader meaning to insight indicating an advance in knowledge [21]. The two types of insight support one another in VDE [21]. To be concrete, Karer et al. [77] defined visualization insight as “a step forward in the interpretation and analysis in the form of a change of the user’s knowledge or understanding”, which could be further distinguished as insight into the visualization/data/domain.

Other than providing an explicit definition, several researchers attempted to characterize insight. Through interviewing professional visualization users, Law et al. [85] characterized insight as actionable, collaboratively-refined, unexpected, confirmatory, spontaneous, trustworthy, and interconnecting, which are similar to Chang et al. [21] and North's [110] discoveries. Others characterized insight in a bottom-up way (e.g., [56, 129]). Based on the collected think-aloud data from users interacting with visualizations, Saraiya et al. [129] quantified insights by domain value, directed versus unexpected, breadth versus depth, correctness, etc. Our studies adopt this characterization, which measures an insight from multiple perspectives based on practice.

Existing visualization tools provide limited support for insight, such as insight automation and insight export [2, 6, 75, 157]. Through reviewing existing work on insight automation, Law et al. [84] characterized 12 types of insight and four purposes of insight automation. Auto-insight can generate comprehensive discoveries about the data and does not have biases that could result from humans' limited attention and belief systems. However, automated insights are usually simple data facts, such as outliers and trends, not aligning with the concept of insights being deep and complex [84]. Also, insight about the data needs to be interpreted in the problem domain to provide actionability, which is difficult to automate [77, 128]. Click2Annotate enables users to semi-automate insight by selecting templates of common types of insight [27]. In this case, users have more flexibility to involve their domain knowledge in the annotation.

Related work also supports manual creation of insight in two main approaches: 1) providing users with a canvas to externalize insight as node-link diagrams matching their mental models (e.g., [88, 107, 138]), and 2) enabling users to input texts as insight and attach visualizations (e.g., [69, 96, 149, 156]) or data sources (e.g., [155, 165]) as insight provenance, or vice versa, enabling users to embed texts in visualizations as insight/annotations (e.g., [123]). Other interaction modalities have also been explored for insight externalization, such as digital pen and touch [80, 124, 125].

To enhance the manual creation of insight, Pike et al. [116] challenged research on the machine-readable externalization of user thinking process rather than being just narratives so that new mixed-initiative systems are possible, e.g., the machine could help to reason and collect evidence to validate or falsify user insight. *Data-aware annotation* is a simple form of machine-readable insight in which annotations could be applied to different views of the same data [68], such as the features of scented insight browsing and faceted insight retrieval in Click2Annotate. We propose to extract entities from insight narratives / attached visualizations to structure insight and support entity-aware annotations. In this way, visualization could support bi-directional exploration: visualization with scented entities could promote exploration of relevant insight, and extracted entities from insight could be added to the visualization for exploration.

2.5 Summary and Open Questions

VDE complements *automatic data analysis* by incorporating human knowledge for insight discovery. The basic feature of VDE involves browsing and search; visualization should seamlessly and intuitively incorporate search functionalities to support VDE, such as search by example/pattern. During VDE, users also alternate between open-ended and focused exploration and between top-down and bottom-up exploration. Most existing visualizations support open-ended exploration, whereas more support on focused and top-down exploration is required.

Interaction design is a weak spot in visualization research. Researchers borrowed methods from other fields, such as HCI and social science, and proposed user-centered frameworks for visualization design [55, 60, 72]. However, a lack of actionability inhibits their application. Besides, interaction beyond traditional desktop settings, such as multi-modal and multi-user interactions, needs further research.

Nonetheless, interaction plays a critical role in visualization. Interaction reflects the user reasoning/sensemaking process and could support visualization evaluation. Learning from user interaction, systems could make predictions and adaptations to assist VDE, which has been studied under the term *semantic interaction*. As the primary goal of VDE is to discover insight, the analysis of interaction needs to be combined with the resulting insight to provide a holistic understanding of VDE.

Besides VDE, visualization needs to support provenance and insight in practice. Provenance data include user interactions, eye movement, thinking processes, etc. Most studies are confined to the analysis of interaction data. Building a community standard to support the transfer of provenance and insight across platforms could facilitate analysis with various tools. Automatic ways to elicit user thought processes, such as inferring higher-level activities from low-level interaction data, could empower machines to guide users through VDE, which needs further investigation.

Regarding insight, while auto-insight could discover data-related insight without inherent human bias, isolated from domain knowledge, insight loses the context to provide actionability and in-depth knowledge. On the other hand, manual externalization of insight narratives is challenged by recording insight in a machine-readable manner so that machines can support reasoning in a mixed-initiative manner. Moreover, with the multiple comparisons problem, discovered insights need to be further validated through VDE/automation, which is not well studied in related work.

Within one dissertation, it is difficult to address all of the above challenges. This research focuses on 1) proposing an interaction design approach for visualization to provide actionability, an anchor point in interaction design thinking (RQ1), and 2) linking interaction to the resulting insight to understand how users generate insights through interactions and provide implications on knowledge-assisted visualization (RQ2).

Chapter 3

Entity-Based Design for VDE

As discussed in Chapter 2, to provide actionability in interaction design, this chapter introduces the approach abstracting data into entities and devising entity-based interactions. Entities are widely used in text analysis [1] and information retrieval [82] to represent any real-world objects and concepts to facilitate VDE. Tools like Jigsaw [141] and Analyst's Workspace [4] extract named entities from documents and represent the entity and document relations using various visualization techniques to support analysis and annotation. Exploration Wall [81] and the topic-relevance map [115] visualize entities, such as keywords and topics, along with search results to help users comprehend the search space and direct search.

According to the entity-relationship model from the database field, an entity denotes a "thing" that can be distinctively identified, such as a person and an event, whereas a relationship is an association among entities [25]. Therefore, we can use entities to represent information in various domains. For instance, Ojha et al. [111] suggest handling open data through entities to create domain-independent and user-centric visualizations. Their entity-centric representation of open data is domain-independent as they modeled types of entities individually to be used in different domains and is user-centric as people intuitively perceive things as entities and categorize entities by their similarities and differences.

Focusing on the interactivity of entities, Klouche et al. [82] proposed a design template of entity-based information exploration: an entity can yield other relevant entities to support information discovery; entities can be organized to assist sensemaking; entities can be saved and shared to support collaboration. This framework implies the flexibility of entity-based interactions and their applicability to various data types. For instance, PivotPaths visualizes entity relations in layered node-link diagrams and supports pivot actions to trigger the re-organization of the entity layouts for information discovery and sensemaking [37]. Their entity-based interaction can be applied to various datasets, such as movie collections and YouTube videos [37]. Andolina et al. [3] and Bier et al. [9] utilized entities to support collaboration. Individual entities [3] or customized entity views [9] can be shared among collaborators to support group sensemaking.

In the remainder of this chapter, Sections 3.1-3.3 present three case studies elaborating how we apply the entity-based interaction design approach in practical visualization design projects to answer RQ1.1-1.3. Each case presents design requirements (DRs) in order to fulfill the various VDE goals and discusses the transferability of the resulting entity-based interactions to other types of data. As stated by Hayes [60] and Sedlmair et al. [134], the goal of visualization design is "transferability, not reproducibility." Section 3.4 concludes this chapter by answering the RQs and providing guidelines to improve the devised entity-based interactions.

3.1 Case 1: Interacting with Information Facets

Search is an essential activity we perform on a daily basis. Research shows that facets are necessary in search to help users navigate the information space, especially when user needs are not well formulated [126, 154]. Information facets, which are orthogonal sets of categories [65], can be considered classes of entities [18], e.g., the people facet consists of individual people entities. Faceted search provides facets to assist search results browsing from multiple perspectives besides the traditional query search. This case study demonstrates the interaction design of a faceted search interface and the result's transferability to other contexts based on the data abstraction to entities and facets (Article I).

Starting with visualizing emails, we extracted the important factors, such as timestamps, people, and keywords, to represent the information space of a collection of emails. Entities of timestamps represent linear facets, whereas entities of people and keywords denote categorical facets. The two types of facets are coordinated in the visualization with the linear facet displaying the distribution of items and the categorical facets summarizing a set of items (Figure 3.1). The interaction design fulfills the two DRs derived from prior work to address the limitations of existing tools in supporting fluid exploratory search.

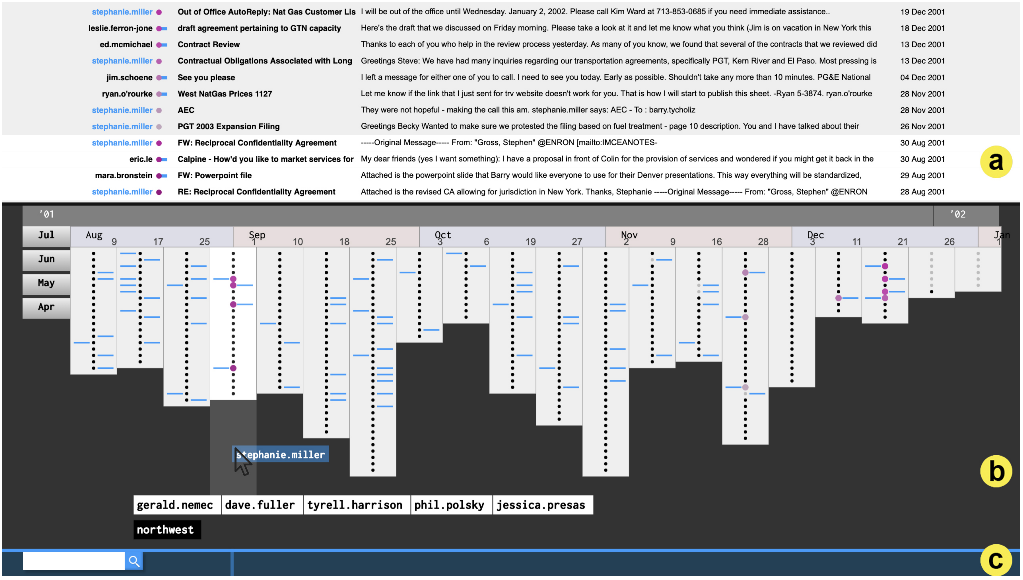

Figure 3.1: The faceted search interface visualizes the selected items (a), a linear facet where each dot represents a data item (b), categorical facets of, e.g., people and keywords (b), and a query field to filter facets and items (c). A categorical entity, "stephanie.miller", is under focus such that the linear facet shows the distribution of relevant items through blue lines. In the case of emails, left-side lines indicate sender relations, and right-side lines denote co-recipient relations. The entity, "stephanie.miller", is dragged on to a linear facet bar (filter-swipe) such that items in the intersection of the two facet values are selected indicated by dark purple dots and a white background color (a) and the categorical facet displays relevant entities to the selected items.

DR1.1: *Provide contextual information for faceted exploration.* Contextual information can avoid users getting lost in the search experience. Visualizing facets per se provides context about the information space. Further, coordinated views are often used to support exploration of facet relations (e.g., [35,160]). To provide a more systematic view on exploration within context, we identified time- and space-related contexts. A time-related context positions the user in the exploration process. We used the color encodings of the item dots to indicate that the items were, are, or have not been selected by the user. A space-related context informs users about the current search space. Facet exploration through coordinated views falls into this category. Similarly, we achieved this through devising the interaction between the linear and the categorical facets. Mousing over the linear facet bars triggers the categorical facets to dynamically summarizing the items in the bars; mousing over the categorical entities shows the distribution of relevant items in the linear facet (Figure 3.1).

DR1.2: *Use facets to support rapid transitions between search criteria.* As user queries are often tentative, user interaction needs to allow easy query transitions with low cognitive load to provide a fluid search experience. Query preview can support tentative queries. However, most tools are limited to preview the number or sample of items related to a facet value (e.g., [65, 130]); more advanced preview techniques could be devised to address this requirement. To support rapid query transitions, the tool features using categorical entities to select items without filtering the item space, i.e., keeping the current search context. One way is to select items by clicking on a categorical entity. The other way is to use a filter-swipe technique by dragging a categorical entity over a linear facet bar; as a result, the items in the intersection of the two facet values will be selected and the categorical facet will show entities relating to those items (Figure 3.1). Figure 3.2 captures the design rationale through the blocks of data/task abstraction and interaction techniques. A video demonstration of the entity-based interactions is available at `https://youtu.be/v0tUAxPjqfg`.

The abstraction of data into facets and entities allows us to transfer the design to other exploration contexts, such as tweets, which also contain linear and categorical facets. To demonstrate the transferability of the design, Article I presents use cases of the design with two other datasets, which are tweets for serendipitous discovery and patient genetic mutation profiles for age-related oncogene co-occurrence recognition (Table 3.1).

3.2 Case 2: Interacting with Data from Multiple Sources

In many real-world situations, such as biology [150] and clinical research [143], relevant data are dispersed in various sources, hindering hypothesis formulation, decision-making, etc. Data integration can make the value of data explode [101] and is identified as necessary for practical data analysis, as mentioned at the beginning of Chapter 2. Visualization is required to integrate data from multiple sources and facilitate analysis (e.g., [10,11,51,87]). For instance, Domino integrates heterogeneous and high-dimensional datasets by creating and linking various data

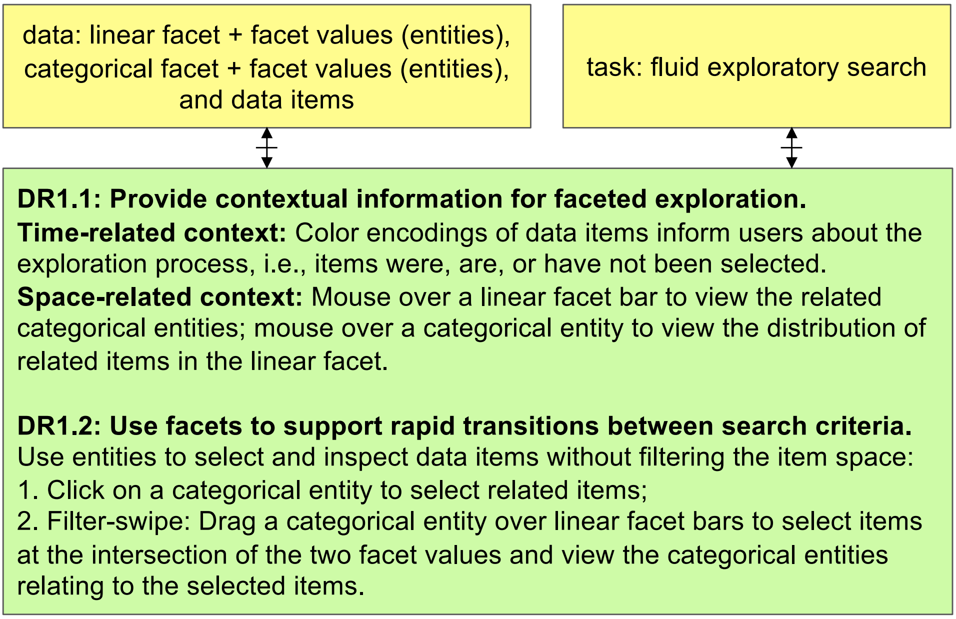

Figure 3.2: The data/task abstraction and interaction technique blocks of the faceted search interface design.

Table 3.1: Transferability: Three use cases of visualizing information facets.

Case	Linear facet	Categorical facet
Email finding	Time	Sender, co-recipient, and keyword
Serendipitous tweet discovery	Time	Username, keyword
Recognition of age-related oncogene co-occurrences	Age	Mutated gene

blocks [51]; StratomeX visualizes datasets in columns and connects columns using ribbons to show relations [87].

In this case study, we devised MediSyn, which integrates drug-target relations from multiple sources. The drug-target relations here mean that various tumor types with certain mutations could be resistant or responsive to certain drugs. The multi-source drug-target data have similar structures and can share the same coordinate space in representation to expose *data uncertainties*. The visualization adopts a matrix-based view to expose *missing data*, which depicts mutations

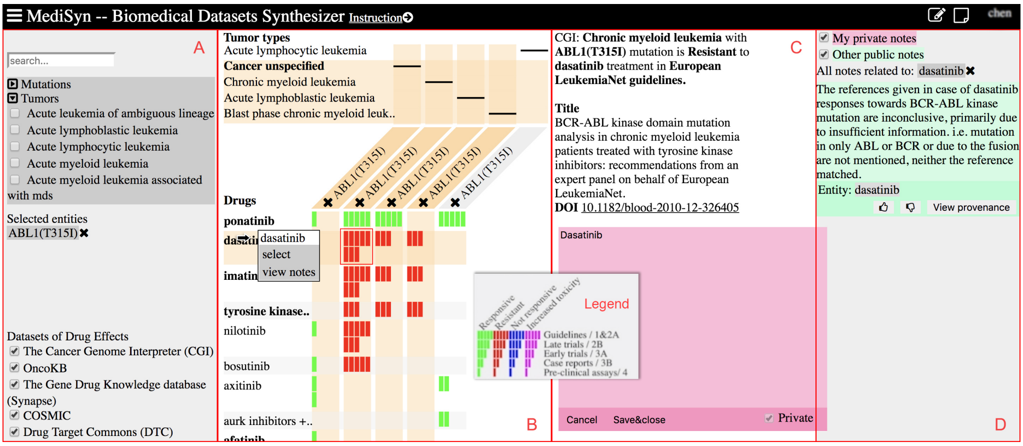

Figure 3.3: The MediSyn interface. Users can select entities of interest from the list (A) and explore relations to other entities in the matrix-based view (B). In the view, columns represent mutations, upper rows are tumor types, and lower rows show drugs. Table cells depict entity relations from various sources in bars where hues indicate drug effects and lengths of the bars denote evidence levels. Users can click on a bar to view its description (C). Entity labels in bold indicate the existence of relevant notes. Through a context menu on hovering, users can choose to explore its entity relations by selecting it and view its relevant notes on the right side (D).

in columns, drugs in lower rows, and tumor types in upper rows (Figure 3.3). Table cells show the drug-target relations from multiple sources to help identify *data consistencies* and display evidence levels of the relations to indicate *data credibility*, such as clinical studies and case reports. The goal of the interaction design is to support biologists to generate and share insight about the data, which are broken down into the two DRs.

DR2.1: *Enable exploration from multiple perspectives to facilitate insight.* The more ways users can explore the data (by changing the forms or perspectives), the more insights they will generate [116]. A similar statement from Sacha et al. [128] is that enabling users to look at data from different perspectives is "the best way to support knowledge generation," which provides "the possibility to collect versatile evidence and increases the level of trust in findings."

DR2.2: *Support the bi-directional exploration of insight and visualization.* Data visualization could promote the exploration of relevant insight; meanwhile, inspired by the insight, users could explore the relevant data view. Data-aware insight mentioned in Section 2.4 in a simple way to address this requirement.

Through an iterative design process, this case demonstrates how we applied the entity-based interaction design to fulfill the two requirements. In the initial design iteration, we focused on designing VDE of mutations without using entity-based design thinking, as the domain expert we collaborated with commented that they were interested in drug activities toward certain mutations in the datasets. Article II presents the design decisions of MediSyn. To explore the data, users can interact with the mutations by selecting mutations of interest, highlighting their relations to drugs, sorting relevant drugs based on clicked mutations, and retrieving the details of a drug-mutation relation. See a video demonstration of the interactions at `https://youtu.be/Bg_YvhBs1sg`.

In the second iteration, we redesigned the interactions by abstracting drugs, mutations, and tumor types to entities. This abstraction enables us to generalize the interaction on mutations to drugs and tumor types so that users can explore the data from multiple perspectives, centering not only on mutations but also on drugs and tumor types (DR2.1). For example, initially, users can click on a mutation to reorganize the rows to view the most relevant drugs; after we generalize the connect action to drugs, users can also explore drug-mutation relations by clicking on a drug to sort columns and view related mutations.

To support the collaboration and communication among biologists, MediSyn allows users to share their insights as notes. We designed an entity-based insight-sharing module, which supports the bi-directional exploration of entities and insights by automatically extracting entities from user notes, such as mutations and drug names (DR2.2). To entice insight exploration, visual cues are provided in the view on entities mentioned in the notes; users can choose to view its relevant notes through a context menu on hovering (Figure 3.3 (B)). Meanwhile, to support entity exploration, MediSyn enables users to select mentioned entities from the notes to explore the entity relations from multiple data sources in the view (Figure 3.3 (D)). To help rationalize insights, MediSyn automatically records user interactions that lead to insights as provenance; it visualizes interaction steps by drawing the resulting views of the interactions linearly when users open the provenance view of an insight.

Figure 3.4 depicts the entity-based interactions to explore drug-target relations. Figure 1 and Section 4.2 of Article III illustrate the resulting MediSyn system and detail the interaction redesign. A video demonstrates the resulting interactions at `https://youtu.be/9NjXvJlqamQ`.

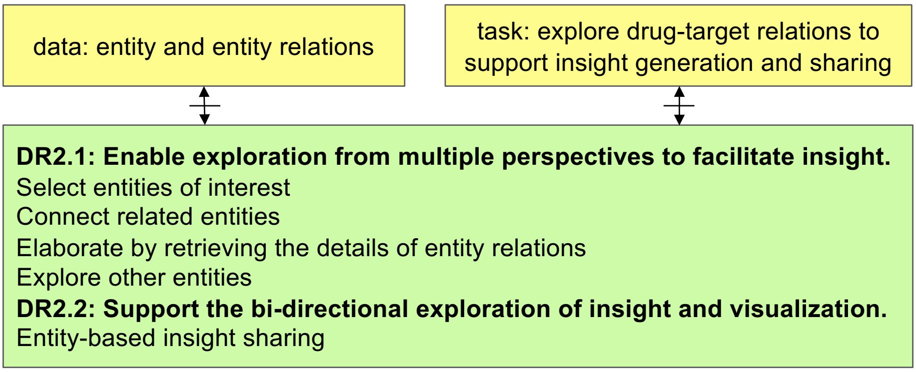

Figure 3.4: The data/task abstraction and interaction technique blocks of MediSyn visualizing multi-source drug-target data.

The resulting visualization and interaction can be transferred to other contexts, such as university rankings by subjects from multiple sources. In this case, the entities of universities, subjects, and countries can replace the entities of mutations, drugs, and tumor types in the visualization, respectively. Table cells depict universities' subject rankings from multiple sources, such as the academic ranking of world universities and the Times higher education world university rankings. Users can select, for instance, a country to explore its universities and subjects, connect relevant entities in the view through highlighting, elaborate on the detailed information of a table cell, explore, e.g., a subject and its relevant entities by selecting it from the view, and share insights on entities of interest by posting notes.

3.3 Case 3: Entity-Based Insight Externalization

With a primary goal of supporting insight, visualization needs to consider insight externalization as an integral part of VDE. Externalizing visualization insight often requires users to link their narrative to the relevant visualization (e.g, [69,149,156]). However, during VDE, an analyst usually works on multiple tasks at the same time in a "chaotic or spontaneous" nature [128], whereas a derived insight could be relevant to part of the visualized data. Allowing users to refer to visualization components, such as a line in the line chart, in their insight as provenance rather than the entire view could make the externalization more relevant and focused.

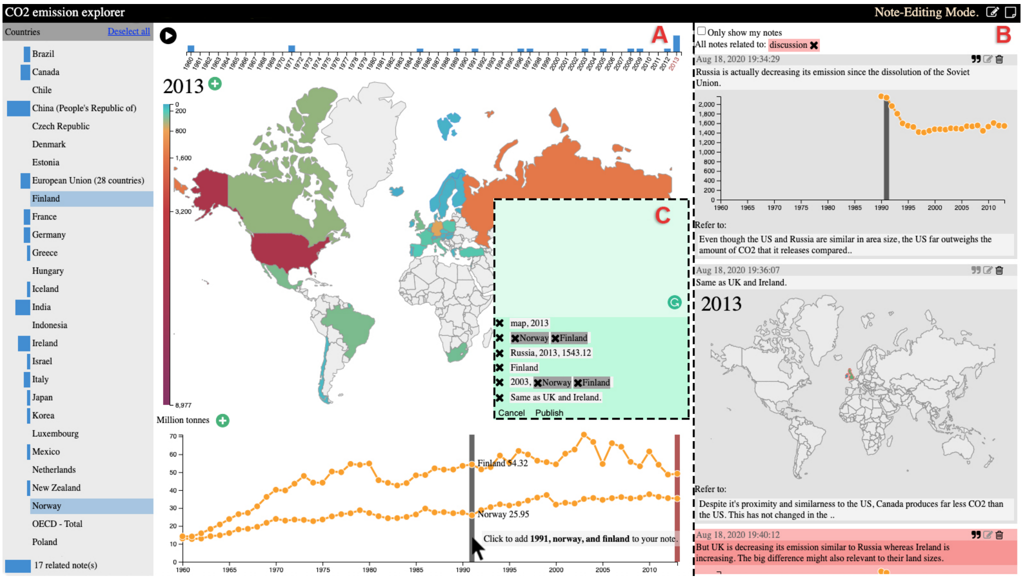

Figure 3.5: The CO_2 Explorer (A) with an insight component (BC). Users can select a year to explore that year's global CO_2 emissions on the map and select countries to explore their CO_2 emissions over the years in the line chart (A). The insight component enables users to compose an insight through inputting notes and referring to six types of entities (C) as well as explore others' insights (B).

With this purpose in mind, this case *supports insight externalization by enabling users to cite relevant visualization components to their narratives (**DR3.1**)*. To achieve this, in contrast to the previous cases in which we considered only nouns as entities, such as emails, keywords, and drugs, this case abstracts visualization components into entities (Article IV). We identified three types of entities in a visualization: individual-level entities denote basic visual elements, such as individual lines in a line chart; a group-level entity depicts a group of visual elements from one or more dimensions, such as a group of bars in a bar chart and lines in a line chart; and a chart-level entity represents various charts, such as a bar chart and a line chart. If a user discovers a trend regarding a line in the line chart, the user can cite the specific line in the note and describe the finding.

We implemented this concept in an existing CO_2 Explorer which shows the global CO_2 emission values of a selected year in a choropleth map and the selected countries' CO_2 emissions over the years in a line chart (Figure 3.5). Chart-level entities are choropleth maps of various years and line charts, group-level entities

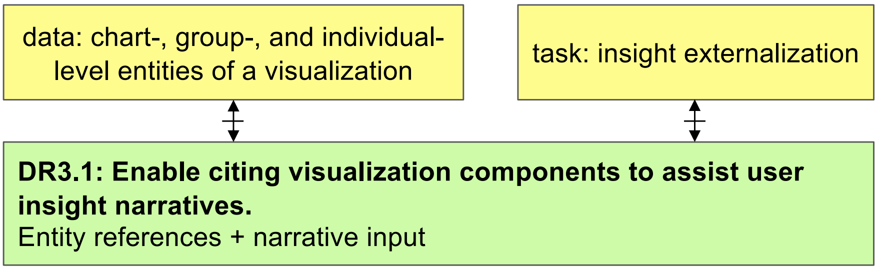

Figure 3.6: The data/task abstraction and interaction technique blocks using entity references for insight externalization.

Table 3.2: Generalizability of the entity references for insight externalization.

Entity types	CO_2 Explorer	MediSyn
Chart-level	A choropleth map, a line chart	An entire resulting view
Group-level	A line / a vertical reference line in a line chart	A table cell with multiple bars, a column / a row of the matrix
Individual-level	A map point	A bar in a table cell, a publication source

include lines and vertical reference lines in the line charts, and individual-level entities are map points (Table 3.2). Additionally, users can refer to public notes as the sixth entity type of the CO_2 Explorer to assist their narratives, which creates a unified mental model in referring to visualizations and notes.

Similar to DR2.2 of Case 2, the CO_2 Explorer supports *scented insight browsing* [27] by attaching the number of related insights to the country and year entities as blue bars in the visualization (Figure 3.5). Users can click on the bars next to the entities to view their related insights. See the video demonstrating the insight externalization feature of the CO_2 Explorer at `https://youtu.be/WX7NmGjBK2s`. Figure 3.6 illustrates the abstraction and interaction blocks of this design.

A crowdsourced study asking users to freely explore the data and externalize insights through writing notes and citing relevant entities (Chapter 4) revealed that group and individual-level entities were more frequently used than chart-level entities in insight externalization.

We can generalize entity references to other visualizations, such as MediSyn of Case 2. A chart-level entity includes the entire resulting visualization, a group-level entity could be a table cell with multiple bars and a column or a row of the matrix, and an individual-level entity denotes a bar in a table cell or a publication source (Table 3.2). Biologists can refer to relevant entities in their notes to help clarify their discoveries.

Compared with logging interaction steps as insight provenance, attaching a customized view might also suffice based on the three considerations. First, the analysis of the interaction steps might not be necessary (too complex) to a simple generic visualization. Second, Ragan et al.'s [120] experiment with text analysis revealed that a final view with analysis visual cues without individual analysis steps could also significantly improve the memory of the analysis process. Third, the automatic method catches all critical/trivial steps, which could be inefficient in communicating analysis rationale. More studies are required to explore the benefits and drawbacks of both methods.

3.4 Discussion and Conclusion

In summary, the three cases support five types of entity-based interactions based on Yi et al.'s [162] interaction taxonomy. Table 3.3 exemplifies the interactions and their implementations across the cases. The *externalize* and *share* interactions are added to indicate user intent of externalizing and exploring insights. We could say that entity-based interaction is flexible in design to support various user intent; the resulting design is inherently user-centric as people naturally perceive things as entities and relate entities to create mental models to understand external information [9, 111]. The following answers RQ1.1-1.3 through the three case studies and proposes design guidelines (DGs) to help improve the entity-based interaction design.

RQ1.1: *How to support VDE of information facets through entity-based interactions?* We considered information facets as classes of entities and devised flexible interactions between linear and categorical facets to enable user retrieval of items of interest. The interface addresses the two requirements derived from prior work to support fluid exploratory search. First, it enables exploration within time- and space-related contexts (DR1.1). Time-related context puts users in the exploration process by visualizing items that were, are, or have not been selected in different color encodings; space-related context is realized through the interactive coordination between the linear and categorical facets. Second, it

Table 3.3: The seven types of interactions and their example instantiations across the three cases.

Interaction	Instantiations
Select	Click on a categorical entity to select related data items (Case 1). Filter-swipe to select items (Case 1). Select an entity of interest from the list to explore entity relations in the visualization (Case 2).
Filter	Drag & drop an entity as a query to filter the item space (Case 1).
Connect	Mouse-over a categorical entity to highlight relevant items in the linear facet (Case 1). Mouse-over an entity to highlight its relevant entities (Case 2).
Elaborate	Mouse-over/click on an entity relation to view its details (Case 2).
Explore	Select an entity from the visualization/note to explore entity relations (Case 2).
Externalize	Cite entities (individual-, group-, and chart-level visual components and prior notes) to assist insight narratives (Case 3).
Share	Click on the visual cue attached to an entity to explore relevant notes (Cases 2 & 3).

allows users to select items through categorical facets while keeping the current search space to support rapid transitions between search criteria (DR1.2). The resulting design has been transferred from emails to tweets and genetic mutation records to demonstrate the generalizability of the entity-based interactions.

To improve the design, we can organize the linear facet to support semantic zoomings, such as organizing by weeks and months, so that patterns could be discovered on various scales of grouped items. This elicits ***DG1:*** *group entities hierarchically for operation.*

RQ1.2: *How to design interaction with drug-target data using entity-based thinking?* Through abstracting drugs, mutations, and tumor types into entities, Case 2 demonstrated how we generalized interactions devised on one type of data (mutations) to other types of data (drugs and tumor types) to increase the perspectives in exploration to support insight discovery (DR2.1). To support the bi-directional exploration of insight and data (DR2.2), we extracted entities of drugs, mutations, and tumor types mentioned in the insights. Users can select the extracted entities from the insight to explore entity relations in the visualization;

meanwhile, visual cues are provided on the entities in the visualization to promote the exploration of relevant insights. A university rankings case demonstrated how the resulting interactions could be transferred to datasets with similar structures.

To improve the design, DG1 could be applied to the visualization. Several domain experts from the MediSyn user study suggested that the drugs could be classified based on their similarities so that experts can explore groups of drugs to discover patterns (Article III). Similarly, mutations could be grouped by genes and tumor types classified by tissues to support interaction in a group manner.

DG2: *Display possible operations next to the entity for clarity.* MediSyn shows possible entity operations in a context menu on hovering, which can be inconvenient according to the user feedback. Like in PivotPaths, we suggest explicitly laying out the operation buttons next to the operating entities. For instance, MediSyn could show action buttons for selecting the entity ☑ and exploring related notes 🗒 beside the entity label to give the user a clear view about possible actions on an entity.

DG3: *Provide entity relation preview.* Query preview is a favored feature in search to help improve search effectiveness [118]. The same goes for entity exploration, as a VDE process resembles a search process. For instance, MediSyn could preview the number of entities related to each entity so that users can select entities for exploration more effectively.

RQ1.3: *How could entity-based interactions support insight externalization?* To support insight externalization, we enabled users to cite visualization components to assist their narratives by abstracting visualization components and insights to entities and devising entity references (DR3.1). We identified individual-, group-, and chart-level visual entities for reference and devised scented insight browsing through providing visual cues on entities to promote insight exploration. The concept of entity references could be applied to other visualizations, such as MediSyn, to support insight externalization. To further improve the design, we could enable users to select groups of visual elements using, e.g., legends [66] for reference (DG1).

DG4: *Distinguish the interaction for VDE from the interaction for entity references.* Referring to entities for insight externalization could interfere with VDE sometimes. For example, with the CO_2 Explorer, clicking on a country on the map to cite a map point conflicts with selecting the country to explore in the line chart. Special concerns need to be taken to harmonize the various interactions, such as using different gestures and single/double clicks for different purposes in the same interaction space.

DG5: *Construct the result view during insight externalization.* During entity references, the representation of the referred entities in the sticky note is text-based, describing the referred data of countries, years, and corresponding emission values. After the user publishes the note, a resulting view is constructed based on the referred data. From visual to text and back to visual again increases cognitive loads. A better alternative is to construct the result view directly on the sticky note replacing the text description.

Chapter 4

Evaluating VDE in Supporting Insight

To investigate RQ2, we conducted two user studies exploring the relations between various interaction types/patterns and insight characteristics (Figure 4.1). Study 1 is a lab study of MediSyn with 14 domain experts (Article III); Study 2 involves a crowdsourced study of the CO_2 Explorer with 158 online workers (Article IV). During the studies, users freely explored the data using the visualizations and wrote notes on their discoveries about the data; we collected user interaction logs and characterized notes (insights) for analysis. In the remainder of this chapter, Section 4.1 explains the study design rationale. Section 4.2 compares the similarities and differences of the two studies, whereas Section 4.3 reports findings. Finally, Section 4.4 discusses study limitations and implications.

4.1 Study Design Rationale

The traditional task-based evaluation that measures task time and error could not evaluate how well visualization supports insight. To measure visualization's ability in supporting open-end discoveries, Saraiya et al. [129] proposed an insight-based method that measures the characteristics of user insights. Based on empirical studies, they characterized insights by direct versus unexpected, breadth versus depth, correctness, domain value, etc. However, measuring visualization insight alone provides limited support in improving visualization; looking into the holistic exploration process, involving interaction logs and user thought processes, could provide better implications about visualization design [97].

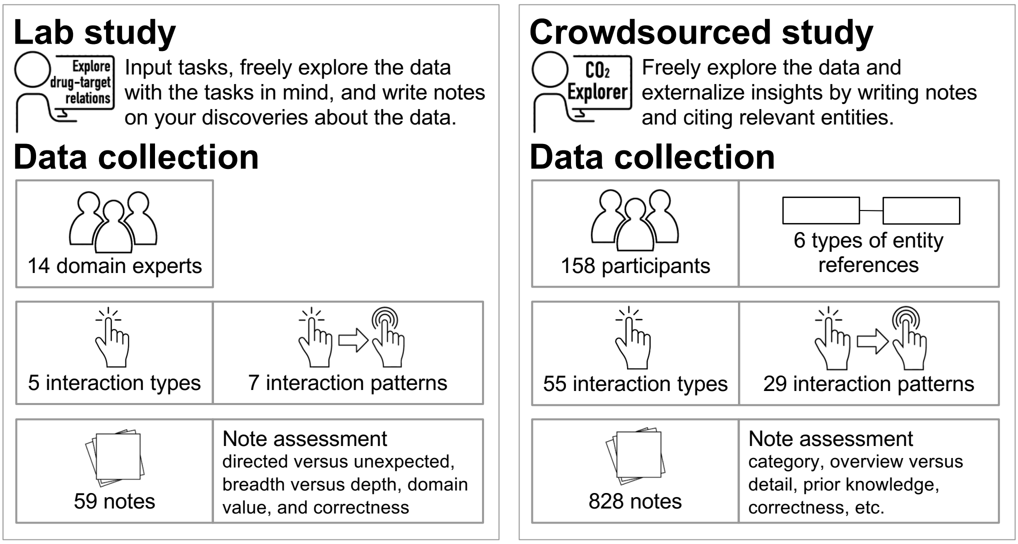

Figure 4.1: Two studies investigating how users derive insights through VDE.

Interaction reflects user intent and contains information on user reasoning/sensemaking processes, which is directly related to generated insights. Guo et al. [56] logged user interactions and insights during VDE and analyzed the correlations between interaction types and insight categories. They found that filtering actions hindered the generation of facts, whereas exploration actions promoted it. We took a similar approach in these two studies. In contrast to Guo et al.'s insight categorization as facts, hypotheses, and generalizations, we measured an insight from multiple perspectives following Saraiya et al.'s [129] method. Study 1 analyzed the correlations between interaction types/patterns and insight characteristics to answer RQ2.1; Study 2 used interaction types/patterns and entity references (Section 3.3) to predict insight characteristics through advanced machine learning models to answer RQ2.2.

4.2 Study Comparisons

This section compares and contrasts the two studies from the three aspects to reflect on the evaluation method: prototype design, evaluation task, and data analysis.

4.2.1 Prototype Design: Domain-Specific versus Generic Visualization and Insight Provenance

Study 1 involved domain experts deriving insights from a domain-specific visualization, MediSyn, which limited the number of target users we can evaluate to gain insightful results. Meanwhile, study results showed that levels of domain knowledge played a critical role in VDE. So with Study 2, we evaluated a generic visualization (CO_2 Explorer), which alleviates the effects of prior knowledge on VDE and allows us to conduct a crowdsourced study with more participants of diverse backgrounds.

In both studies, participants had access to a fixed set of public notes collected from pilot studies. To support insight provenance, MediSyn automatically records interaction steps that lead to insights, whereas the CO_2 Explorer enables users to manually cite entities to attach a result view, as explained in Sections 3.2 and 3.3, respectively. As a result, MediSyn supports five types of interactions based on the categorization by Yi et al. [162], which are *select*, *connect*, *elaborate*, *explore*, and *share* (Figure 3.4); the CO_2 Explorer supports 55 types of interactions (recognized as the action and its operating entity, such as *select a year*) and six types of entity references (Section 3.3).

4.2.2 Evaluation Task: With/Without Inputting a Task

We studied the two visualizations with open-ended tasks by asking users to freely explore the data and write at least five notes on their discoveries about the data. Each note should mention at least one entity name with MediSyn, such as a drug name, or refer to at least one type of entity with the CO_2 Explorer.

To assess the exploratory feature of visualizations, in Study 1, we evaluated whether users' discoveries went beyond their initial tasks (unexpectedness of insights). To achieve this, we asked users to take an extra step in the task by first inputting a task and then exploring the data with their task in mind. If they switched to another task, they needed to open the task window to input the new task. We then evaluated the relations between the discoveries and tasks to rate the unexpectedness of insights.

However, this setting had some limitations that resulted in the collected insights being mostly categorized as direct rather than unexpected. Several participants tried to stick with the inputted task during data exploration, whereas some others inputted a new task inspired by the exploration. Also, to keep the simplicity of the evaluation, Study 2 omitted this step and did not evaluate the unexpectedness of insights. An appropriate method to measure the exploratory feature of visualization in supporting serendipitous discoveries remains an open question here, e.g., a post-task interview by Thudt et al. [145] could potentially help evaluate users' serendipitous discoveries qualitatively.

4.2.3 Data Analysis: Interaction Patterns, Insight Characteristics, and Analysis Methods

During the two studies, we logged user interactions with timestamps and recorded user-inputted notes. To analyze interactions in relation to insight characteristics, we extracted interaction patterns from the collected interaction trails. An interaction pattern denotes a sequence of interactions that frequently appear in interaction trails. In Study 1, we extracted patterns manually (Section 5.4 of Article III), which resulted in seven patterns; in Study 2, we used a more systematic way to extract patterns automatically (Section 5.3.1 of Article IV) and collected 29 patterns.

After removing unqualified notes, such as interface suggestions, and low-quality notes, we collected 59 notes for Study 1 and 828 notes for Study 2. Referring to Saraiya et al.'s [129] insight characterization, Study 1 assessed four insight characteristics: *directness versus unexpectedness*, *breadth versus depth*, *domain value*, and *correctness.* All were rated on 5-point Likert scales, such as from 1 = direct to 5 = unexpected. Study 2 evaluated notes on the *category*, *overview versus detail*, *using prior knowledge*, *correctness*, *the relation between notes and referred charts*, and *the number of entities mentioned.* Notes were *categorized* in a bottom-up manner as statements, comparisons, and groupings. The assessment of *the number of entities mentioned* was tailored to our entity-based insight externalization technique, which counted the numbers of countries, years, and values mentioned in notes. The *overview versus detail* matched the *breadth versus depth* characterization in Study 1, except that we evaluated it on a 3-point Likert scale (0 = overview, 1 = overview + detail, and 2 = detail). The remaining three characteristics (using prior knowledge, correctness, and the relation between notes and referred charts) were evaluated on binary scales (e.g., 0 = no prior knowledge and 1 = with prior knowledge) to keep the assessment simple. The assessment of using prior knowledge was inspired by Study 1, as we discovered that the use of domain knowledge could affect VDE and the quality of generated insight.

Detailed grading criteria on each characteristic is critical to ensure objective and consistent ratings between the evaluators. Two evaluators assessed a subset of the notes independently and then had their grading consistency checked;

evaluators kept revising the grading criteria until they reached a strong grading consistency based on the statistical assessment.

To explore the relations between interactions and insights, Study 1 used correlation analysis to link the number of various interaction types/patterns to insight characteristics, whereas Study 2 went one step further: Study 2 leveraged machine learning models to predict insight characteristics using interactions and entity references. The predictors of interactions were 35 interaction types after removing from the total 55 interaction types the ones we considered not contributing to the prediction. The predictors of entity references included the six types of references the CO_2 Explorer supports and the number of unique countries and years referred to which were extracted from the referred entities. To help explain the prediction performance, we analyzed the participant-wise correlations as we did in Study 1 and used the SHapley Additive exPlanations (SHAP) approach [93] which measures each predictor's contribution to the prediction outcome on individual cases.

4.3 Study Findings

RQ2.1: *What are the correlations between interactions and insight characteristics?* Study 1 analyzed the correlations of five action types (Figure 3.4) and six interaction patterns (omitting one pattern with insufficient data) to the three characteristics of insights without the characteristic of correctness, as most insights we collected were correct. Results provided evidence that exploration actions led to unexpected insights and the drill-down pattern related to insights with higher domain values. Study 2 explored the correlations of three groups of interactions (data/note exploration and edit actions) to the insight characteristics of category, overview versus detail, and using prior knowledge. Regarding the three insight categories of statements, comparisons, and groupings, results showed that statements tended to have fewer data exploration actions, whereas groupings appeared to relate to more data/note exploration actions. Detailed insights seemed to have more mouse-overs in the chart area; using prior knowledge tended to positively relate to note exploration actions.

RQ2.2: *To which extent could we characterize insight through interactions / entity references?* Table 4.1 presents the results of using interactions and entity references to characterize insights. Compared with interaction types, entity references increased the accuracy of insight characterization. Interaction patterns did not produce promising results for the characterization. The reason may be that

Table 4.1: Comparison of characterizing insights by interaction types and entity references. Results were measured by Kappa values on a scale of poor, slight, fair, moderate, substantial, and almost perfect.

	Category	Overview vs. detail	Prior knowledge
Interaction types	fair	slight	fair
Entity references	moderate	fair	fair

what insight is to be discovered during VDE is uncertain, whereas, during entity references, the discovery is made certain. Subsection 4.4.1 discusses possibilities to improve insight characterization using interactions.

To understand the prediction results using entity references, SHAP feature importance and correlation analysis revealed that the comparison and detailed types of insights tended to reference more vertical reference lines in the line charts, whereas groupings appeared to reference more entire charts. Meanwhile, statements seemed to refer to fewer countries, whereas the comparison and grouping types of insights tended to cite more countries, consistent with the correlations between note categories and the assessment of the number of mentioned countries in notes.

A qualitative analysis of insight in Study 1, comparing the insights of two groups of participants with high and low domain values, revealed that insights with high domain values tended to involve domain knowledge, relate to less exploration of public notes, and progress from broad to in-depth discoveries. However, Study 2 with a generic visualization shows that insights involving prior knowledge were inclined to relate to more note exploration actions and, consequently, more note references.

4.4 Discussion and Implications

This section discusses improving insight characterization using interactions and study implications on knowledge-assisted visualization.

4.4.1 Interaction & Insight

To improve insight characterization using interactions, we need to recognize which interactions contribute to the specific insight. The two studies collected interactions from the start till the note editing action as provenance for correlation and prediction analysis, which potentially involved irrelevant steps in the insight characterization. The need grows when users conduct long exploration sessions, which is not rare in practice.

However, this is a challenging task. User analysis could involve multiple tasks at the same time [128]; an interaction could contribute to one or more insights. Even interactions that do not result in explicit insights may have implications on user understanding of the data [164].

We present some directions for thought. A technique of segmenting interactions to higher-level activities by Yan et al. [161], which considers the goal and intent of the interaction and the data attributes, may help insight characterization. Interaction segments that involve the entities referred to in the insight may be more relevant to the insight than those that do not. Further, involving interaction paces in the analysis, i.e., leveraging the time dimension of interaction [7, 47], may improve the interaction segmentation. Besides, the level of details of interaction data could impact the prediction results. For instance, *select a year* and *select the year 1990* indicate different levels of interaction details. Battle and Heer [7] found that revisitation of states signaled significant analysis states. More detailed interaction data could be evaluated in Study 2.

4.4.2 Knowledge-Assisted Visualization

Knowledge-assisted visualization aims to share expert knowledge among users to support VDE [23]. Federico et al. [45] proposed a conceptual model of knowledge-assisted visual analytics where systems could utilize users' implicit/explicit knowledge externalization, such as parameter settings and note taking, to support visual analysis.

This research implies that automatic insight characterization based on user interactions and entity references could support VDE through guidance and insight recommendation. Learning about the interaction patterns that lead to insight, systems could guide users through the VDE to help discover more insights [33]. For instance, Study 1 showed that the drill-down pattern could lead to insight with high domain values; systems could guide users through this pattern to discover insights.

The type of insight could be called interaction-aware insights. Compared to data-aware annotations, with interaction-aware insights, systems are aware of the interaction states that lead to insight. As mentioned earlier, capturing interaction automatically records all trivial/non-trivial steps. Interactions that directly relate to insight generation, such as parameter tuning for model building and settings for data filtering, can take more weight in user guidance. Also, hybrid methods involving user intervention to identify significant steps for reasoning and sensemaking could be studied.

On the other hand, systems could recommend insights based on user interaction and operating entities. For instance, assume that elaboration actions lead to deep insight; when a user frequently elaborates on certain entities, the system could suggest in-depth insights referring to these entities inferred by the interactions. Leveraging data mining and machine learning techniques, systems could construct recommendation models with complex interaction patterns.

Moreover, machine-readable note-taking as discussed in Section 2.4, such as using note templates and natural language processing, could further enhance VDE. Constructing notes as entities and entity relations is not only intuitive for users to annotate but also convenient for machines to interpret. For instance, MediSyn provides data-aware annotations by extracting entities from notes and supports the bi-directional exploration of entities and notes, as explained in Section 3.2.

Chapter 5

Discussion

This chapter answers the two RQs we proposed in Chapter 1, discusses limitations and future work of this research, and concludes this dissertation.

5.1 Answers to RQs

RQ1: *How could we design interaction to support VDE?* To facilitate actionability of interaction design in visualization, this research proposes to abstract data into entities and design entity operations to realize higher-level interaction goals. Following the nested model of visualization design, Chapter 3 demonstrated the entity-based interaction design in practice through three case studies. The cases detailed how this approach could address the various DRs derived from prior work to support various task goals and the resulting design's transferability to other datasets.

The first case depicted the interaction design of information facets to support fluid exploratory search. Visual cues on the items and the interactive coordination of facets provide time- and space-related contexts (DR1.1); the interaction of selecting items through entities of interest without filtering the item space supports rapid transitions between search criteria (DR1.2). The resulting design could be transferred from emails to other similar datasets, such as tweets and patient oncogene profiles. The second case illustrated the generalizability of the entity-based interaction design from one type of data (mutations) to other types of data (drugs and tumor types) to support exploration from multiple perspectives to assist insight discovery (DR2.1), detailed the use of entities to bridge the bi-directional exploration of insight and data (DR2.2), and also discussed the transferability of the design to the university rankings dataset. The third case exemplified the interaction of citing relevant entities (visualization components and prior notes) for insight externalization (DR3.1) in a CO_2 Explorer with choropleth maps and line charts. The abstraction of visualization components into three levels of entities enables the technique of entity references to be applied to all types of visualizations to support insight externalization.

To improve the entity-based interactions resulting from the case studies, we provided five guidelines: 1) grouping entities hierarchically for operation, 2) displaying possible operations next to the entity for clarity, 3) providing entity relation preview, 4) distinguishing the interaction for VDE from the interaction for entity references, and 5) constructing the result view during insight externalization.

RQ2: *How do users discover insights through interacting with the visualization?* Study results of MediSyn provided evidence that exploration actions led to unexpected insights; the drill-down pattern related to insights with higher domain values. The study of the CO_2 Explorer collected insight categories of statements, comparisons, and groupings and revealed that simple statements, mostly mentioned one country entity, tended to relate to fewer data exploration actions, whereas groupings, mostly discussed more than two country entities, appeared to have more data/note exploration actions and referred to more charts. The comparison and detailed types of insights tended to reference more vertical reference lines in the line charts; meanwhile, detailed notes appeared to have more mouse-overs in the chart area.

A qualitative analysis of insight generation indicated that participants who derived insights with high domain values tended to explore others' notes less and advance from broad to in-depth discoveries using their domain knowledge. However, with a generic visualization, Study 2 found that using prior knowledge in insights tended to relate to more note exploration actions and more note references.

5.2 Limitations

We discuss limitations in the evaluation of the entity-based interaction and the generalizability of the user study.

5.2.1 Evaluation of the Interaction Design

The case studies lack an evaluation with baseline systems to inform the merits and demerits of the entity-based interaction. The challenge lies in the design of a baseline system. Keeping the same visual design, the evaluating systems could be 1) entity-based interaction on one type of entity versus entity-based interaction generalized to all types of entities in the visualization, or 2) with/without a type of entity-based action. In the former case, the choice of the entity type could induce bias in evaluation results, whereas the latter is concerned with the task that needs to motivate users to use the specific action. For instance, to evaluate the *scented insight browsing* of the CO_2 Explorer in supporting insight, we can use the same visualization without the functionality as the baseline. In this case, how the task would motivate users to explore public insights would be a central issue. Special consideration needs to be paid to a fair comparison.

5.2.2 Generalizability of the User Study

The evaluation method could be generalized to study other visualizations to validate the results. Although the bottom-up insight categories and extracted interaction patterns could be specific to a certain type of visualization, the ways to characterize insights and to extract interaction patterns are general. Entity references could also be generalized, as discussed in the case study. Due to the controlled experiment settings, study results could be biased considering the following aspects: 1) the training program before the actual experiment introduced the features of the visualization in a certain order, which could influence the way users interact with the visualization as noted by Wesslen et al. [153]; 2) the tasks required users to mention or reference at least one entity, which could induce bias in the collected insights; and 3) users could access a fixed set of public notes in both studies, which could inspire, but could also hinder user thinking and the diversity of the resulted insights despite the open-ended task settings.

5.3 Future Directions

Apart from the future work hinted at in Section 4.4 as study implications, we present two directions building on current research. To further support insight, visualization could structure insight by entities and record uncertainty-aware insight and insight provenance across platforms.

5.3.1 Structuring Insight by Entities

As mentioned earlier, entities are widely used in text analysis and information retrieval. Borrowing the knowledge from text analysis, we can structure insight by entities instead of simple narratives so that novel mixed-initiative systems are possible [76]. For instance, systems could generate a graph/matrix of co-occurring entities in insights to expose over-explored/under-explored relations. This research extracted simple entities, including drugs, tumor types, and mutations in MediSyn, and countries and years in the CO_2 Explorer, to support insight exploration; more complex structures among entities could be built to further support reasoning, e.g., drawing relations between nouns and verbs.

Leveraging the knowledge from information retrieval, systems with entity-based insight could retrieve evidence from external sources to support insight validation, which can potentially alleviate the multiple comparisons problem mentioned in Chapter 2 and expand/direct information exploration (e.g., [20, 28, 155, 167]). For instance, CAVA enables users to expand their datasets using a broad set of attributes crawled from external knowledge graphs [20]. If the crawling is aware of user interactions and insights, the system could then provide more focused assistance addressing user needs.

5.3.2 Supporting Insight and Insight Provenance

Developing a platform that records insight and insight provenance enables data exploration across analytic tools, which would be helpful in practice [2, 6, 75, 157]. Moreover, uncertainty always accompanies analysis, passed on from the data source to the model and then to the visualization, but is often overlooked in the actual analysis [71]. Quantifying and visualizing uncertainty along the analytic provenance can support users to calibrate their trust and make informed decisions [127].

5.4 Conclusion

A review of related work on VDE revealed that, among other findings, a lack of actionability in interaction design hinders the application of the proposed visualization design models and frameworks; the community lacks an understanding of the user insight generation process during VDE. This research attempted to address these two challenges by proposing an interaction design approach and conducting two user studies to explore user insight generation.

To provide an anchor point in interaction design thinking, this research proposed abstracting data to entities and designing entity-based interaction. Three case studies demonstrated the applicability of this approach in addressing the various DRs derived from related work to fulfill the various task goals, i.e., fluid exploratory search, insight generation and sharing, and insight externalization through entity (visualization components and prior notes) references. Three cases also illustrated the transferability of the resulting interaction design to other datasets with similar structures. Reflecting on the case studies, we proposed five guidelines to help improve the entity-based interactions.

To understand the user insight generation process, we presented two user studies asking users to freely explore the visualization and write notes on their discoveries about the data. Through correlating the types/patterns of user interactions with the characteristics of their generated insights, the first study indicated the potential of using interactions to characterize insights. However, the second study of the CO_2 Explorer showed that entity references were better at characterizing insights than interactions. Recognizing significant interaction states that contribute to the insight could potentially improve their insight prediction. Study results imply the potential of building novel knowledge-assisted visualizations that automatically characterize insights to provide exploration guidance and insight recommendations.

As future work, research could build on current results to support uncertainty-aware insight and insight provenance across platforms. Visualization tools could also structure insight by entities to support novel mixed-initiative systems, such as systems that could analyze insights and retrieve external resources to support VDE.

References

[1] Mohammad Alharbi and Robert S. Laramee. SoS TextVis: An extended survey of surveys on text visualization. *Computers*, 8(1):17, 2019.

[2] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A. Hearst. Futzing and moseying: Interviews with professional data analysts on exploration practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):22–31, 2019.

[3] Salvatore Andolina, Khalil Klouche, Tuukka Ruotsalo, Patrik Floréen, and Giulio Jacucci. Querytogether: Enabling entity-centric exploration in multi-device collaborative search. *Information Processing and Management*, 54(6):1182–1202, 2018.

[4] Christopher Andrews and Chris North. Analyst's Workspace: An embodied sensemaking environment for large, high-resolution displays. In *IEEE Conference on Visual Analytics Science and Technology*, pages 123–131. IEEE Computer Society, 2012.

[5] Ali Baigelenov and Paul Parsons. Interactivity factors in visualization-based exploratory search. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, 2018.

[6] Andrea Batch and Niklas Elmqvist. The interactive visualization gap in initial exploratory data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):278–287, 2018.

[7] Leilani Battle and Jeffrey Heer. Characterizing exploratory visual analysis: A literature review and evaluation of analytic provenance in tableau. *Computer Graphics Forum*, 38(3):145–159, 2019.

[8] Roger Beecham, Jason Dykes, Chris Rooney, and William Wong. Design exposition discussion documents for rich design discourse in applied visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(8):3451–3462, 2021.

[9] Eric A. Bier, Stuart K. Card, and John W. Bodnar. Principles and tools for collaborative entity-based intelligence analysis. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):178–191, 2010.

[10] Tanja Blascheck, Fabian Beck, Sebastian Baltes, Thomas Ertl, and Daniel Weiskopf. Visual analysis and coding of data-rich user behavior. In *IEEE Conference on Visual Analytics Science and Technology*, pages 141–150. IEEE Computer Society, 2016.

[11] Tanja Blascheck, Markus John, Kuno Kurzhals, Steffen Koch, and Thomas Ertl. VA^2: A visual analytics approach for evaluating visual analytics applications. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):61–70, 2016.

[12] Christian Bors, Theresia Gschwandtner, and Silvia Miksch. Capturing and visualizing provenance from data wrangling. *IEEE Computer Graphics and Applications*, 39(6):61–75, 2019.

[13] Christian Bors, John E. Wenskovitch, Michelle Dowling, Simon Attfield, Leilani Battle, Alex Endert, Olga Kulyk, and Robert S. Laramee. A provenance task abstraction framework. *IEEE Computer Graphics and Applications*, 39(6):46–60, 2019.

[14] Edward M. Bowden, Mark Jung-Beeman, Jessica Fleck, and John Kounios. New approaches to demystifying insight. *Trends in Cognitive Sciences*, 9(7):322–328, 2005.

[15] Eli T. Brown, Alvitta Ottley, Helen Zhao, Quan Lin, Richard Souvenir, Alex Endert, and Remco Chang. Finding Waldo: Learning about users from their interactions. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1663–1672, 2014.

[16] Steven P. Callahan, Juliana Freire, Emanuele Santos, Carlos E. Scheidegger, Cláudio T. Silva, and Huy T. Vo. VisTrails: Visualization meets data management. In *the SIGMOD International Conference on Management of Data*, pages 745–747. ACM, 2006.

[17] Akhilesh Camisetty, Chaitanya Chandurkar, Maoyuan Sun, and David Koop. Enhancing web-based analytics applications through provenance. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):131–141, 2019.

[18] Nan Cao, Jimeng Sun, Yu-Ru Lin, David Gotz, Shixia Liu, and Huamin Qu. FacetAtlas: Multifaceted visualization for rich text corpora. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1172–1181, 2010.

[19] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in information visualization - using vision to think*. Academic Press, 1999.

[20] Dylan Cashman, Shenyu Xu, Subhajit Das, Florian Heimerl, Cong Liu, Shah Rukh Humayoun, Michael Gleicher, Alex Endert, and Remco Chang. CAVA: A visual analytics system for exploratory columnar data augmentation using knowledge graphs. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1731–1741, 2021.

[21] Remco Chang, Caroline Ziemkiewicz, Tera Marie Green, and William Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009.

[22] Min Chen and David S. Ebert. An ontological framework for supporting the design and evaluation of visual analytics systems. *Computer Graphics Forum*, 38(3):131–144, 2019.

[23] Min Chen, David S. Ebert, Hans Hagen, Robert S. Laramee, Robert van Liere, Kwan-Liu Ma, William Ribarsky, Gerik Scheuermann, and Deborah Silver. Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1):12–19, 2009.

[24] Min Chen and Darren J. Edwards. "Isms" in visualization. In *Foundations of Data Visualization*, pages 225–241. Springer, 2020.

[25] Peter P. Chen. The entity-relationship model - toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36, 1976.

[26] Siming Chen, Natalia V. Andrienko, Gennady L. Andrienko, Linara Adilova, Jérémie Barlet, Jörg Kindermann, Phong H. Nguyen, Olivier Thonnard, and Cagatay Turkay. LDA ensembles for interactive exploration and categorization of behaviors. *IEEE Transactions on Visualization and Computer Graphics*, 26(9):2775–2792, 2020.

[27] Yang Chen, Scott Barlowe, and Jing Yang. Click2Annotate: Automated insight externalization with rich semantics. In *IEEE Conference on Visual Analytics Science and Technology*, pages 155–162. IEEE Computer Society, 2010.

[28] Isaac Cho, Wenwen Dou, Derek Xiaoyu Wang, Eric Sauda, and William Ribarsky. VAiRoma: A visual analytics system for making sense of places, times, and events in roman history. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):210–219, 2016.

[29] Eun Kyoung Choe, Bongshin Lee, and M. C. Schraefel. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE Computer Graphics and Applications*, 35(4):28–37, 2015.

[30] In Kwon Choi, Taylor Childers, Nirmal Kumar Raveendranath, Swati Mishra, Kyle Harris, and Khairi Reda. Concept-driven visual analytics: An exploratory study of model- and hypothesis-based reasoning with visualizations. In *the CHI Conference on Human Factors in Computing Systems*, pages 1–14. ACM, 2019.

[31] Zhe Cui, Sriram Karthik Badam, Mehmet Adil Yalçin, and Niklas Elmqvist. DataSite: Proactive visual data exploration with computation of insight-based recommendations. *Information Visualization*, 18(2):251–267, 2019.

[32] Zach Cutler, Kiran Gadhave, and Alexander Lex. Trrack: A library for provenance-tracking in web-based visualizations. In *IEEE Visualization Conference - Short Papers*, pages 116–120. IEEE, 2020.

[33] Filip Dabek and Jesus J. Caban. 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):41–50, 2017.

[34] Evanthia Dimara and Charles Perin. What is interaction for data visualization? *IEEE Transactions on Visualization and Computer Graphics*, 26(1):119–129, 2020.

[35] Marian Dörk, Sheelagh Carpendale, Christopher Collins, and Carey Williamson. Visgets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1205–1212, 2008.

[36] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. Fluid views: A zoomable search environment. In *the International Working Conference on Advanced Visual Interfaces*, pages 233–240. ACM, 2012.

[37] Marian Dörk, Nathalie Henry Riche, Gonzalo A. Ramos, and Susan T. Dumais. PivotPaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2709–2718, 2012.

[38] Wenwen Dou, Dong Hyun Jeong, Felesia Stukes, William Ribarsky, Heather Richter Lipford, and Remco Chang. Recovering reasoning processes from user interactions. *IEEE Computer Graphics and Applications*, 29(3):52–61, 2009.

[39] Wenwen Dou, William Ribarsky, and Remco Chang. Capturing reasoning process through user interaction. In *International Symposium on Visual Analytics Science and Technology, EuroVAST@EuroVis*, pages 33–38. Eurographics Association, 2010.

[40] Cody Dunne, Nathalie Henry Riche, Bongshin Lee, Ronald A. Metoyer, and George G. Robertson. GraphTrail: Analyzing large multivariate, heterogeneous networks while supporting exploration history. In *the CHI Conference on Human Factors in Computing Systems*, pages 1663–1672. ACM, 2012.

[41] Omar ElTayeby and Wenwen Dou. A survey on interaction log analysis for evaluating exploratory visualizations. In *the Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 62–69. ACM, 2016.

[42] Alex Endert, Remco Chang, Chris North, and Michelle X. Zhou. Semantic interaction: Coupling cognition and computation through usable interactive analytics. *IEEE Computer Graphics and Applications*, 35(4):94–99, 2015.

[43] Alex Endert, Patrick Fiaux, and Chris North. Semantic interaction for sensemaking: Inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, 2012.

[44] Alex Endert, Patrick Fiaux, and Chris North. Semantic interaction for visual text analytics. In *the CHI Conference on Human Factors in Computing Systems*, pages 473–482. ACM, 2012.

[45] Paolo Federico, Markus Wagner, Alexander Rind, Albert Amor-Amoros, Silvia Miksch, and Wolfgang Aigner. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In *IEEE Conference on Visual Analytics Science and Technology*, pages 92–103. IEEE Computer Society, 2017.

[46] Jean-Daniel Fekete, T. J. Jankun-Kelly, Melanie Tory, and Kai Xu. Provenance and logging for sense making (report from Dagstuhl seminar 18462). *Dagstuhl Reports*, 8(11):35–62, 2019.

[47] Mi Feng, Evan M. Peck, and Lane Harrison. Patterns and pace: Quantifying diverse exploration behavior with visualizations on the web. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):501–511, 2019.

[48] Ana Figueiras. Towards the understanding of interaction in information visualization. In *International Conference on Information Visualisation*, pages 140–147. IEEE Computer Society, 2015.

[49] David Gotz and Zhen Wen. Behavior-driven visualization recommendation. In *the International Conference on Intelligent User Interfaces*, pages 315–324. ACM, 2009.

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

[51] Samuel Gratzl, Nils Gehlenborg, Alexander Lex, Hanspeter Pfister, and Marc Streit. Domino: Extracting, comparing, and manipulating subsets across multiple tabular datasets. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2023–2032, 2014.

[52] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, N. Cosgrove, and Marc Streit. From visual exploration to storytelling and back again. *Computer Graphics Forum*, 35(3):491–500, 2016.

[53] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. LineUp: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013.

[54] Tera Marie Green, William Ribarsky, and Brian D. Fisher. Visual analytics for complex concepts using a human cognition model. In *IEEE Symposium*

on Visual Analytics Science and Technology, pages 91–98. IEEE Computer Society, 2008.

[55] Tera Marie Green, Ron Wakkary, and Richard Arias-Hernández. Expanding the scope: Interaction design perspectives for visual analytics. In *Hawaii International International Conference on Systems Science*, pages 1–10. IEEE Computer Society, 2011.

[56] Hua Guo, Steven R. Gomez, Caroline Ziemkiewicz, and David H. Laidlaw. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):51–60, 2016.

[57] Yi Guo, Shunan Guo, Zhuochen Jin, Smiti Kaul, David Gotz, and Nan Cao. Survey on visual analysis of event sequence data. *CoRR*, abs/2006.14291, 2020.

[58] Andreas Hall, Paula Ahonen-Rainio, and Kirsi Virrantaus. Insight provenance for spatiotemporal visual analytics: Theory, review, and guidelines. *Journal of Spatial Information Science*, 15(1):65–88, 2017.

[59] Kyle Wm. Hall, Adam James Bradley, Uta Hinrichs, Samuel Huron, Jo Wood, Christopher Collins, and Sheelagh Carpendale. Design by immersion: A transdisciplinary approach to problem-driven visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):109–118, 2020.

[60] Gillian R. Hayes. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction*, 18(3):15:1–15:20, 2011.

[61] Chen He, Luana Micallef, Liye He, Gopal Peddinti, Tero Aittokallio, and Giulio Jacucci. Characterizing the quality of insight by interactions: A case study. *IEEE Transactions on Visualization and Computer Graphics*, 27(8):3410–3424, 2021.

[62] Chen He, Luana Micallef, Barış Serim, Tung Vuong, Tuukka Ruotsalo, and Giulio Jacucci. Interactive visual facets to support fluid exploratory search. In *the International Symposium on Visual Information Communication and Interaction*. ACM, 2021.

[63] Chen He, Luana Micallef, Ziaurrehman Tanoli, Samuel Kaski, Tero Aittokallio, and Giulio Jacucci. MediSyn: Uncertainty-aware visualization of multiple biomedical datasets to support drug treatment selection. *BMC Bioinformatics*, 18(S-10):393:1–393:12, 2017.

[64] Chen He, Tung Vuong, and Giulio Jacucci. Characterizing visualization insights through entity-based interaction: An exploratory study. Submitted.

[65] Marti A. Hearst, Ame Elliott, Jennifer English, Rashmi R. Sinha, Kirsten Swearingen, and Ka-Ping Yee. Finding the flow in web site search. *Communications of the ACM*, 45(9):42–49, 2002.

[66] Jeffrey Heer, Maneesh Agrawala, and Wesley Willett. Generalized selection via interactive query relaxation. In *the CHI Conference on Human Factors in Computing Systems*, pages 959–968. ACM, 2008.

[67] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: Using Mechanical Turk to assess visualization design. In *the CHI Conference on Human Factors in Computing Systems*, pages 203–212. ACM, 2010.

[68] Jeffrey Heer and Ben Shneiderman. Interactive dynamics for visual analysis. *Communications of the ACM*, 55(4):45–54, 2012.

[69] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Communications of the ACM*, 52(1):87–97, 2009.

[70] Enamul Hoque and Giuseppe Carenini. ConVis: A visual text analytic system for exploring blog conversations. *Computer Graphics Forum*, 33(3):221–230, 2014.

[71] Jessica Hullman. Why authors don't visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, 2020.

[72] Stefan Jänicke, Pawandeep Kaur, Pawel Kuzmicki, and Johanna Schmidt. Participatory visualization design as an approach to minimize the gap between research and application. In *the Workshop on the Gap between Visualization Research and Visualization Software*, pages 35–42. Eurographics Association, 2020.

[73] T. J. Jankun-Kelly, Kwan-Liu Ma, and Michael Gertz. A model and framework for visualization exploration. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):357–369, 2007.

[74] Waqas Javed and Niklas Elmqvist. ExPlates: Spatializing interactive analysis to scaffold visual exploration. *Computer Graphics Forum*, 32(3):441–450, 2013.

[75] Eser Kandogan, Aruna Balakrishnan, Eben M. Haber, and Jeffrey S. Pierce. From data to insight: Work practices of analysts in the enterprise. *IEEE Computer Graphics and Applications*, 34(5):42–50, 2014.

[76] Eser Kandogan and Ulrich Engelke. Towards a unified representation of insight in human-in-the-loop analytics: A user study. In *the Workshop on Human-In-the-Loop Data Analytics, HILDA@SIGMOD*, pages 3:1–3:7. ACM, 2018.

[77] Benjamin Karer, Hans Hagen, and Dirk J. Lehmann. Insight beyond numbers: The impact of qualitative factors on visual data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1011–1021, 2021.

[78] Daniel A. Keim. Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

[79] Daniel A. Keim, Gennady L. Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. In *Information Visualization - Human-Centered Issues and Perspectives*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer, 2008.

[80] Yea-Seul Kim, Nathalie Henry Riche, Bongshin Lee, Matthew Brehmer, Michel Pahud, Ken Hinckley, and Jessica Hullman. Inking your insights: Investigating digital externalization behaviors during data analysis. In *the ISS International Conference on Interactive Surfaces and Spaces*, pages 255–267. ACM, 2019.

[81] Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. Designing for exploratory search on touch devices. In *the CHI Conference on Human Factors in Computing Systems*, pages 4189–4198. ACM, 2015.

[82] Khalil Klouche, Tuukka Ruotsalo, and Giulio Jacucci. From Hyperlinks to Hypercues: Entity-based affordances for fluid information exploration. In *the Designing Interactive Systems Conference*, pages 401–411. ACM, 2018.

[83] Ratanond Koonchanok, Parul Baser, Abhinav Sikharam, Nirmal Kumar Raveendranath, and Khairi Reda. Data prophecy: Exploring the effects of belief elicitation in visual analytics. In *the CHI Conference on Human Factors in Computing Systems*, pages 18:1–18:12. ACM, 2021.

[84] Po-Ming Law, Alex Endert, and John T. Stasko. Characterizing automated data insights. In *IEEE Visualization Conference - Short Papers*, pages 171–175. IEEE, 2020.

[85] Po-Ming Law, Alex Endert, and John T. Stasko. What are data insights to professional visualization users? In *IEEE Visualization Conference - Short Papers*, pages 181–185. IEEE, 2020.

[86] Bongshin Lee, Petra Isenberg, Nathalie Henry Riche, and Sheelagh Carpendale. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2689–2698, 2012.

[87] Alexander Lex, Marc Streit, Hans-Jörg Schulz, Christian Partl, Dieter Schmalstieg, Peter J. Park, and Nils Gehlenborg. StratomeX: Visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization. *Computer Graphics Forum*, 31(3):1175–1184, 2012.

[88] Jianping Kelvin Li, Shenyu Xu, Yecong (Chris) Ye, and Kwan-Liu Ma. Resolving conflicting insights in asynchronous collaborative visual analysis. *Computer Graphics Forum*, 39(3):497–509, 2020.

[89] Heather Richter Lipford, Felesia Stukes, Wenwen Dou, Matthew E. Hawkins, and Remco Chang. Helping users recall their reasoning process. In *IEEE Conference on Visual Analytics Science and Technology*, pages 187–194. IEEE Computer Society, 2010.

[90] Zhicheng Liu and John T. Stasko. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):999–1008, 2010.

[91] Zhicheng Liu, Yang Wang, Mira Dontcheva, Matthew Hoffman, Seth Walker, and Alan Wilson. Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):321–330, 2017.

[92] Jie Lu, Zhen Wen, Shimei Pan, and Jennifer C. Lai. Analytic trails: Supporting provenance, collaboration, and reuse for visual data analysis by business users. In *the International Conference on Human-Computer Interaction - INTERACT, Part IV*, volume 6949 of *Lecture Notes in Computer Science*, pages 256–273. Springer, 2011.

[93] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *the Conference on Neural Information Processing Systems*, pages 4765–4774, 2017.

[94] Jock D. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, 1986.

[95] Karthic Madanagopal, Eric D. Ragan, and Perakath C. Benjamin. Analytic provenance in practice: The role of provenance in real-world visualization and data analysis environments. *IEEE Computer Graphics and Applications*, 39(6):30–45, 2019.

[96] Andreas Mathisen, Tom Horak, Clemens Nylandsted Klokmose, Kaj Grønbæk, and Niklas Elmqvist. InsideInsights: Integrating data-driven reporting in collaborative visual analytics. *Computer Graphics Forum*, 38(3):649–661, 2019.

[97] Eva Mayr, Michael Smuc, and Hanna Risku. Many roads lead to rome: Mapping users' problem-solving strategies. *Information Visualization*, 10(3):232–247, 2011.

[98] Sean McKenna, Dominika Mazur, James Agutter, and Miriah D. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200, 2014.

[99] Miriah D. Meyer, Michael Sedlmair, P. Samuel Quinan, and Tamara Munzner. The nested blocks and guidelines model. *Information Visualization*, 14(3):234–249, 2015.

[100] Shayan Monadjemi, Roman Garnett, and Alvitta Ottley. Competing models: Inferring exploration patterns and information relevance via bayesian model selection. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):412–421, 2021.

[101] Michael J. Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. How data science workers work with data: Discovery, capture, curation, design, creation. In *the CHI Conference on Human Factors in Computing Systems*, pages 1–15. ACM, 2019.

[102] Tamara Munzner. A nested process model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):921–928, 2009.

[103] Tamara Munzner. *Visualization Analysis and Design*. A K Peters visualization series. A K Peters/CRC Press, 2014.

[104] Mathieu Nancel and Andy Cockburn. Causality: A conceptual model of interaction history. In *the CHI Conference on Human Factors in Computing Systems*, pages 1777–1786. ACM, 2014.

[105] Phong H. Nguyen, Rafael Henkin, Siming Chen, Natalia V. Andrienko, Gennady L. Andrienko, Olivier Thonnard, and Cagatay Turkay. VASABI: hierarchical user profiles for interactive visual user behaviour analytics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):77–86, 2020.

[106] Phong H. Nguyen, Cagatay Turkay, Gennady Andrienko, Natalia Andrienko, Olivier Thonnard, and Jihane Zouaoui. Understanding user behaviour through action sequences: From the usual to the unusual. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2838–2852, 2019.

[107] Phong H. Nguyen, Kai Xu, Andy Bardill, Betul Salman, Kate Herd, and B. L. William Wong. SenseMap: Supporting browser-based online sensemaking through analytic provenance. In *IEEE Conference on Visual Analytics Science and Technology*, pages 91–100. IEEE Computer Society, 2016.

[108] Phong H. Nguyen, Kai Xu, Ashley Wheat, B. L. William Wong, Simon Attfield, and Bob Fields. SensePath: Understanding the sensemaking

process through analytic provenance. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):41–50, 2016.

[109] Donald A. Norman. *The Design of Everyday Things.* Basic Books, New York, 2nd edition, 2013.

[110] Chris North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006.

[111] Sajan Raj Ojha, Mladjan Jovanovic, and Fausto Giunchiglia. Entity-centric visualization of open data. In *the International Conference on Human-Computer Interaction - INTERACT, Part III*, volume 9298 of *Lecture Notes in Computer Science*, pages 149–166. Springer, 2015.

[112] Alvitta Ottley, Roman Garnett, and Ran Wan. Follow the clicks: Learning and anticipating mouse interactions during exploratory data analysis. *Computer Graphics Forum*, 38(3):41–52, 2019.

[113] Paul Parsons and Kamran Sedig. Adjustable properties of visual representations: Improving the quality of human-information interaction. *Journal of the Association for Information Science and Technology*, 65(3):455–482, 2014.

[114] Paul Parsons, Kamran Sedig, Arman Didandeh, and Arash Khosravi. Interactivity in visual analytics: Use of conceptual frameworks to support human-centered design of a decision-support tool. In *Hawaii International Conference on System Sciences*, pages 1138–1147. IEEE Computer Society, 2015.

[115] Jaakko Peltonen, Kseniia Belorustceva, and Tuukka Ruotsalo. Topic-relevance map: Visualization for improving search result comprehension. In *the International Conference on Intelligent User Interfaces*, pages 611–622. ACM, 2017.

[116] William A. Pike, John T. Stasko, Remco Chang, and Theresa A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009.

[117] Margit Pohl, Michael Smuc, and Eva Mayr. The user puzzle - explaining the interaction with visual analytics systems. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2908–2916, 2012.

[118] Pernilla Qvarfordt, Gene Golovchinsky, Tony Dunnigan, and Elena Agapie. Looking ahead: Query preview in exploratory search. In *the SIGIR International Conference on Research and Development in Information Retrieval*, pages 243–252. ACM, 2013.

[119] Eric D. Ragan, Alex Endert, Jibonananda Sanyal, and Jian Chen. Characterizing provenance in visualization and data analysis: An organizational framework of provenance types and purposes. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):31–40, 2016.

[120] Eric D. Ragan, John R. Goodall, and Albert Tung. Evaluating how level of detail of visual history affects process memory. In *the CHI Conference on Human Factors in Computing Systems*, pages 2711–2720. ACM, 2015.

[121] Khairi Reda, Andrew E. Johnson, Michael E. Papka, and Jason Leigh. Modeling and evaluating user behavior in exploratory visual analysis. *Information Visualization*, 15(4):325–339, 2016.

[122] Dylan Rees and Robert S. Laramee. A survey of information visualization books. *Computer Graphics Forum*, 38(1):610–646, 2019.

[123] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Höllerer, and Eun Kyoung Choe. ChartAccent: Annotation for data-driven storytelling. In *IEEE Pacific Visualization Symposium*, pages 230–239. IEEE Computer Society, 2017.

[124] Hugo Romat, Emmanuel Pietriga, Nathalie Henry Riche, Ken Hinckley, and Caroline Appert. SpaceInk: Making space for in-context annotations. In *the ACM Symposium on User Interface Software and Technology*, pages 871–882. ACM, 2019.

[125] Hugo Romat, Nathalie Henry Riche, Ken Hinckley, Bongshin Lee, Caroline Appert, Emmanuel Pietriga, and Christopher Collins. ActiveInk: (Th)Inking with data. In *the CHI Conference on Human Factors in Computing Systems*, pages 1–13. ACM, 2019.

[126] Tuukka Ruotsalo, Giulio Jacucci, and Samuel Kaski. Interactive faceted query suggestion for exploratory search: Whole-session effectiveness and interaction engagement. *Journal of the Association for Information Science and Technology*, 71(7):742–756, 2020.

[127] Dominik Sacha, Hansi Senaratne, Bum Chul Kwon, Geoffrey P. Ellis, and Daniel A. Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, 2016.

[128] Dominik Sacha, Andreas Stoffel, Florian Stoffel, Bum Chul Kwon, Geoffrey P. Ellis, and Daniel A. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, 2014.

[129] Purvi Saraiya, Chris North, and Karen Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):443–456, 2005.

[130] Monica M. C. Schraefel, Max L. Wilson, Alistair Russell, and Daniel A. Smith. mspace: Improving information access to multimedia domains with multimodal exploratory search. *Communications of the ACM*, 49(4):47–49, 2006.

[131] Kamran Sedig and Paul Parsons. Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction*, 5(2):84–133, 2013.

[132] Kamran Sedig, Paul Parsons, and Alex Babanski. Towards a characterization of interactivity in visual analytics. *Journal of Multimedia Processing and Technologies*, 3(1):12–28, 2012.

[133] Kamran Sedig, Paul Parsons, Mark Dittmer, and Robert Haworth. Human-centered interactivity of visualization tools: Micro- and macro-level considerations. In *Handbook of Human Centric Visualization*, pages 717–743. Springer, 2014.

[134] Michael Sedlmair, Miriah D. Meyer, and Tamara Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.

[135] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *the IEEE Symposium on Visual Languages*, pages 336–343. IEEE Computer Society, 1996.

[136] Ben Shneiderman. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6):495–504, 2020.

[137] Yedendra Babu Shrinivasan, David Gotz, and Jie Lu. Connecting the dots in visual analysis. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 123–130. IEEE Computer Society, 2009.

[138] Yedendra Babu Shrinivasan and Jarke J. van Wijk. Supporting the analytical reasoning process in information visualization. In *the CHI Conference on Human Factors in Computing Systems*, pages 1237–1246. ACM, 2008.

[139] Yedendra Babu Shrinivasan and Jarke J. van Wijk. Supporting exploration awareness in information visualization. *IEEE Computer Graphics and Applications*, 29(5):34–43, 2009.

[140] Arjun Srinivasan, Steven Mark Drucker, Alex Endert, and John T. Stasko. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):672–681, 2019.

[141] John T. Stasko, Carsten Görg, and Zhicheng Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[142] Holger Stitz, Samuel Gratzl, Harald Piringer, Thomas Zichner, and Marc Streit. KnowledgePearls: Provenance-based visualization retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):120–130, 2019.

[143] Jing Tang, Zia ur Rehman Tanoli, Balaguru Ravikumar, Zaid Alam, Anni Rebane, Markus Vähä-Koskela, Gopal Peddinti, Arjan J. van Adrichem, Janica Wakkinen, Alok Jaiswal, Ella Karjalainen, Prson Gautam, Liye He, Elina Parri, Suleiman Khan, Abhishekh Gupta, Mehreen Ali, Laxman Yetukuri, Anna-Lena Gustavsson, Brinton Seashore-Ludlow, Anne Hersey, Andrew R. Leach, John P. Overington, Gretchen Repasky, Krister Wennerberg, and Tero Aittokallio. Drug Target Commons: A community effort to build a consensus knowledge base for drug-target interactions. *Cell Chemical Biology*, 25:224–229, 2018.

[144] Maria Taramigkou, Dimitris Apostolou, and Gregoris Mentzas. Leveraging exploratory search with personality traits and interactional context. *Information Processing and Management*, 54(4):609–629, 2018.

[145] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. The bohemian bookshelf: Supporting serendipitous book discoveries through information visualization. In *the CHI Conference on Human Factors in Computing Systems*, pages 1461–1470. ACM, 2012.

[146] Christian Tominski. *Interaction for Visualization*. Synthesis Lectures on Visualization. Morgan & Claypool Publishers, 2015.

[147] John W. Tukey. *Exploratory data analysis*. Addison-Wesley series in behavioral science: quantitative methods. Addison-Wesley, 1977.

[148] John W Tukey. We need both exploratory and confirmatory. *The American Statistician*, 34(1):23–25, 1980.

[149] Fernanda B. Viégas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matthew M. McKeon. ManyEyes: A site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, 2007.

[150] Jamie Waese, Jim Fan, Asher Pasha, Hans Yu, Geoffrey Fucile, Ruian Shi, Matthew Cumming, Lawrence A Kelley, Michael J Sternberg, Vivek Krishnakumar, et al. ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell*, 29(8):1806–1821, 2017.

[151] Jagoda Walny, Christian Frisson, Mieka West, Doris Kosminsky, Søren Knudsen, Sheelagh Carpendale, and Wesley Willett. Data changes everything: Challenges and opportunities in data visualization design handoff. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):12–22, 2020.

[152] Yun Wang, Zhida Sun, Haidong Zhang, Weiwei Cui, Ke Xu, Xiaojuan Ma, and Dongmei Zhang. DataShot: Automatic generation of fact sheets from tabular data. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):895–905, 2020.

[153] R. Wesslen, S. Santhanam, A. Karduni, I. Cho, S. Shaikh, and W. Dou. Investigating effects of visual anchors on decision-making about misinformation. *Computer Graphics Forum*, 38(3):161–171, 2019.

[154] Ryen W. White and Resa A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan & Claypool Publishers, 2009.

[155] Wesley Willett, Shiry Ginosar, Avital Steinitz, Björn Hartmann, and Maneesh Agrawala. Identifying redundancy and exposing provenance in crowdsourced data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2198–2206, 2013.

[156] Wesley Willett, Jeffrey Heer, Joseph M. Hellerstein, and Maneesh Agrawala. CommentSpace: structured support for collaborative visual analysis. In *the CHI Conference on Human Factors in Computing Systems*, pages 3131–3140. ACM, 2011.

[157] Kanit Wongsuphasawat, Yang Liu, and Jeffrey Heer. Goals, process, and challenges of exploratory data analysis: An interview study. *CoRR*, abs/1911.00568, 2019.

[158] Kai Xu, Simon Attfield, T. J. Jankun-Kelly, Ashley Wheat, Phong H. Nguyen, and Nallini Selvaraj. Analytic provenance for sensemaking: A research agenda. *IEEE Computer Graphics and Applications*, 35(3):56–64, 2015.

[159] Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John E. Wenskovitch. Survey on the analysis of user interactions and visualization provenance. *Computer Graphics Forum*, 39(3):757–783, 2020.

[160] Mehmet Adil Yalçin, Niklas Elmqvist, and Benjamin B. Bederson. Keshif: Rapid and expressive tabular data exploration for novices. *IEEE Transactions on Visualization and Computer Graphics*, 24(8):2339–2352, 2018.

[161] Jing Nathan Yan, Ziwei Gu, and Jeffrey M. Rzeszotarski. Tessera: Discretizing data analysis workflows on a task level. In *the CHI Conference on Human Factors in Computing Systems*, pages 20:1–20:15. ACM, 2021.

[162] Ji Soo Yi, Youn ah Kang, John T. Stasko, and Julie A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1224–1231, 2007.

[163] Ji Soo Yi, Youn-ah Kang, John T. Stasko, and Julie A. Jacko. Understanding and characterizing insights: How do people gain insights using information visualization? In *the Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. ACM, 2008.

[164] Emanuel Zgraggen, Zheguang Zhao, Robert C. Zeleznik, and Tim Kraska. Investigating the effect of the multiple comparisons problem in visual analysis. In *the CHI Conference on Human Factors in Computing Systems*, pages 1–12. ACM, 2018.

[165] Jian Zhao, Michael Glueck, Simon Breslav, Fanny Chevalier, and Azam Khan. Annotation graphs: A graph-based visualization for meta-analysis of data based on user-authored annotations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):261–270, 2017.

[166] Jian Zhao, Michael Glueck, Petra Isenberg, Fanny Chevalier, and Azam Khan. Supporting handoff in asynchronous collaborative sensemaking using knowledge-transfer graphs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):340–350, 2018.

[167] Zhilan Zhou, Ximing Wen, Yue Wang, and David Gotz. Modeling and leveraging analytic focus during exploratory visual analysis. In *the CHI Conference on Human Factors in Computing Systems*, pages 21:1–21:15. ACM, 2021.