A Grammar-Based
Approach for Applying Visualization Taxonomies to Interaction Logs

Sneha Gathani[1], Shayan Monadjemi[2] , Alvitta Ottley[2] , Leilani Battle[3]

[1]University of Maryland, College Park, MD
[2] Washington University in St. Louis, St. Louis, MO
[3] University of Washington, Seattle, WA

Abstract

*Researchers collect large amounts of user interaction data with the goal of mapping user's workflows and behaviors to their high-level motivations, intuitions, and goals. Although the visual analytics community has proposed numerous taxonomies to facilitate this mapping process, no formal methods exist for systematically applying these existing theories to user interaction logs. This paper seeks to bridge the gap between visualization task taxonomies and interaction log data by making the taxonomies more actionable for interaction log analysis. To achieve this, we leverage structural parallels between how people express themselves through interactions and language by reformulating existing theories as **regular grammars**. We represent interactions as **terminals** within a regular grammar, similar to the role of individual words in a language, and patterns of interactions or **non-terminals** as **regular expressions** over these terminals to capture common language patterns. To demonstrate our approach, we generate regular grammars for seven existing visualization taxonomies and develop code to apply them to three public interaction log datasets. In analyzing these regular grammars, we find that the taxonomies at the low-level (i.e., terminals) show mixed results in expressing multiple interaction log datasets, and taxonomies at the high-level (i.e., regular expressions) have limited expressiveness, due to primarily two challenges: inconsistencies in interaction log dataset granularity and structure, and under-expressiveness of certain terminals. Based on our findings, we suggest new research directions for the visualization community to augment existing taxonomies, develop new ones, and build better interaction log recording processes to facilitate the data-driven development of user behavior taxonomies.*

CCS Concepts

*• **Theory of computation** → **Regular languages;** Algebraic language theory; • **Human-centered computing** → **Visualization theory, concepts and paradigms;***

1 Introduction

A clear understanding of the user's visual analytic process is critical for designing and evaluating visualization systems. To this end, the visualization community has comprehensively captured their knowledge of user's visual analytic processes via multiple theoretical frameworks, typologies, and taxonomies [AES05; GZ09; BM13]. We refer to these kinds of structures as *taxonomies* in this paper. In parallel, researchers are collecting more and more interaction log data to learn how humans analyze information via visualization systems in more data-driven ways [HMSA08; XOW*20; CGL20; PW18]. The interaction log data can reveal the user's sensemaking process, analytical strategies, and reasoning behavior empirically much like taxonomies have aimed to capture them theoretically. By enabling data-driven approaches to testing, validating, and extending longstanding theoretical taxonomies in the visualization community.

However, taxonomies generalize our understanding of user analysis behavior as *high-level* user goals and strategies, whereas interaction logs aim to capture *low-level* actions and system events. As a result, inferring high-level goals and analysis strategies from interaction log data often requires an explicit mapping between lower-level interactions captured with the visualization interface and a model of the user's task. One solution from the literature is to manually

define user tasks based on the data and visualization system design. For example, Cook et al. [CCI*15] defined tasks models such as *InvestigateCrime* and *InvestigateSuspectsBehavior* to map low-level data interactions to potential high-level goals. This formulation enabled them to create a mixed-initiative system that infers the user's task as it evolves throughout their analysis and provides suggestions to aid the process. Similarly, Heer et al. [HMSA08] and Battle and Heer [BH19] categorized their observed actions into task types such as `analysis-filter`, `undo`, `navigate`, as well as interface specific actions like `shelf-add`, `show-me` and `worksheet-add`. Customized categorizations can help researchers reveal patterns in user's analysis strategies with specific systems, but fail to generalize to other visual interfaces [PW18].

Other works have leveraged existing visualization task taxonomies to systematize the analysis of collected interaction logs. For example, Pohl et al. [PWM*12], Torsney et al. [TSM17] and Guo et al. [GGZL15] demonstrate the potential of using theoretical taxonomies by mapping interaction logs they gathered to pre-defined task categories. Their process first involved selecting the most appropriate theoretical taxonomy for their collected interaction data and then transforming the taxonomy into an actionable task model encoding. All three mapped their application- and task-specific

actions to the same set of abstract analytic activities proposed by Yi et al. [YaKSJ07] — select, explore, reconfigure, encode, abstract/elaborate, filter, and connect. Similarly, Kahng et al. [KC20] characterize their collected interaction logs using Gotz & Zhou's [GZ09] taxonomy. Although visualization taxonomies were not developed for the purpose of analyzing interaction log data, these mappings still enabled researchers to verify and compare log analysis methods and determine further stages of system design. Thus, existing mappings from logs to taxonomies point to an exciting opportunity to *generalize* the log analysis process. However, key decision points within this mapping process are still unclear, such as which taxonomies to use or how to translate mappings for one set of system logs to another.

In this paper, we seek to bridge the gap between high-level visualization taxonomies and low-level interaction logs programmatically. In this way, we aim to extend the applicability of taxonomies by making them more actionable on interaction log datasets collected from real-world visualization systems. Our approach draws parallels between how users express themselves through interactions with analytic systems and how they express themselves via natural language by reformulating existing visualization taxonomies as *regular grammars*. We represent user's recorded interactions as *terminals* within a regular grammar, similar to the role of individual words in a language, and common sequences of user interactions as *non-terminals*, which are defined as regular expressions over the terminals. We demonstrate the viability of our approach by generating regular grammars for seven well-known visualization taxonomies (Amar et al. [AES05], Brehmer & Munzner [BM13], Gotz & Zhou [GZ09], Yi et al. [YaKSJ07], Guo et al. [GGZL15], Shneiderman [Shn96], and Gotz & Wen [GW09]) and developing code to apply them to three publicly available interaction log datasets (Battle & Heer [BH19], Liu & Heer [LH14], and Wall [Wal20]). All of our code is available in our supplemental materials[†] and provides a starting point for future works to apply the process to other taxonomies and interaction logs.

To demonstrate the utility of our grammars, we take first steps in using them to analyze the corresponding taxonomies. Specifically, we explore two new measures for analyzing the expressiveness of visualization taxonomies: *coverage* and *diversity*. The coverage-based measure captures the fraction of interaction log events that can be successfully mapped into a given taxonomy. The diversity-based measure examines the frequency and variety of symbols observed after mapping the taxonomy to interaction logs. Understanding such measures for various taxonomies could enable researchers to choose suitable taxonomies for their log analyses.

In analyzing the coverage and diversity of selected taxonomies, we find that the terminal-level taxonomies (low-level) show good coverage-based results over the selected interaction log datasets but are skewed in terms of the diversity-based results, resulting in mixed overall expressiveness and the non-terminal-level taxonomies (high-level) have limited expressiveness, primarily due to two challenges: *inconsistencies in log granularity and structure* and *under-expressiveness of specific terminals*. We use our initial measures of expressiveness to highlight the strengths of using existing taxonomies for interaction log analysis and the limitations of their use. Furthermore, we discuss how this framework can be used to come up with new analysis measures, enrich theoretical taxonomies

and build better interaction logging mechanisms to facilitate actionable and data-driven development of user behavior models.

In summary, we make the following contributions -

- We **reformulate theoretical taxonomies as regular grammars** to make them more actionable for interaction log analysis.

- We demonstrate our approach by **generating regular grammars for seven taxonomies** and developing code to apply them to three real-world interaction log datasets.

- Based on an analysis of our derived grammars, we **suggest new research directions** for augmenting existing taxonomies, developing new ones, and generalizing log recording processes to facilitate data-driven development of user behavior models.

2 Background

Although taxonomies are designed to characterize the space of user interactions, they are increasingly used for log data analysis, reinforcing the importance of taxonomies in designing and evaluating visualization tools [KK17]. Below we summarize the literature on visualization taxonomies and their use in log data analysis.

2.1 Visualization Taxonomies

Researchers generally view user interactions with visualization systems as a hierarchical construct with multiple levels of granularity, as shown in Figure 1. These granularities are often derived in a bottom-up manner and categorized into four levels, as summarized by Gotz & Zhou [GZ09]: individual user interactions (e.g., [SMG*20]), sequences of user interactions (e.g., [GGZL15]), user tasks (e.g., [BH19]), and high-level reasoning constructs or goals (e.g., [BOZ*14; LTM17]). There exist many visualization taxonomies and applications of these taxonomies at each of these four levels of granularity, which we describe below.

Individual Interactions. Taxonomies at the lowest-level categorize the user's most primitive interactions. Example taxonomies include those proposed by Amar et al. [AES05], Yi et al. [YaKSJ07] and Brehmer & Munzner [BM13]. Amar et al. proposed ten categories of interactions derived from explicit questions asked by students as they visually explored various datasets [AES05]. Yi et al. clustered the interaction capabilities of visualization tools into seven categories of interaction [YaKSJ07]. Brehmer & Munzner used a similar approach to derive a multi-level typology for user interaction [BM13], categorizing not only low-level interactions, but also the user's motivations for performing these interactions, such as to present information or to discover new hypotheses.

Sequences of Interactions. Low-level interactions are often chained together into sequences or patterns and examples of such taxonomies include ones proposed by Shneiderman [Shn96],
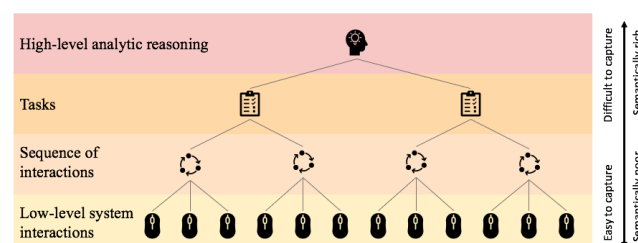

[†] https://tinyurl.com/regular-grammar-taxonomies

Figure 1: *User interaction log data is viewed as a hierarchical construct similar to the visualization analytic activity or taxonomy structure proposed by Gotz & Zhou [GZ09].*

Grammel et al. [GTS10], and Guo et al. [GGZL15]. Shneiderman's information-seeking mantra: "overview first, zoom and filter, then details on demand" describes an explicit progression of interactions performed when visually exploring data [Shn96]. Grammel et al. conducted a user study to understand how novices construct visualizations in Tableau [GTS10] and used their collected data to tabulate transitions made between attribute and encoding selection interactions. Guo et al. build on their analysis to identify sequences of interactions that frequently lead to insights in systems [GGZL15].

Tasks. Related interaction sequences can be clustered together to infer the intent of a user's analysis, often referred to as tasks. Notable examples at the task level include the taxonomies proposed by Pirolli and Card [PC05], Battle and Heer [BH19], Kang et al. [KGS09], Sedig and Parsons [SP13], and Alspaugh et al. [AZL*18]. Pirolli and Card propose a pipeline of data analysis tasks encompassed within two high-level loops, foraging and sensemaking [PC05]. Battle and Heer [BH19] survey the literature to identify common tasks completed during visual exploration. Kang et al. [KS11], Kandel et al. [KPHH12], and Alspaugh et al. [AZL*18] interview industry professionals to summarize common steps and challenges in the data analysis process. Sedig and Parsons [SP13] use a set of tasks which are broader patterns to characterize user's mental cognitive process. Yan et al. segment sequential event logs which combines data, interaction, and user features into high-level user tasks [YGR21]. Other works such as [Hib99] characterize specific non-exploration tasks for information visualization of real-world data. Further, [SNHS13] characterize broad visualization tasks based on their roles such as developers, authors and end users. However, although tasks are richer semantically, they often require in-depth and arguably laborious analysis of the underlying log data to extract meaningful user activities [YGR21].

Goals and Reasoning. At the highest-level of the hierarchy, taxonomies aim to capture how the users *organize* their analysis process into tasks and broader analysis goals. For example, Karer et al. build a formal rule-based model to reason about creation of a visualization and represent the analyst's information and knowledge flow graphically [KSHL20]. Lam et al. [LTM17] survey design study papers to understand how user's goals are broken down into actionable analysis tasks within visualization tools. Gotz et al. [GZA06] and Shrivnivasan et al. [SvW08] represent analyst's mental models as links or cycles between insight discovery and knowledge understanding when tracking a user's interactions with their tools. Sedig and Parsons [SP13] aim to speak to the user's cognitive processes via the use a patterns characterizing patterns in their analysis with visualization tools. However, we observe very few works that develop models of the user's high-level analysis intents. We believe this stems from the challenge of modeling high-level constructs in general. Tasks and analytic reasoning structures at the highest level are difficult to capture but semantically rich as they give insight into human analytic process, while primitive interactions captured at the lowest level of the hierarchy are easier to capture but semantically poor as they provide few details about the human analytic process.

2.2 Previous Use of Taxonomies for Interaction Log Analysis
The literature shows that many works have used taxonomies to analyze interaction logs. For example, Guo et al. manually apply the taxonomy proposed by Yi et al. [YaKSJ07] to analyze user interaction logs from a text document exploration tool [GGZL15]. Similarly, Satyanarayan et al. [SMWH17] also use Yi et al. [YaKSJ07] to define interaction categories to evaluate Vega-Lite. Battle and Heer [BH19] use a similar approach to map analyst's interactions with Tableau to

the Tableau-focused taxonomy of Heer et al. [HMSA08]. Likewise, Gotz & Wen use the interaction taxonomy proposed by Gotz & Zhou [GZ09] as a means for extracting common sequences of visualization interactions [GW09]. Battle and Scheidegger [BS21] use Brehmer & Munzner [BM13] to guide literature review on capturing distinctions in how data management technology can be applied in interactive analysis systems. Further, Battle et al. use Pirolli and Card's sensemaking loop taxonomy [PC05] and Shneiderman's information-seeking mantra [Shn96] to distinguish common interaction sequences as users visually explore massive array data [BCS16]. In all cases, these applications focus less on using taxonomies for their intended use as *design guidelines* and more on the unintended use of *evaluation*. Only a few interaction-focused taxonomies are designed to accommodate the complex tasks observed in visual analytics [LFB*14]. Further, all these taxonomies still need to be *manually* applied to individual log records, which can be a tedious process. Our goal is to work towards automation by formalizing the process of applying taxonomies to interaction logs.

2.3 Grammar-Based Approaches to Modeling User Behavior
Although few, there have been some interesting works in the visualization community that have taken a language-based or grammar-based perspective to understand user behavior and analytic activity. For instance, several works use Markov models [OGW19; BCS16; RJPL16] and finite automata [DC16] to infer user's common analysis and exploration behaviors. Dabek et al. in particular derive a grammar-based model to learn user interactions and determine common patterns for guiding new users for their visual analytic process [DC16]. Expressing taxonomies as formal grammars can enable researchers to express theories of user analytic activity *using a single, consistent language* thereby encouraging a formal means to analyze taxonomies, compare them, and either refine existing taxonomies or derive new ones. Our goal is to bridge the gap between theoretical taxonomies and data-driven analysis of visualization tools. In addition to formalizing the process of mapping taxonomies to interaction logs, we aim to enable the generation of data-driven guidelines for the design of future taxonomies *and* visualization tools.

3 Visualization Taxonomies as Regular Grammars
Inspired by ideas from linguistics and theory of computation, we leverage structural parallels between how people express themselves through interactions with analytic systems and language structure. We reformulate the hierarchical structure of visualization taxonomies (Figure 1) as *regular grammars*. Low-level user interactions can be represented as individual words in a sentence, i.e., *terminal* symbols within regular grammars. More complex user behaviors captured in higher levels of the hierarchy (sequences, tasks, goals) can be viewed as *non-terminal* symbols within the regular grammar. We formulate non-terminals as *regular expressions* comprised of terminals and/or other non-terminals. These regular expressions are synonymous with *production rules* which are functions defined over the same terminal and/or non-terminal symbols. In this section, we motivate our approach through a concrete example of generating a regular grammar for a well-known taxonomy.

3.1 Defining a Regular Grammar for a Taxonomy
We define a taxonomy t as being at either the low level, i.e., terminal level (T), or high level, i.e., non-terminal level (NT): $t \in \{T, NT\}$. We use t to define a regular grammar, which consists of three parts:

- a set of terminal symbols Σ,
- a set of non-terminal symbols N, and
- a set of production rules

– Each production rule f is a function over terminal and/or non-terminal symbols $f : N \rightarrow \{\Sigma \cup N\}^*$, where $\cup$ denotes the union of two symbol sets and $*$ denotes repetition of items.

The terminal symbols, Σ, are the most primitive building blocks in a grammar, i.e., the lowest level of the visualization taxonomy hierarchy in Figure 1. Therefore, the taxonomies at the lowest level (T) can be seen as Σ. In the context of interaction log data, the set of distinct user interactions captured represent the primitives to be mapped to a target set of terminal symbols, similar to mapping observed words to a target dictionary.

The non-terminal symbols, N, capture the syntax of a grammar. These symbols resemble more complex user analysis behavior captured at higher levels of the visualization hierarchy. For example, recurring patterns or sequences of user interactions within log data can be represented as functions over terminal and non-terminal symbols, akin to deriving common sentence structures from observed word sequences. Thus taxonomies at the higher levels (NT) can be represented as N. Note that non-terminals are not limited to expressing sequences of terminals and in fact can express *all levels* of the hierarchy, which we modulate through production rules. In the simplest case, non-terminals can be defined as a mapping to a single terminal, i.e., mapping a single log record to a taxonomy category. In the most complex case, non-terminals can be defined as a function of other non-terminal symbols, i.e., a function of functions. In this way, we can leverage the recursive power of regular grammars to express user analysis behavior at multiple levels of granularity.

3.2 An Example of Generating a Regular Grammar

Here we walk through our approach of generating a regular grammar for two well-known visualization taxonomies: *Brehmer & Munzner's multi-level typology [BM13] (BM)* and *Shneiderman's information-seeking mantra [Shn96] (S)*. Note that we focus on the *how* level of *BM*. We also demonstrate its application on an interaction log dataset collected by Wall [Wal20]. This dataset captures user interactions from a visualization system intended to select a committee of politicians, enabling further study of the user's potential biases in decision making. We represent our approach in Figure 2.

Brehmer & Munzner classify individual interactions into 11 categories, which we represent as a set of terminal symbols Σ_{BM}:

$\Sigma_{BM} = \{$encode,select,navigate,arrange,change,filter,

aggregate,annotate,import,derive,record$\}$

Note that we can represent any low-level visualization taxonomy as a set of terminal symbols in a similar fashion.

The Wall dataset has 11 distinct log record categories representing individual interaction types, all of which can be represented using an equivalent terminal from Σ_{BM}. As shown in Figure 2(1), examples of these log records include mouseover_from_list, change_attribute_distribution, filter_changed, etc., which allow users to retrieve details for specific politicians within the current committee list, change the rendered distribution measures computed over the politician attributes, and change the politician filtering criteria, respectively. These distinct log records can be mapped to corresponding terminals in Σ_{BM} using production rules defined over the Wall dataset $D_W \mapsto \Sigma_{BM}$, which we label as wall2020-brehmermunzner2013-mapping (Figure 2(2)). For example, we can use these functions to map the individual log records listed earlier to select, aggregate, and filter terminals respectively in Σ_{BM} (Figure 2(3)).
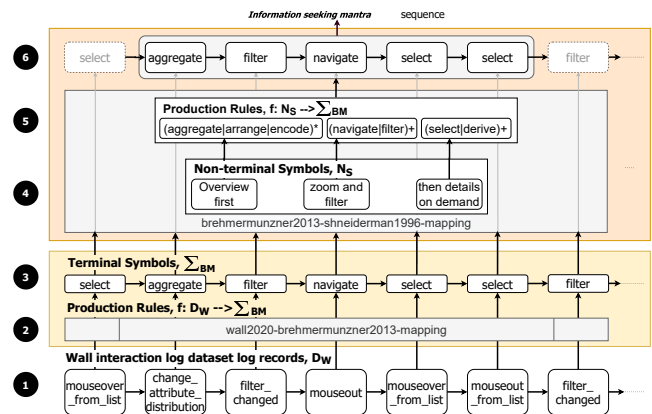

Figure 2: *Example grammars for the Brehmer & Munzner [BM13] taxonomy and Shneiderman's information-seeking mantra [Shn96] applied to the Wall [Wal20] interaction log dataset.*

High-level taxonomies such as Shneiderman's information-seeking mantra ("overview first, zoom and filter, then details-on-demand" [Shn96]) express common sequences of user interactions observed during visual exploration. The components of this mantra ("overview", "zoom", "filter" and "details-on-demand") can together be represented as a set of non-terminal symbols N_S (see Figure 2(4)):

$N_S = \{$overview, zoom, filter, details_on_demand$\}$

Other high-level taxonomies can also be represented using non-terminal symbols in a similar fashion.

Each non-terminal in N_S can be defined as a function of one or more Brehmer & Munzner terminals: $N_S \mapsto \Sigma_{BM}$ (Figure 2(5)), with left-hand side showing non-terminal symbol and right-hand side showing regular expression of possible terminal symbols. For example, the overview non-terminal symbol occurs when users transform the data, arrange the data differently or visualize the data in various ways. These transformations can be represented using the aggregate, arrange or encode terminals in Σ_{BM}. The corresponding production rule is as follows, defined as a regular expression:

$$\text{overview} \rightarrow (\text{aggregate}|\text{arrange}|\text{encode})^*$$

Similarly, the zoom, filter and details-on-demand non-terminal symbols can be defined as regular expressions over Σ_{BM}:

$$\text{zoom} \rightarrow (\text{navigate})^+$$

$$\text{filter} \rightarrow (\text{filter})^+$$

$$\text{details\_on\_demand} \rightarrow (\text{select}|\text{derive})^+$$

We label these production rules as brehmermunzner2013-shneiderman1996-mapping. Similar production rules can be generated for the same non-terminal using different underlying terminals, e.g., terminals defined by Yi et al. [YaKSJ07] instead of Brehmer & Munzner [BM13].

Since production rules can be represented as regular expressions, we can easily apply them to user interaction sequences within logged analysis sessions to determine whether relevant patters arise within this data. For example, Figure 2(6) shows one occurrence of the information-seeking mantra in the user session data. Similarly, other non-terminals represented by their corresponding production rules can be used to examine patterns in user interaction log data.

4 Mapping Taxonomy Grammars to Interaction Logs

We use best practices in qualitative coding to first map low-level interaction log records to the low-level taxonomy categories which form

Table 1: *Statistics characterizing the three selected interaction log datasets (*Battle & Heer [BH19], Liu & Heer [LH14] *and* Wall [Wal20]*).*

Interaction Log Datasets	Datasets	Total Tasks for each Participant	No. of Participants	Min Interactions	Max Interactions	Mean of Interactions	Median of Interactions
Battle & Heer	flight performance		15	39	139	78.6	84
	wildlife strikes	4	17	46	122	87.71	89
	weather		16	46	123	85.19	84
Liu & Heer	brightkite	1	16	8353	52266	36236.19	38807
	flight performance		16	11181	42496	22969.25	21271.5
Wall	politicians	1	24	122	783	370.5	331.5

the terminal symbols–Σ. Then, the high-level sequences or patterns of user's behaviors established in taxonomies are represented as non-terminal symbols–N. Then, functions over these non-terminals are developed as production rules of combinations of simpler terminal symbols. Finally, the production rules which form regular expressions are used to observe more meaningful user sequences and patterns in the interaction log data. In order to develop our approach, we select few representative analytic visualization taxonomies and interaction log datasets which we describe first before elaborating on our approach.

4.1 Representative Analytic Visualization Taxonomies

We select **seven** taxonomies: *four* taxonomies occurring at the lower granularity of the hierarchical structure and *three* taxonomies occurring at the higher granularity as representative taxonomies on which we demonstrate the generation of regular grammar. Although we select only a subset of taxonomies here, we note that the same process can be applied directly to other taxonomies. The representative taxonomies are selected based on two factors. First, we look for the implementation and application of the taxonomies in application systems to demonstrate their empirical use. We use the number of citations measure, with a minimum threshold of 300 as a quantitative measure of the widespreadness of the taxonomies. Second, we examine the feasibility of taxonomies such that they have enough detail to be mapped to interaction log data in order to successfully generate a regular grammar for it. We observe the number of systems that apply or use these taxonomies for either their design, analysis (e.g., GAN Lab [KC20]) or evaluation (e.g., Sliceplorer [TSM17]) as a measure of feasibility of the taxonomies. In accordance to these factors, the four representative low-level taxonomies selected are:

[T1] **Amar et al.'s** [AES05] analytic task taxonomy.

[T2] **Brehmer & Munzner's** [BM13] multi-level typology of visualization tasks.

[T3] **Gotz & Zhou's** [GZ09] characterization of visual activity.

[T4] **Yi et al.'s** [YaKSJ07] interaction technique category.

And, the three representative high-level taxonomies selected are:

[NT1] **Shneiderman's** [Shn96] information-seeking mantra.

[NT2] **Gotz & Wen's** [GW09] behavior patterns to infer user's intended visual task.

[NT3] **Guo et al.'s** [GGZL15] common analysis patterns.

4.2 Representative Interaction Log Datasets

We performed a search for available and pre-collected interaction log datasets online as well as reached out to our network. Following the retrieval of interaction log datasets, we specified three criteria to short-list usable datasets. First, the interaction log datasets captured needed to be of visualization-based systems. Second, the interaction log datasets needed to capture the lowest level of user interactions

with visualization systems and last, the systems needed to have features that led to interaction log records mapping to at least 60% distinct categories of the representative taxonomies. Some interaction log datasets that were ruled out due to either their unavailability or limited functionalities of the tools were Patterns and Pace dataset [FPH18], HindSight dataset [FDPH16], Anchoring Effect dataset [CWK*17], etc. We found three interaction log datasets that matched our criteria, encouraging us to apply our regular grammar's approach to them. However, our process can also be applied to other interaction log datasets as well. We list our representative datasets here, describe their tasks and provide statistics about them in Table 1.

[D1] **Battle & Heer:** Data exploration using Tableau [BH19].

[D2] **Liu & Heer:** Big data exploration using imMens [LH14].

[D3] **Wall:** Visual analytics for decision making [Wal20].

4.3 Terminal Symbols (Σ)

The individual user interactions in an interaction log dataset can be synonymous to low-level taxonomy categories, that is terminal symbols (Σ).

Overview. Therefore, for a given interaction log dataset D and terminal symbol Σ, we can translate each log record $d \in D$ to its corresponding terminal symbol $i \in \Sigma$, producing the mapping $f : D \mapsto \Sigma$. For example, suppose the dragging of a slider is being mapped to the Brehmer & Munzner terminal, represented by Σ_{BM} (see subsection 3.2), f may map this recorded event to the filter terminal $\in \Sigma_{BM}$. We develop functions over Σ and the interaction log dataset D which we call *mappings*. These mappings are formulated as JSON objects, one for every interaction log dataset and low-level taxonomy that is a terminal symbol pair. Therefore, we develop a total of 12 unique mappings = 3 sets of interaction log datasets (D) $\times$ 4 sets of terminal taxonomies (Σ_T).

Process. We follow best practices in qualitative coding to derive our mappings. Similar to prior work [GTS10; LTM17], we used an iterative approach to establish a code-book of the mappings for all the terminal taxonomy and interaction log dataset pairs. To measure agreement among authors in applying the code-book, we calculate inter-rater reliability scores [Wan09], and find that all coders were consistently in close agreement, i.e., achieved scores of 0.99 out of 1.0.

In the first iteration, for each interaction log dataset, two researchers on the project individually used the elimination approach [Smi43] to map each distinct log record $d \in D$ to a single terminal of the low-level taxonomy, that is $i \in \Sigma$. This was repeated for all four terminal symbol taxonomies. The two researchers mapped an interaction log record to a special *null* terminal if they found multiple or no terminals from the taxonomy that mapped to the log record. At the end of the first iteration, the mappings of both the researchers were aligned to calculate an inter-coder reliability score of 0.47. In order to reconcile on the

code-book, the second iteration constituted of both the researchers discussing the conflicting mappings and reasoning on their choices. This discussion led to either a consensus on the mappings or finalizing the conflicts. The second iteration of forming the code-book resolved the majority of the conflicts and increased the inter-coder reliability score to 0.91. In the final round of iteration, the final conflicts were resolved by having the other two researchers on the project follow the same process which further raised the inter-coder reliability score to 0.99.

The special *null* terminal was used for interaction log records that could not be mapped to a single terminal even after discussions with all four researchers. Examples of such terminals are the log records for resetting of the interface, which are commonly observed in the Tableau tool and thus in Battle & Heer's [BH19] interaction log dataset. The Gotz & Zhou [GZ09] taxonomy does not include an interaction category that resets interfaces. Therefore, we assignment *null* to reset actions when mapping to the Gotz & Zhou terminal symbol Σ_{GZ}.

We use this final iterated and established code-book of mappings (inter-coder reliability agreement = 0.99 out of 1.0) to perform further analysis. A justification for each mapping is provided as an additional nested `description` property in the JSON object mappings explaining our reasoning process. Along with the mappings, we provide a Python script that takes as input a interaction log dataset file (D) and the terminal symbol mapping (Σ) and outputs another file containing a list of corresponding terminal symbols, $i \in \Sigma$ of all log records of the interaction log dataset, $d \in D$.

4.4 Non-terminal Symbols (N)

The user patterns or sequences observed in interaction log datasets can be synonymous to high-level sequence taxonomy categories, that is non-terminal symbols (N). Therefore, we represent every high-level taxonomy describing user's sequences (NT) as a distinct set of non-terminal symbols, N_{NT}.

Overview. Each pattern or sequence in a high-level taxonomy can be represented as consecutive interactions of simpler symbols, which are terminals (Σ_T). Therefore, we simply degenerate each pattern into simpler terminals and assign them to be included in the set of non-terminal symbols. These degenerations are either explicitly provided by the researchers who find these patterns or are described by them in words. Thus, based on these explicit representations or descriptions, we come up with a set of consecutive simpler terminals, $t \in NT$ that express the non-terminal sequence. Therefore, we develop a total of 3 sets of non-terminal symbols (N_{NT}) each consisting of one or multiple sequences or non-terminals which are represented using terminal symbols (Σ_t).

Process. We follow the same iterative process to establish a code-book for the non-terminal symbols mappings as followed for the terminal symbols. In the first iteration, two researchers on the project understand the descriptions of the founder researchers coming up with the sequences to build the set of non-terminal symbols. This is repeated for all three sets of representative non-terminal taxonomies (N_{NT}). A full inter-coder reliability score of 1.0 was achieved after the first iteration, thereby, saving further discussion and iterations with the other two researchers. Once again, this iterated and established code-book of non-terminals mappings was used to perform further analysis.

4.5 Production Rules

Empirically in interaction log datasets, complex behaviors of users are perceived to be patterns of underlying simpler interactions. In regular grammar, this can be paralleled to forming sets of non-terminal symbols (N) using combinations of simpler sets of terminal symbols (Σ) put together. Even more complex sequences, which speak to the more abstract tasks of the user can be formulated as combinations of not only the sets of terminal symbols but also the sets of non-terminal symbols. Our approach adopts this approach by the ability to generate production rules over non-terminal symbols. Production rules can be functions or regular expressions of both terminal and non-terminal symbols to generate complex non-terminal symbols thus informing user's common behaviors as sequences or tasks in interaction log datasets.

Overview. Each non-terminal can be represented as a function of terminal symbols, Σ_T. Because each low-level taxonomy or terminal symbol has a unique mapping represented by Σ, the same high-level taxonomy may be generated using different terminals, depending on which underlying low-level terminal is used. Thus, we develop functions or mappings of all possible pairings of low-level taxonomy (i.e., terminals) and high-level taxonomy (i.e., non-terminals) listed in section 3. These mappings too are formulated as a JSON object, one for every terminal and non-terminal pair. Therefore, we develop a total of 12 unique mappings = 4 sets of terminal symbols (Σ_T) × 3 sets of non-terminal symbols (N_{NT}).

Process. We follow a similar iterative and best qualitative process to establish the code-book for the production rule mappings as followed for the sets of terminal and non-terminal symbols. The mappings for non-terminal symbols are produced by building functions of mapping each individual non-terminal of the non-terminal sequence to its corresponding terminal mapping. This is repeated for all non-terminal taxonomy and terminal-taxonomy pairs. After the first iteration, where two researchers on the project individually produced the mappings for the non-terminals, a inter-coder reliability score of 0.73 was achieved. After a second iteration of discussion and resolving conflicts, a full inter-coder reliability score of 1 was achieved, thereby, saving a third iteration with the other two researchers. Once again, this final iterated and established code-book of regular expressions mappings was used to perform further analysis.

For some non-terminals, additional information was needed about the attributes and dimensions on which interactions were captured. For example, two non-terminals observed by Gotz & Wen [GW09]: `scan` and `drill-down` mean that users continuously perform the `inspect` iterations over a series of similar (i.e., on the same dimension or attribute) and different (i.e., on different dimensions or attributes) visual data objects respectively. To meet these needs, we modified the underlying terminal symbols to include `inspectsame` and `inspectdifferent` for `scan` and `drill-down` respectively.

Similar to the previous mappings, regular expressions also use the special *null* category if the non-terminals cannot be represented as regular expressions using underlying terminals. For instance, `scan` and `drill-down` Gotz & Wen non-terminals cannot be represented with Amar et al. [AES05] terminal, since there is no `inspect`-like interaction in Amar et al. taxonomy. We also provide another Python script that takes as input the terminal mappings file (Σ) and non-terminal mappings file (N), and outputs another file containing a list of sequences found.

5 Analysis

A natural question stemming from this work is: what can we learn from translating taxonomies into regular grammars? To answer this question, we applied our derived grammars to three different interaction log datasets, and used the results to evaluate the

corresponding taxonomies. Our objective is to identify *measurable differences* between grammars that may reveal the *most suitable taxonomies* for a given analysis context.

Inspired by prior work in evaluating visual encoding grammars [SMWH17; Mac86], our analysis focuses on measuring the *expressiveness* of selected taxonomies. We propose two measures of expressiveness for visualization taxonomies, *coverage* and *diversity*. When choosing taxonomies for their log analyses, we find that researchers tended to favor taxonomies where all distinct log records could be mapped to a valid terminal, i.e., taxonomies that generate mappings with high coverage of all log records. In response, our proposed *coverage* measure calculates the fraction of interaction log records that can successfully be mapped to the symbols of a given taxonomy. Similarly, researchers also seemed to favor taxonomies that would avoid mapping many different log records to the same taxonomy category. For instance, if every log record maps to a single terminal, then it becomes impossible to extract meaningful interaction sequences or patterns. To this end, our *diversity* measure captures the frequency and variety of symbols observed after mapping a taxonomy to the interaction log data.

5.1 Analyzing the Expressiveness of Terminal Taxonomies

Before we can extract patterns from log data, we first need to map it to a relevant set of terminals (Σ). However, the expressiveness of a terminal-level taxonomy can influence our ability to extract patterns, such as by having low coverage that causes us to lose data records, or by having low diversity that causes us to lose semantic meaning across interaction sequences. To evaluate the expressiveness of terminal-level taxonomies, we analyze the coverage they provide when mapped to each of our three datasets, as well as the diversity of terminals observed across the resulting mappings. We evaluate the four different representative terminal-level taxonomies in this analysis.

5.1.1 Coverage-based Analysis

For each terminal-level taxonomy (see section 3), we map each distinct log record from our datasets to its corresponding terminal symbol. To measure coverage, we calculate the percentage of successfully mapped interaction log records or the percentage of "non-null" mappings for each interaction log dataset and taxonomy pairing, summarized in Table 2. We observe relatively high coverage for two out of three datasets: 100% coverage of the Wall dataset, 85-100% coverage of the Liu & Heer dataset, and 46-68% coverage of the Battle & Heer dataset. We find that the Brehmer & Munzner taxonomy [BM13] provides the best coverage across all three datasets, followed by the taxonomy of Gotz & Zhou [GZ09]. The Amar et al. taxonomy [AES05] provided the lowest coverage, 65.56% on average.

Table 2: *Average coverage percentage of the four sets of terminal symbols (Amar et al. [AES05],* Brehmer & Munzner *[BM13],* Gotz & Zhou *[GZ09],* Yi et al. *[YaKSJ07]) over three interaction log datasets (*Battle & Heer *[BH19],* Liu & Heer *[LH14],* Wall *[Wal20]).*

	Battle & Heer	Liu & Heer	Wall	Avg.
Amar et al.	46.67%	50%	100%	65.56%
Brehmer & Munzner	68.89%	100%	100%	**89.63%**
Gotz & Zhou	67.78%	100%	100%	89.26%
Yi et al.	74.44%	91.67%	100%	88.7%
Avg.	64.45%	85.42%	**100%**	

Number of terminals alone did not seem to be a strong predictor of coverage. For example, Brehmer & Munzner propose fewer terminals than Gotz & Zhou but the Brehmer & Munzner taxonomy provides (slightly) higher coverage. Amar et al. propose fewer terminals as well, but their taxonomy provides lower coverage. Instead, we see a consistent difference in coverage based on dataset, suggesting that for the taxonomies and datasets studied, coverage is inversely proportional to the complexity of the visualization systems used to capture the interaction logs. For instance, the Wall dataset was collected using an interface with fewer features, leading to high coverage, while the Battle & Heer dataset was captured using the feature-rich Tableau tool, resulting in the lowest observed coverage. Thus these taxonomies seem to cover simpler interfaces well but not necessarily more complex ones. We encourage further investigation in future work.

5.1.2 Diversity-based Analysis

To measure the diversity of our terminal-level taxonomies, we analyze the distribution of terminals observed for each interaction log dataset. Specifically, we evaluate the percentage of log records mapped to each terminal within a taxonomy to measure terminal utilization, as well as the fraction of the dataset mapped to the most popular terminal to gauge redundancies in the resulting mappings. We represent these distributions in Figure 3 as four stacked bar charts, one chart per taxonomy. Each stack represents a log dataset, and each bar an individual terminal symbol from the corresponding terminal taxonomy.

Terminal Under-Utilization. First, we consider under-utilization of terminals, which can help us understand which taxonomies may contain terminals that do not provide meaningful context for log analyses. We find that across the four representative terminal-level taxonomies we evaluated, only two had obvious under-utilized terminals. Across the representative three log datasets we used, the Amar et al. taxonomy had two terminals that were never used: *find anomalies* and *find extremum*. Similarly, across all three log datasets, the Gotz & Zhou taxonomy had three terminals that were never used: *restore*, *split*, and *bookmark*. Further, some system actions such as *formatting actions* (e.g., changing the size or color of the legends in the Tableau tool) and *setting default operations* (e.g., setting default color, size, etc.) were not represented across any of the taxonomies.

Terminal Over-Utilization. Next, we consider over-utilization of terminals, which can help us determine when we may be losing semantic meaning across consecutive interactions. For example, if three consecutive but different interactions all mapped to the exact same terminal, we effectively lose the meaning behind these interactions—and as a consequence, the sequences derived from these interactions—in later analyses. To measure this, we identify the most popular terminal for a given log dataset and taxonomy pairing and calculate the percentage of log records mapped to this terminal. For the Battle & Heer dataset, we find that the Brehmer & Munzner taxonomy had the lowest measure of 19.2% (for the *aggregate* terminal), and the Gotz & Zhou taxonomy had the highest measure of 49% (for the *remove* terminal). For the Liu & Heer dataset, the Amar et al. taxonomy had the lowest measure of 86.99% (for the *filter* terminal), and the Yi et al. taxonomy had the highest measure of 95.43% (for the *explore* terminal). For the Wall dataset, the Brehmer & Munzner taxonomy had the lowest measure of 47.86% (for the *navigate* terminal), whereas the Gotz & Zhou taxonomy and Yi et al. taxonomy tied for the highest measure of 81.61% (for the *inspect* and *explore* terminals, respectively). Overall, we found that the Brehmer & Munzner taxonomy provided the best measures. We also stress that although skew in itself is not inherently bad, extreme skew, e.g., a measure of 95.43%,

Figure 3: *Diversity percentage for each of the terminal symbols for the four low-level taxonomies (*Amar et al. [AES05]*, Brehmer & Munzner [BM13], Gotz & Zhou [GZ09] and Yi et al. [YaKSJ07]) for the three interaction log datasets (*Battle & Heer, Liu & Heer, Wall*) (top to bottom).*

means that the overwhelming majority of log records were mapped to the exact same terminal, which likely will not lead to a meaningful analysis of interaction sequences, since most interaction events will appear to be identical. If extreme skew is observed across different log datasets, then the corresponding taxonomy may not be suitable for log analysis. Further, some system actions are too broad and can override taxonomy categories. For example, the geographic filter, legend filter, filter by value, etc. in Tableau are all attributed to the *filter* taxonomy category. The Wall dataset also suggests that adding or removing politicians from committee via card click or scatterplot point click are all actions that are attributed to *select* taxonomy category.

We acknowledge that since none of these taxonomies were designed to analyze the specific logs we used, there are likely dataset effects that come to play in our analysis. For example, we found that terminal over-utilization was consistently high for all of our representative taxonomies when analyzing the Liu & Heer and Wall datasets, but not for the Battle & Heer dataset. These findings suggest that some log datasets may not be as complex as others, likely because the original interfaces themselves contained proportionally fewer features, e.g., the research prototypes developed by Liu & Heer and Wall versus Tableau, the tool used by Battle & Heer. That being said, we do still observe differences between taxonomies even when dataset differences are considered, with the Brehmer & Munzner taxonomy providing overall best over- and under-utilization results. From these skewed findings of terminals across the interaction log datasets,

we see that some taxonomies have poor diversity-based measures of expressiveness. With high coverage but low diversity, the terminal taxonomies we studied show mixed overall expressiveness when applied to our selected log data.

5.2 Analyzing Non-Terminal Taxonomies

Similar to the analysis performed for terminal symbols, the coverage-based analysis of non-terminal symbols is analyzed by calculating the number of occurrences of established sequences or non-terminals in the interaction log datasets. And, the diversity-based analysis of non-terminal symbols is informed by finding new sequences or new non-terminals that are observed within and across interaction log datasets.

5.2.1 Coverage-based Analysis

We calculate the coverage of non-terminal sequences observed in the interaction log datasets. The non-terminal symbols are sequences of interactions or patterns that recur within user analysis sessions, and can be represented using simpler terminal symbols. While
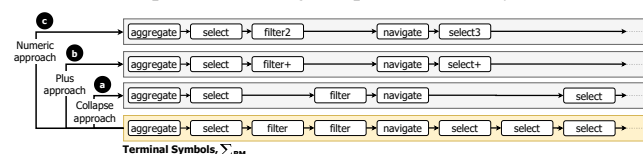

Figure 4: *Different approaches to reduce the interaction log records of the interaction log datasets for data-driven exploration of new sequences: (a) Collapse, (b) Plus, (c) Numeric.*

Table 3: *Average count of non-terminals for the three high-level taxonomies (*Guo et al., Shneiderman *and* Gotz & Wen*) observed **per participant's session** for three interaction log datasets (*Battle & Heer, Liu & Heer, Wall*) when retrieved with four underlying terminal symbols (*Amar et al. (A)*,* Brehmer & Munzner (BM)*,* Gotz & Gotz (GZ)*, and* Yi et al. (Y)*) with a **confidence interval of** ± 95%.*

Provenance Datasets	Interaction Taxonomy	Guo et al.								Shneiderman		Gotz & Wen							
		elaborating		locating		orienting		sampling		ISM		drill-down		swap		flip		scan	
		Count	±95%	Count	±95%	Count	±95%	Count	±95%	Count	±95%	Count	±95%	Count	±95%	Count	±95%	Count	±95%
Battle & Heer	A	0	0	0	0	0	0	0	0	0.04	0.58	0	0	1.12	0.41	3	0.62	0	0
	BM	0	0	0	0	0	0	0	0	0	0	5.75	0.65	1.27	0.58	5.88	0.87	5.19	0.94
	GZ	2.25	0.55	0.62	0.28	0	0	0.17	0.11	0.48	0.2	5.19	0.89	1.38	0.5	8.23	1.22	4.38	0.76
	Y	6.52	1.28	2.62	0.59	0	0	0.38	0.21	0	0	6.33	0.83	4.71	1.06	0	0	5.46	1.02
Liu & Heer	A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	BM	165.5	214.18	0	0	0	0	0	0	40.94	18.41	162.62	203.79	0	0	83.12	23.07	44	11.42
	GZ	252.19	230.16	22.75	5.57	0	0	5.06	2.18	87.38	36.75	152.56	53.19	0	0	141.06	48.81	124.88	38.32
	Y	71.56	21.89	1.38	0.65	0	0	0.81	0.55	0	0	286.56	270.48	0	0	0	0	125.19	32.81
Wall	A	0	0	0	0	0	0	0	0	23.86	4.07	0	0	0	0	7.5	1.9	0	0
	BM	22.58	3.27	0.79	0.26	0	0	5.79	1.72	21.42	3.09	22.21	4.28	0	0	10.04	3.07	19.29	3.36
	GZ	8.5	2.86	3.71	1.49	0	0	3.33	1.48	6.54	2.4	20.08	3.85	0	0	9.29	2.8	18.88	3.28
	Y	1.83	0.55	1.42	0.28	0.17	0.15	1.12	0.21	0.12	0.24	22.88	4.13	7.04	1.92	0	0	18.75	3.28

calculating the number of non-terminals in the interaction log datasets, we observed multiple user sessions having consecutive log records that mapped to the same terminals. Since, having a sequence of similar terminals makes the chain of user terminals mappings very long and verbose for analysis, we use the "collapse approach" to collapse consecutive similar terminal symbols to one, as shown in Figure 4 (a). The sequence of eight terminals are reduced to five terminals after applying the collapse approach on the *filter* and *select* repetitive terminals. We use the regular expressions generated for each of the non-terminal symbols to sum the number of sequences observed for each of the user sessions in the interaction log datasets. We report these in Table 3 where we represent the average number of non-terminal sequences observed per user session with a confidence interval of ±95%. For instance, we observe an average of 2.25±0.55 = 1.7 to 2.8, which is approximately 2 to 3 number of *elaborating* Guo et al. [GGZL15] non-terminal sequences that are expressed using the Gotz & Zhou (GZ) [GZ09] underlying terminals in one user session of the Battle & Heer [BH19] interaction log dataset.

We observe that the coverage of non-terminal symbols is less than that of the terminal symbols for our selected taxonomies and datasets. Of the representative non-terminal taxonomies, we observe the least coverage for the Guo et al. [GGZL15] non-terminals (maximum low counts of sequences), followed by Shneiderman's [Shn96] information-seeking mantra as we rarely observe any of the sequences in user sessions. On the other hand, the Gotz & Wen [GW09] non-terminal sequences seem to be more widely observed across user sessions. We attribute this behavior to the environments and tools used to come up with these non-terminals. For instance, since Guo et al. [GGZL15] realized the common user sequences only using a single and specific text analysis tool, we tend to not observe them in any other datasets. The Gotz & Wen non-terminals seem to be the most prevalent across the interaction log datasets, making them more generalizable for multi-system and multi-task purposes.

5.2.2 Diversity-based Analysis
The analysis uses our regular grammars approach to find new common behaviors of users within and across the interaction log datasets. We use terminal mappings of individual taxonomies and split them per user sessions. Then we find intersection of consecutive sequence of terminals to attain the common sequences within the user sessions of an interaction log dataset. Similar to the previous analysis where we used the "collapse" approach to reduce the vast amounts of log data collected, we use the "plus" and "numeric" approach for this analysis, as shown in Figure 4. In the plus approach, consecutive similar terminal symbols are collapsed to one terminal symbol concatenated with

a "plus" notation. For example, the consecutive *filter* and *select* terminals are replaced with *filter+* and *select+* terminals respectively, as seen in Figure 4. The numeric approach is similar to the plus approach but the plus notation is instead replaced with the number of times the terminal is repeated consecutively. Again as observed in Figure 4, the two consecutive *filter* terminals and three consecutive *select* terminals are replaced with *filter2* and *select3* respectively. We favor the plus and numeric approaches instead of the collapse approach for this analysis since the former approaches preserve the interactions of the users.

The new non-terminal sequences that we observed are shown in Table 4. For the Battle & Heer dataset, we do not observe any common terminal-level sequences across user sessions. For both the Liu & Heer and Wall datasets, we observe different common non-terminal sequences for different terminal symbols as seen in Table 4. Most of these sequences follow the same pattern of alternating between terminals and non-terminals. For instance, when using Amar et al. terminals, we observe that the Wall interaction log dataset alternates between multiple *retrieve-value* and a *filter* terminals followed by a *retrieve-value* terminal. A similar pattern occurs with multiple *brush* and a *delete* terminals followed by a *brush* terminal in the Liu & Heer dataset (with Brehmer & Munzner terminals). However, we see few exact pattern matches for the plus and numeric approaches, suggesting that even when users perform similar patterns, the number of interactions (or terminals) within these patterns often varies. We further extended this analysis to search for common sequences *across* all three datasets, but we did not find any universal patterns. Thus, the studied non-terminals appear to be less expressive.

6 Discussion
The terminal-level taxonomies we studied have mixed *expressiveness* as defined by the *coverage* and *diversity* of the mapped terminals. We observe high coverage for these taxonomies across all three log datasets, suggesting that the current set of low-level taxonomies in the literature provide sufficient coverage of log records within real-world interaction log datasets. However, we find that some of the representative terminal taxonomies tend to under- or over-utilize certain terminals, resulting in skewed distributions of emitted terminals within our mappings and thus limited diversity.

We believe the mixed results stem from the tension between optimizing for *specificity*, or ensuring specific user activities are represented, and *generality*, or designing terminals that can be applied to many tools, when designing taxonomies. On the one hand, our findings demonstrate the utility of popular theoretical taxonomies beyond their intended use as descriptive tools for designers. Our

Table 4: *New sequences observed using the* plus *and* numeric *approaches to analyzing the selected interaction log datasets.*

Interaction Log Data Terminal Symbol	Approach	Amar et al.	Brehmer & Munzner	Gotz & Zhou	Yi et al.
Battle & Heer	Plus	-	-	-	-
	Numeric	-	-	-	-
Liu & Heer	Plus		((brush+, delete)+, brush+)	((filter, navigate+)+, filter, navigate)	((filter, explore+)+, filter, explore)
	Numeric		(delete, brush)	((navigate, change2)19, navigate, change)	((connect, explore2)24, connect, explore)
Wall	Plus	((retrieve-value+, filter)+, retrieve-value)	(inspect+, annotate, inspect+)	(navigate+, select, navigate)	(explore+, select, explore+)
	Numeric	(filter, retrieve-value)	(change-metaphor)	(select, navigate)	(select, explore)

grammar-based approach reveals how taxonomies can be useful tools for analyzing interaction logs from a variety of systems. On the other hand, our approach demonstrates the limitations of taxonomies. For example, our results suggest that highly skewed taxonomies may be too general, sacrifice context, and as a result lead to mappings that may not be meaningful for interaction log analysis. In some cases, these taxonomies produced a single terminal for most (e.g., over 95% of) log records, resulting in homogeneous mappings. However, striving purely for specificity may also produce taxonomies with limited applicability, e.g., taxonomies that apply only to one tool. Our research highlights a potential need for more taxonomies that strike a balance between specificity and generality. It also suggests a potential direction for achieving this balance: augmenting taxonomies to include critical contextual cues as input to the underlying regular grammar; for example, details about the system wherein the interactions are being performed.

In contrast, we rarely observe the interaction patterns proposed in our selected non-terminal taxonomies, and the few we do observe are only a small fraction of the analyzed logs. Although these taxonomies do not directly match the interaction log datasets we analyzed, our regular grammars approach enables deducing new *data-driven taxonomies* at higher levels of user activity such as interaction sequences and analysis tasks. For example, our approach reveals common subsequences within interaction logs which represent more meaningful and complex patterns than those proposed in well-known taxonomies. Finally, we do not observe any common sequences that occur across all three log datasets. We believe these issues stem in part from a mismatch between popular non-terminal taxonomies and log recording strategies, as well as challenges originating from the terminal rather than non-terminal level: over- and under-utilization of certain terminals and lack of important contextual cues in these taxonomies.

However, only a subset of taxonomies and log datasets are analyzed in this work. We encourage the community to extend these ideas to see how they generalize to other taxonomies and log analysis contexts.

Implications for Log Data Analysis. There are two important concerns when applying taxonomies to interaction logs. First is the loss of information in the resulting mappings such as a lack of proper translation between the context or *semantics* of the user action and taxonomy category or *syntax*. Taxonomies are designed to concisely communicate the semantics of user interactions. However, a user's interaction intent is also influenced by the design of the underlying interface,

which is intentionally abstracted away from most taxonomies. As an example, consider the `filter` terminal from the Brehmer & Munzner taxonomy. While mapping the Battle & Heer interaction log dataset, we notice that *eight* of its distinct log records were mapped to the `filter` terminal, as shown in Figure 5. After investigating the underlying system details, we posit that these eight filters actually represent *three* types of filtering interactions, shown in in Figure 5: filter the data, filter the visualization, and "other" filter operations. These findings point to a need to *augment taxonomies* to include system level details to prevent losing user context. A second concern lies in the inability to express timing of interactions because of the use of grammar-based approach. However, to facilitate the development of new taxonomies, we need to ensure that log data collection processes *scale*, which is generally enabled by developing structured languages, as observed in other areas such as databases, distributed systems, and NLP.

Further, although many interaction logs are shared online, this is no guarantee that others will actually be able to use them. It is critical to create a community-wide process for sharing datasets that will be reusable [BAB*18]. Logs are often collected in an ad-hoc manner that is unique to the system being evaluated, making it difficult to translate these logs to a broad range of analysis contexts [PW18; CGL20]. Generalizable logging formats must be adopted to make future log datasets applicable to a wider range of analysis scenarios. We need a shift in how we think about data sharing and consider data useful only if others can and actually *use* it.

7 Conclusion

This paper presents a framework that bridges the gap between theoretical visualization task taxonomies and empirical analysis of interaction log data. to do this, we exploit structural parallels between how people express themselves through interactions and language by reformulating existing theories as regular grammars. We represent interactions as terminals within a regular grammar and patterns of interactions as regular expressions over these terminals to capture common language patterns. Regular grammars provide opportunities to express new taxonomies in exciting ways. For example, this formulation can enable future work to express new taxonomies as a mix of low-level and high-level attributes of the inherently hierarchical structure of visual analysis and human reasoning. Our contributions can help the community to create taxonomies that match a broader range of granularities in user intents and try to strike a subtler balance between capturing the coverage and diversity of interaction log events.
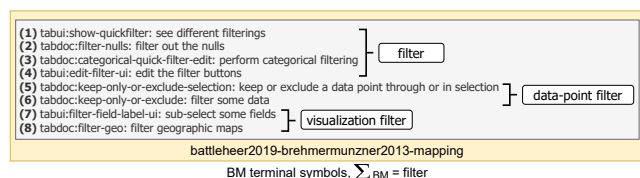
8 Acknowledgements

Figure 5: *Eight Tableau log records mapped to the same terminal in the Brehmer & Munzner taxonomy (*filter*), which could be represented as three terminals:* filter, *data-point filter, and* visualization filter.

References

[AES05] AMAR, ROBERT, EAGAN, JAMES, and STASKO, JOHN. "Low-level components of analytic activity in information visualization". *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. IEEE. 2005, 111–117. DOI: 10.1109/INFVIS.2005.1532136 1, 2, 5–8.

[AZL*18] ALSPAUGH, SARA, ZOKAEI, NAVA, LIU, ANDREA, et al. "Futzing and moseying: Interviews with professional data analysts on exploration practices". *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2018), 22–31. DOI: 10.1109/TVCG.2018.2865040 3.

[BAB*18] BATTLE, LEILANI, ANGELINI, MARCO, BINNIG, CARSTEN, et al. "Evaluating Visual Data Analysis Systems: A Discussion Report". *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*. HILDA'18. Houston, TX, USA: ACM, 2018, 4:1–4:6. ISBN: 978-1-4503-5827-9. DOI: 10.1145/3209900.3209901. URL: http://doi.acm.org/10.1145/3209900.3209901 10.

[BCS16] BATTLE, LEILANI, CHANG, REMCO, and STONEBRAKER, MICHAEL. "Dynamic prefetching of data tiles for interactive visualization". *Proceedings of the 2016 International Conference on Management of Data*. 2016, 1363–1375. DOI: 10.1145/2882903.2882919 3.

[BH19] BATTLE, LEILANI and HEER, JEFFREY. "Characterizing Exploratory Visual Analysis: A Literature Review and Evaluation of Analytic Provenance in Tableau". *Computer Graphics Forum* 38.3 (2019), 145–159. DOI: https://doi.org/10.1111/cgf.13678. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13678. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13678 1–3, 5–7, 9.

[BM13] BREHMER, MATTHEW and MUNZNER, TAMARA. "A multi-level typology of abstract visualization tasks". *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), 2376–2385. DOI: 10.1109/TVCG.2013.124 1–5, 7, 8.

[BOZ*14] BROWN, ELI T, OTTLEY, ALVITTA, ZHAO, HELEN, et al. "Finding waldo: Learning about users from their interactions". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), 1663–1672. DOI: 10.1109/TVCG.2014.2346575 2.

[BS21] BATTLE, LEILANI and SCHEIDEGGER, CARLOS. "A Structured Review of Data Management Technology for Interactive Visualization and Analysis". *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2021), 1128–1138. DOI: 10.1109/TVCG.2020.3028891 3.

[CCI*15] COOK, KRISTIN, CRAMER, NICK, ISRAEL, DAVID, et al. "Mixed-initiative visual analytics using task-driven recommendations". *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE. 2015, 9–16. DOI: 10.1109/VAST.2015.7347625s 1.

[CGL20] CUTLER, ZACH, GADHAVE, KIRAN, and LEX, ALEXANDER. "Trrack: A Library for Provenance-Tracking in Web-Based Visualizations". *2020 IEEE Visualization Conference (VIS)*. 2020, 116–120. DOI: 10.1109/VIS47514.2020.00030 1, 10.

[CWK*17] CHO, ISAAC, WESSLEN, RYAN, KARDUNI, ALIREZA, et al. "The anchoring effect in decision-making with visual analytics". *2017 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE. 2017, 116–126. DOI: 10.1109/VAST.2017.8585665 5.

[DC16] DABEK, FILIP and CABAN, JESUS J. "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 (2016), 41–50. DOI: 10.1109/TVCG.2016.2598471 3.

[FDPH16] FENG, MI, DENG, CHENG, PECK, EVAN M, and HARRISON, LANE. "Hindsight: Encouraging exploration through direct encoding of personal interaction history". *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2016), 351–360. DOI: 10.1109/TVCG.2016.2599058 5.

[FPH18] FENG, MI, PECK, EVAN, and HARRISON, LANE. "Patterns and pace: Quantifying diverse exploration behavior with visualizations on the web". *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2018), 501–511. DOI: 10.1109/tvcg.2018.2865117 5.

[GGZL15] GUO, HUA, GOMEZ, STEVEN R, ZIEMKIEWICZ, CAROLINE, and LAIDLAW, DAVID H. "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 (2015), 51–60. DOI: 10.1109/TVCG.2015.2467613 1–3, 5, 9.

[GTS10] GRAMMEL, LARS, TORY, MELANIE, and STOREY, MARGARET-ANNE. "How information visualization novices construct visualizations". *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), 943–952. DOI: 10.1109/TVCG.2010.164 3, 5.

[GW09] GOTZ, DAVID and WEN, ZHEN. "Behavior-driven visualization recommendation". *Proceedings of the 14th international conference on Intelligent user interfaces*. 2009, 315–324. DOI: 10.1145/1502650.1502695 2, 3, 5, 6, 9.

[GZ09] GOTZ, DAVID and ZHOU, MICHELLE X. "Characterizing users' visual analytic activity for insight provenance". *Information Visualization* 8.1 (2009), 42–55. DOI: 10.1057/ivs.2008.31 1–3, 5–9.

[GZA06] GOTZ, DAVID, ZHOU, MICHELLE X, and AGGARWAL, VIKRAM. "Interactive visual synthesis of analytic knowledge". *2006 IEEE Symposium On Visual Analytics Science And Technology*. IEEE. 2006, 51–58. DOI: 10.1109/VAST.2006.261430 3.

[Hib99] HIBINO, STACIE L. "Task analysis for information visualization". *International Conference on Advances in Visual Information Systems*. Springer. 1999, 139–146. DOI: 10.1007/3-540-48762-X_18 3.

[HMSA08] HEER, JEFFREY, MACKINLAY, JOCK, STOLTE, CHRIS, and AGRAWALA, MANEESH. "Graphical histories for visualization: Supporting analysis, communication, and evaluation". *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), 1189–1196. DOI: 10.1109/TVCG.2008.137 1, 3.

[KC20] KAHNG, MINSUK and CHAU, DUEN HORNG POLO. "How Does Visualization Help People Learn Deep Learning? Evaluating GAN Lab with Observational Study and Log Analysis". *2020 IEEE Visualization Conference (VIS)*. IEEE. 2020, 266–270. DOI: 10.1109/VIS47514.2020.00060 2, 5.

[KGS09] KANG, YOUN-AH, GORG, CARSTEN, and STASKO, JOHN. "Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study". *2009 IEEE Symposium on Visual Analytics Science and Technology*. IEEE. 2009, 139–146. DOI: 10.1109/VAST.2009.5333878 3.

[KK17] KERRACHER, NATALIE and KENNEDY, JESSIE. "Constructing and evaluating visualisation task classifications: Process and considerations". 36.3 (2017), 47–59. DOI: 10.1111/cgf.13167 2.

[KPHH12] KANDEL, SEAN, PAEPCKE, ANDREAS, HELLERSTEIN, JOSEPH M, and HEER, JEFFREY. "Enterprise data analysis and visualization: An interview study". *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), 2917–2926. DOI: 10.1109/TVCG.2012.219 3.

[KS11] KANG, YOUN-AH and STASKO, JOHN. "Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study". *2011 IEEE conference on visual analytics science and technology (VAST)*. IEEE. 2011, 21–30. DOI: 10.1109/VAST.2011.6102438 3.

[KSHL20] KARER, B, SCHELER, I, HAGEN, H, and LEITTE, H. "ConceptGraph: A Formal Model for Interpretation and Reasoning During Visual Analysis". 39.2 (2020), 5–18. DOI: 10.1111/cgf.13899 3.

[LFB*14] LANDESBERGER, TATIANA VON, FIEBIG, SEBASTIAN, BREMM, SEBASTIAN, et al. "Interaction taxonomy for tracking of user actions in visual analytics applications". *Handbook of Human Centric Visualization*. Springer, 2014, 653–670. DOI: 10.1007/978-1-4614-7485-2_26 3.

[LH14] LIU, ZHICHENG and HEER, JEFFREY. "The effects of interactive latency on exploratory visual analysis". *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), 2122–2131. DOI: 10.1109/TVCG.2014.2346452 2, 5, 7.

[LTM17] LAM, HEIDI, TORY, MELANIE, and MUNZNER, TAMARA. "Bridging from goals to tasks with design study analysis reports". *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2017), 435–445. DOI: 10.1109/TVCG.2017.2744319 2, 3, 5.

[Mac86] MACKINLAY, JOCK. "Automating the Design of Graphical Presentations of Relational Information". *ACM Trans. Graph.* 5.2 (Apr. 1986), 110–141. ISSN: 0730-0301. DOI: 10.1145/22949.22950. URL: https://doi.org/10.1145/22949.22950 7.

[OGW19] OTTLEY, ALVITTA, GARNETT, ROMAN, and WAN, RAN. "Follow The Clicks: Learning and Anticipating Mouse Interactions During Exploratory Data Analysis". 38.3 (2019), 41–52. DOI: 10.1111/cgf.13670 3.

[PC05] PIROLLI, PETER and CARD, STUART. "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis". *Proceedings of international conference on intelligence analysis*. Vol. 5. McLean, VA, USA. 2005, 2–4 3.

[PW18] PSALLIDAS, FOTIS and WU, EUGENE. "Smoke: Fine-Grained Lineage at Interactive Speed". *Proc. VLDB Endow.* 11.6 (Feb. 2018), 719–732. ISSN: 2150-8097. DOI: 10.14778/3184470.3184475. URL: https://doi.org/10.14778/3184470.3184475 1, 10.

[PWM*12] POHL, MARGIT, WILTNER, SYLVIA, MIKSCH, SILVIA, et al. "Analysing interactivity in information visualisation". *KI-Künstliche Intelligenz* 26.2 (2012), 151–159. DOI: 10.1007/s13218-012-0167-6 1.

[RJPL16] REDA, KHAIRI, JOHNSON, ANDREW E, PAPKA, MICHAEL E, and LEIGH, JASON. "Modeling and evaluating user behavior in exploratory visual analysis". *Information Visualization* 15.4 (2016), 325–339. DOI: 10.1177/1473871616638546 3.

[Shn96] SHNEIDERMAN, BEN. "The eyes have it: A task by data type taxonomy for information visualizations". *Proceedings 1996 IEEE symposium on visual languages*. IEEE. 1996, 336–343. DOI: 10.1109/VL.1996.545307 2–5, 9.

[SMG*20] SUKUMAR, POORNA TALKAD, MARTINEZ, GONZALO J, GROVER, TED, et al. "Characterizing Exploratory Behaviors on a Personal Visualization Interface Using Interaction Logs". *EuroVis 2020-Short Papers* (2020), 79–83. DOI: 10.2312/evs.20201052 2.

[Smi43] SMITH, MAPHEUS. "The Method of Elimination in Scientific Study". *Philosophy of Science* 10.4 (1943), 250–254. DOI: 10.1086/286816 5.

[SMWH17] SATYANARAYAN, ARVIND, MORITZ, DOMINIK, WONG-SUPHASAWAT, KANIT, and HEER, JEFFREY. "Vega-Lite: A Grammar of Interactive Graphics". *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), 341–350. DOI: 10.1109/TVCG.2016.2599030 3, 7.

[SNHS13] SCHULZ, HANS-JÖRG, NOCKE, THOMAS, HEITZLER, MAGNUS, and SCHUMANN, HEIDRUN. "A design space of visualization tasks". *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), 2366–2375. DOI: 10.1109/TVCG.2013.120 3.

[SP13] SEDIG, KAMRAN and PARSONS, PAUL. "Interaction design for complex cognitive activities with visual representations: A pattern-based approach". *AIS Transactions on Human-Computer Interaction* 5.2 (2013), 84–133. DOI: 10.17705/1THCI.00055 3.

[SvW08] SHRINIVASAN, YEDENDRA BABU and van WIJK, JARKE J. "Supporting the analytical reasoning process in information visualization". *Proceedings of the SIGCHI conference on human factors in computing systems*. 2008, 1237–1246. DOI: 10.1145/1357054.1357247 3.

[TSM17] TORSNEY-WEIR, THOMAS, SEDLMAIR, MICHAEL, and MÖLLER, TORSTEN. "Sliceplorer: 1D slices for multi-dimensional continuous functions". 36.3 (2017), 167–177. DOI: 10.1111/cgf.13177 1, 5.

[Wal20] WALL, EMILY. "Detecting and Mitigating Human Bias in Visual Analytics". PhD thesis. Georgia Institute of Technology, 2020 2, 4, 5, 7.

[Wan09] WANG, PING. "The Inter-Rater Reliability in Scoring Composition." *English language teaching* 2.3 (2009), 39–43 5.

[XOW*20] XU, KAI, OTTLEY, ALVITTA, WALCHSHOFER, CONNY, et al. "Survey on the analysis of user interactions and visualization provenance". 39.3 (2020), 757–783. DOI: 10.1111/cgf.14035 1.

[YaKSJ07] YI, JI SOO, ah KANG, YOUN, STASKO, JOHN, and JACKO, JULIE A. "Toward a deeper understanding of the role of interaction in information visualization". *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), 1224–1231. DOI: 10.1109/TVCG.2007.70515 2–5, 7, 8.

[YGR21] YAN, JING NATHAN, GU, ZIWEI, and RZESZOTARSKI, JEFFREY M. "Tessera: Discretizing Data Analysis Workflows on a Task Level". *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021. ISBN: 9781450380966. DOI: 10.1145/3411764.3445728. URL: https://doi.org/10.1145/3411764.3445728 3.