Study of Factors Affecting the Problem and Task Characterization for Time-Stamped Event Sequences

Master Thesis

Yasara Peiris

19-765-288

Submitted at
August 14th 2022

Thesis Supervisor
Prof. Dr. Jürgen Bernard

INTERACTIVE
VISUAL
DATA
ANALYSIS
GROUP

Master Thesis

Author: Yasara Peiris, yasara.peiris@uzh.ch

Project period: 14.3.2022 - 14.8.2022

Interactive Visual Data Analysis Group
Department of Informatics, University of Zurich

Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisor Prof. Jürgen Bernard for his immense guidance, and continuous support given throughout the project. Also, I would like to sincerely thank Clara-Maria Barth for her contribution in the project. I also thank my colleagues in IVDA lab for the support they extended. My gratitude also goes to all the academic and non-academic staff at UZH who helped me to successfully complete my postgraduate studies. Last, but not least, I thank my parents, my siblings and my husband who supported me in numerous ways during my studies at UZH.

LAUSANNE, 14TH OF AUGUST, 2022

Abstract

Time-stamped event sequences (TSES) are event sequences without values. Analysts are mainly interested in the temporal signatures of phenomena. It is a hardly investigated data type with growing interest since it is observed across a wide range of domains. There are two main problems in TSES that hamper the design of visual-interactive solutions. First, lack of awareness of affecting factors for problem characterization for TSES and second, lack of specific task characterization for TSES. Consequently, designers have a hard time making correct design decisions when building data analysis solutions for TSES, which ultimately influence the effectiveness of the tool to be built. We conducted two types of studies to address these problems. To address the lack of awareness of affecting factors, we did a systematic characterization of TSES-oriented real-world problems structured by four main aspects: (1) domain context & users, (2) data characteristics, (3) tasks, (4) metrics. In our study approach, we systematically identified a diverse set of factors associated with these above-mentioned main four aspects initially. Then, we collected 65 TSES-oriented real-world problems spanning a wide range of domains, focusing on identified factors using a user-based survey study. Lastly, we systematically analyzed the discovered factors and then related them to identify the relationships between factors. To address the second problem of lack of specific task characterization, we presented a generalized problem characterization for TSES. In our study approach, we used two complementary survey sources: a User-based survey study and a Survey of design studies. Initially, we built a generalization of tasks that are currently supported in vis tools related to TSES based on 16 design studies for TSES and related, which resulted in 26 tasks. Then we built a generalization of user tasks based on 65 survey responses, which resulted in 25 tasks. For the generalization, we used coding and affinity diagramming, well-known techniques of qualitative research. Finally, we unified these two sources and proposed a task characterization consisting of 28 tasks for TSES. We found that questionnaire answers on TSES are extremely heterogeneous, and no two answers are similar or even equal with some degree of abstraction. Results of the second part of the study show that 80% of the user tasks are similar to tasks extracted from design studies focused on TSES or related data types.

Contents

Chapter 1

Introduction

As data volumes increase, it is impossible to analyze data manually to gain useful information. Consequently, there is an increasing interest in gaining new insights from data analysis support. It has become crucial for many fields such as chemistry, software, cybersecurity, and environmental sciences to analyze their own data, so they can gain new insights. Visualization (vis) tools show data visually in a way that makes insights easier to comprehend for humans. As a result, vis tools became popular, driving research in multiple fields including machine learning, human-computer interaction, and visualization proposing different guidelines, methodologies, and techniques that could make those tools more effective for the user. Design studies are a common approach for building vis tools. There are two main types of design studies to create vis tools: problem-driven design studies and data-first design studies, each focusing on users and data respectively, proving that both aspects and their characteristics are equally important. Having a clear idea of both users and data makes the design process efficient and the final output more effective. In such design processes, there are a number of **factors** to take into consideration in multiple **aspects** besides user characteristics and data characteristics such as user needs and domain context characteristics. This is called *problem characterization* [66, 86] in the vis community which involves consideration of multiple aspects such as learning the target domain and the problems, data, and needs of the domain experts in order to have a shared understanding between vis researcher and the domain experts. According to Sedlmair et al.[86] this is a crucial step in design study and even argued this as a first-class contribution of a design study. Further, the authors have pointed out multiple potential benefits of problem characterization including future researchers who propose different solutions to the same problem can benefit from this knowledge, as well as this allows more automation which doesn't require humans in the loop as a result of externalizing and articulated domain knowledge.

The data type we mainly focus on in our research is 'Time-Stamped Event Sequences'. We refer to "time-stamped event sequences" as *TSES* for both singular and plural from this point onward. A TSES is a sequence of time-stamped events with no particular value domain (refer Figure 1.1), where the analytical focus is mainly on the temporal signatures of the events. TSES can be found across a wide variety of application domains including *manufacturing, healthcare, education, finance, software, retail, or geological phenomena*. Example cases of TSES include *git commits, hospi-*

talizations due to Covid-19, customer complaints, traffic accidents, sleep behavior etc. This special data type contains three natural granularities (refer Figure 1.2): events, event sequences, and groups of event sequences. In addition to time signatures, TSES often are accompanied by a rich set of relevant attributes to encode the context in the data. We refer to these attributes as metadata and there can be a variety of metadata types, either categorical, numerical, or multivariate, and the number of metadata can vary as well. There are many examples of metadata, including *geographical location of some Earth phenomena, device type for measuring blood pressure, and vehicle type in accidents*, and all play a fundamental role in building the context for either event, sequence of events, or group of event sequences. Given three granularities in the data and the heterogeneous nature of metadata, TSES are a complex and specific data type. Users carry out special tasks on it due to its unique nature. There has been a diverse set of tasks [32, 75] identified in the vis community for event sequence data such as compare, grouping, identify trends, etc. The diversity of the tasks increases when the tasks are associated with the different granularities of the data and the metadata. Furthermore, due to their unique characteristics, TSES are often assessed using a set of metrics that translate data characteristics into the numerical format, which allows easy comparison, trend identification, etc. Some common metrics found are peaks, gaps, recurrences, accelerations, density, or outliers.

Even though TSES is a hardly investigated data type, there is a growing interest towards analyzing TSES data since it spans across a wide range of domains. Therefore, proper problem characterization of TSES is important when building vis tools focusing on TSES. Existing research revealed three main aspects that need special attention: *domain context* & *users*, *data*, and *tasks.* Six qualitative interviews with domain experts focusing on TSES in our previous work and existing work on TSES revealed *metrics* are also a special aspect of TSES that requires special attention. To explain these aspects in detail: (1) Domain context & user aspect refers to learning about the problem, target domain, goals, existing practices, and user characteristics, (2) data aspect refers to learning about the general discipline that the dataset describes, (3) task aspect refers to learning about users' data analysis and visualization needs, (4) metrics aspect refers to learning about users' interested features. However, designers are not informed about these different influencing factors for problem characterization.

For TSES two main problems hamper the design of interactive visual analysis solutions.

- **Problem 1**: Lack of awareness of affecting factors for problem characterizations for TSES
- **Problem 2**: Lack of specific task characterizations for TSES

The lack of awareness of designers on different factors can lead to poor and incomplete problem characterization, which ultimately affects the effectiveness of the final tool to be designed. This is due to three main reasons. First, there are no existing works discussed all four aspects which mentioned earlier, and if so not at the same balance or level of depth. Second, the factors existing problem driven work paid attention to are quite diverse, so are their emphasis. Even though reflection on the existing design studies often guide new design studies and gives clues as to which path to follow, this is not an option for TSES due to the limited existing design studies focusing on TSES. Third, designers have a lack of solid understanding of the data type and overview of TSES oriented problems, since existing work doesn't provide many insights about the data type and also about diverse TSES problems. Consequently, designers are limited in making proper design decisions.

Further, we identified that the awareness for analysis task is crucial for the design of usable and useful data analysis systems. However, for TSES, this awareness hardly exists and the reflection on related work reveals 3 reasons. First, there are general task taxonomies, but they aren't specific enough to apply for a specific data type. Second, there are data-specific task taxonomies that recognize the value of data beside the tasks. But, there is no data-specific taxonomy for TSES. Third, design studies provide domain specific task taxonomies, but nice to have a domain agnostic task taxonomy. Unfortunately, generalizability is not possible in the context of TSES, due to limited existing design studies exist for TSES or related. Consequently, general users may have heterogeneous tasks when working with TSES that have not yet been taken into account.

In this thesis, we conducted two types of studies to address these problems. We did (1) a systematic characterization of problems to address the lack of awareness of affecting factors for problem characterizations for TSES using a user-based survey study and (2) a systematic identification of tasks to address the lack of task characterization for TSES using two complementary studies: a user-based survey study and survey on existing literature study. The research questions on the lack of problem characterization were (1) RQ1: How do domain-related factors impact the problem characterization?, (2) RQ2: How do diverse data characteristics impact the problem characterization?, (3) RQ3: How do tasks impact the problem characterization?, (4) RQ4: How do metrics impact the problem characterization?, (5) RQ5: What effect do domain context, user characteristics, data characteristics, tasks and metrics of TSES have on each other? while the research questions of the lack of task characterization were (1) RQ1: Can tasks derived from surveys be described in a common language?, (2) RQ2: Do common tasks exist?, (3) RQ3: How well do user study tasks match those from related design studies? (4) RQ4: Can two sources be unified? Overall, this work addresses these research questions with the following two main contributions.

1. Systematic characterization of TSES-oriented real-world problems structured by four major aspects: *domain context & users*, *data characteristics*, *tasks*, and *metrics* using a user-based survey
   - Systematic identification, analysis of factors associated with domain context/users, data characteristics, tasks, and metrics
   - Identification of relationships between different factors based on the systematically identified real-world problems.
2. Systematic domain-agnostic task characterization for TSES based on two complementary survey sources.
   - Generalization of tasks that are currently supported in vis tools related to TSES based on a survey on design studies.
   - Systematic review of the current state of user tasks based on user-based survey.
   - Unification of two sources to present a task characterization.

The remainder of this thesis is structured as follows. In Chapter 2, we provide an overview of three core research areas that inspired this study: Information Visualization and Visual Analytics, Vis Design Methodology, and Time-Stamped Event Sequences. In chapter 3, we will reflect on the two main problems discovered and lay the foundation for the study work. Methodology & results related to problem 1 are explained in Chapter 4, while problem 2 is discussed in Chapter 5. Limitations of this work and discussion items and the future work are presented in Chapter 6. Finally, the conclusions along with recommendations for future work are presented in Chapter 7.

Data Type	Values	Time	Example
Time-stamped event sequences	none		Heartbeats
Classical event sequences	1x cat		Treatment plan
Univariate time series	1x num		Stock chart
Multivariate time series	nx num/cat		Sensor data

Figure 1.1: Differences between TSES and other event sequence data and time-oriented data

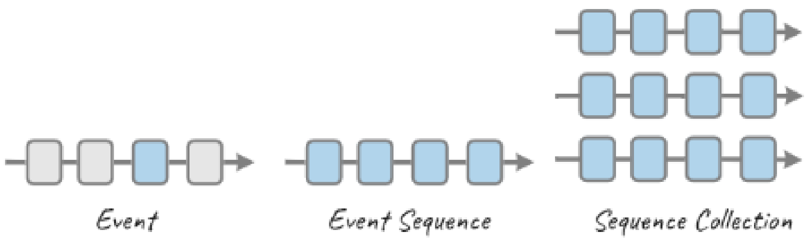


Figure 1.2: Granularities naturally exist in TSES data. Individual events represent the finest granularity of event sequences, an event sequence is the complete record of events and a group of event sequences represent a collection of event sequences.

Chapter 2

Related Work

We start with providing an overview of Info Vis(IV) and Visual Analytics(VA) as an overarching research field. Next, we emphasize vis design methodologies as it is similar to our type of study. Then, we give a profound overview of TSES, which is the data type at hand with unique characteristics that need to be considered in the study of design/user support. Finally, we shift our focus towards tasks by providing an overview of tasks related to TSES. We further reviewed existing task characterizations and methodological aspects to building task characterizations, since in our study we are planning to present a task characterization.

2.1 Information Visualization and Visual Analytics

Information visualization typically focuses on abstract data by showing it visually in a way that makes insights easier to comprehend for humans [90] while visual analytics(VA) combines computational techniques with information visualization to support analytical reasoning about the data. In recent years, VA has gained popularity due to the heterogeneous, massive, and complex nature of the data in a wide range of domains. VA is "the science of analytical reasoning facilitated by interactive visual interfaces", which integrates human knowledge in the data analysis process through visualizations [21, 95, 44]. In contrast, it can be viewed as an integrated approach to decision-making, combining visualization, human factors, and data analysis [21, 44]. Ergonomics et al. have pointed out that user plays an integral part of the visual analytics process by interacting with the tool [24]. At the end, it is the user who observes the visualizations to generate new insights through reasoning. Therefore, user's perception and interaction with a visualization tool has a significant effect on their understanding of data as well as the tool's usefulness. Therefore, significant efforts have been made to develop visualization tools that can be used by users to gain relevant information.

There have been numerous research in many fields including machine learning, cognitive sciences, and advanced algorithms suggesting different methodologies, design guidelines, and algorithms to provide users with VA tools that meet their expectations. Cui and Sun reviewed the recent research in VA and presented a comprehensive overview on research advances in the field to provides a clear understanding on current edge techniques in the field [21, 91]. Accordingly, there are a number of methodologies, guidelines proposed in the literature which can be utilized when building vis tools. Methodologies and guidelines are a crucial research line of research in VA, leading to effective VA tools (cf. Section 2.2).

A number of vis tools have been proposed in the literature to assist users in solving complex problems with large data volumes by helping them gain new knowledge by incorporating guidelines and techniques proposed in the literature. VA tools can be divided into two main types, depending on whether they were developed in response to a set of problems or a particular problem. (1) Generalized vis tools are developed as a general solution to a group of problems. Therefore, requirements and constraints spans across all the problems were given attention when building these tools. For example, these tools can be either data specific or domain specific or data agnostic or domain agnostic solutions. (2) Specialized vis tools are developed as a solution to a specific problem in the real-world by giving attention to all the constraints and requirements only relevant to that problem. In our study, we aim to find out different important factors that facilitate the development of both specialized and generalized vis tools focusing TSES. The results of this study may benefit future research on a variety of aspects when building analytic tools such as problem characterization, design, and evaluation.

2.2 Vis Design Methodologies for Data Driven and Problem Driven Research

Literature offers many methodologies for visual design, which plays a crucial role in achieving an effective tool in the end. Section 2.1 reveals, it is the user who observes and gain insights which they are not aware of earlier to formulate new questions, hypotheses, and models. Therefore, a good visualization design should also include a valuable exploration process in addition to visualizing data [80]. Several author groups, such as Roger et al. [82] and Saket et al. [84], have stressed the importance of both usability and user experience in visualization designs. Due to the major role users play in the analysis process, it was identified that users should be included in the design process as well.

Design Studies User-centered design has been extensively discussed and applied in visualization [48, 88, 8, 58]. A common method for building these vis tools are design studies. Two types of design studies exist, data-first design studies and problem-driven design studies, of which problem-driven design studies are commonly used as it was proposed 10 years ago while data-first design studies was an outcome of a recent research. Problem-driven studies address real-world problems by involving users early in the process. There are two possible ways to do the problem driven design studies in order to build an effective vis tool efficiently. Either investigate existing design studies to gain a better understanding of the current state of the problems and critical factors that need special attention prior to undertaking the study, or utilize existing design study methodologies for the design process [86]. Due to the little guidance on how to do problem-driven design studies effectively, Sedlmair et al. proposed a nine stage methodology [86]. Collaboration between visualization researchers and domain experts is a fundamental and mandatory part of this nine-stage framework. There have been several enhancements suggested to the above framework [62, 63] while keeping collaboration between the researcher and the domain expert unaltered. The other method of building vis tools is data-first design studies which is applicable when problem-driven research motivated and inspired by data, which is similar to our user-based survey study which was inspired by TSES. Compared to the problem-driven visualization research, stakeholders join later in the process. Despite data being the main focus of data-first studies, visualization researchers should be experts in the domain and act as primary stakeholders in developing visualization tools based on their own experiences. Therefore, it is evident that user play a vital role, besides data being the main focus in these data-first design studies [71]. Oppermann et al. proposed methodology for data first design studies consists of 10 stages by making some changes to existing problem driven methodology. Despite differences in process order, both methodologies emphasize tasks, data, and domains in their respective studies [71, 86]. Thus, we decided to utilize two types of data sources to answer our research questions, one of which focuses on data-first studies and the other focusing on problem-driven studies using a user-based survey studies focusing on TSES and survey on problem driven design studies related to TSES respectively.

Aspects & Factors influencing Visualization Design There is a rich source of existing work which proposed guidelines that can be applied to design studies. There are many factors that were taken into account during different studies, so in this section we discuss how they were taken into consideration during different studies along with the different factors. We identified a set of factors stems from domain context, data, and task aspects. Many studies have emphasized the importance of learning tasks and data of target users in some particular target domain [66, 86, 10]. Clear understanding of *existing tools and workflows* allows designers to understand the data and also the actions carried out by target users [66]. Brehmer also pointed out that it is crucial to understand how *individuals*, *existing tools*, *data*, and *contextual factors* interact before developing a tool suitable for the target audience and its unique contexts of use [10] and have presented different methods that can be used to learn those factors. Similarly, Plaisant argues that information visualization designers should study the design context for visualization tools, including *tasks*, *environments*, and *current practices* [74]. Lloyd et al. highlighted the significance of common understanding of *context of use*, *domain data* to successful vis design based on their design study in the context of geovis [58]. Crisan et al. explored two types of external constraints in depth, *regulatory and organizational constraints*, and describe how these constraints impact visualization design and evaluation [20]. Many task taxonomies and design spaces are proposed in the vis community, confirming that *user needs* are the most important aspect of designing a vis tool [3, 85, 100]. On the other hand, Pretorius et al. argued the importance of understanding *data characteristics* when designing a visualization, more than designing for user needs [78] revealing the equal importance of both tasks and data. There are data specific task taxonomies proposed in the literature, claiming significance of paying attention to the *data characteristics* besides tasks [12, 1, 32]. Hackos et al. listed a wide range of factors which affect the task analysis when building a system. This list contains many factors such as *personal,social, cultural characteristics* of user, *user preferences, goals*, *experience*, *physical environment*, *tasks* etc. [33]. Our work is inspired by this work, and identified the possibility to have a similar set of factors that is specific enough to characterize TSES. We identified different factors stems from 4 main aspects namely domain context, data, tasks, and metrics. To summarize, domain context aspect includes existing tools [66, 10, 74], existing work flows [66], contextual factors [10, 74, 58, 33], regulatory constraints [20], goals [33], organizational constraints [20] while user characteristics aspect includes individuals [10], personal,social, cultural characteristics of user [33], user preferences [33]. Further, many studies have researched on task aspect [74, 3, 85, 100, 33]. Importance of data characteristics aspect is also highlighted [10, 58, 78, 12, 1, 32]. Furthermore, we discovered metrics also can contribute significantly to problem characterization, although this is a very specific aspect that is not applicable to most data types. In section 3.1, we will discuss the problem of lack of awareness of these influencing factors and then summarize different factors identified under different aspect in more detail.

2.3 Time-Stamped Event Sequences

Event sequences are an ordered list of events [32]. There are different types of event sequence data as shown in Figure 1.1 such as classical event sequences (categorical, value), univariate (numeric value), and multivariate time series (typically multiple numeric values). However, in our study, our focus is on TSES. In contrast to classical event sequences, in TSES each event contains a timestamp without any value information. TSES spans across a wide range of domains including *manufacturing, healthcare, education, finance, software, retail, or geological phenomena, education, food. Git commits, hospitalizations due to Covid-19, customer complaints, traffic accidents, sleep behavior, earthquake aftershocks, or heartbeat* are some example scenarios of TSES. Domain context is also playing an important role, since most of the user needs are highly specific to the given domain and the context.

Following that, we will examine data characteristics of TSES. TSES consists of three natural granularities: event, event sequences, and group of event sequences 1.2. The size of the data depends on the number of events and the number of event sequences of the dataset, and can vary in both sides. Data analysis is even more challenging due to the large data volume when event sequence and event count are high. Du et al. proposed different strategies to cope up with growing data volume [23] by sharpening analytic focus. Some of the interesting strategies that are relevant to TSES are *'aligning' 'temporal folding', 'partitioning'*. Guo et al. also listed different interaction techniques to support users with enough flexibility [32]. Some interesting techniques relevant to TSES and this study are *'segmentation','aligning'*. These different strategies and techniques relevant to TSES will be discussed in greater detail in Section 3.1.

Most related approaches for time-oriented data support classical event sequences with categorical events or numerical time series of univariate or multivariate types. Guo et al. presented a review of state-of-the-art visual analysis approaches for classical event sequences categorized by analytical use and application domain [32]. Research work in this field spans across a wide range of domains serving a variety of tasks such as health informatics [15, 15, 34, 70], social media [65, 87, 102], computer systems [65, 87, 102]. According to Guo et al. some of the frequently applied event sequence tasks are *'summarization', 'prediction and recommendation', 'anomaly detection', and 'causality analysis', 'compare'* [32]. However, due to the differences in data structure, all existing techniques for event sequences can't be directly integrated into this data type. This allows for great exploration, which led to its growing popularity in diverse research fields. There exist only a few visualization approaches and design studies for TSES[103, 68]. Unique data characteristics of TSES opens up new task space such as metadata introduce a new set of analysis tasks on 'relation seeking' due to the ability to relating to the context compared to the existing event sequence analysis. Tasks play the central role in vis design. Even though TSES holds a different set of tasks due to it unique nature, there is no task characterization for TSES. This inspired us in the second part of the study.

Furthermore, in data mining research, TSES are often assessed with metrics which produce features to support the analysis. Metrics address in this study are size, subsequence length, regularity, distance, acceleration, outliers in event sequence level, outliers in event level, gaps, skewness, density, entropy, peakiness. However, this is a unique aspect compared to other data types and will be discussed further in the Section 3.1.

Even though there is a large exploration space exist, there is very little existing work on TSES. Therefore, reflection of design studies are not an option to learn about different characteristics of TSES which leads to poor design decisions, ultimately affecting the effectiveness of the tool to be built. Therefore, we identified the need of a problem characterization for TSES will be discussed further in the Section 3.1.

2.4 Tasks, Task Characterizations & Building Methodologies

Tasks for Time-Stamped Event Sequences In this section, we will look at the tasks of TSES. Most of the existing work focuses on classical event sequence data even though our main focus is on TSES. Guo et al. reviewed state-of-the-art visual analytic approaches for event sequences and categorized them based on application domain and analytic tasks [32]. According to the study, event sequence data is frequently observed and discussed in medical, internet, and industry 4.0 applications. The most prominent tasks found across these domains are patterns of individual sequences [53, 73, 28, 29], cluster event sequences [29] and cohort comparison [13, 49, 50, 61, 76], cohort summarization [30, 31, 77, 108], progression analysis [97, 73], detect anomalous sequences [14, 102]. Despite not being prominent, the following tasks are also present across these domains such as modify individual sequences [51] and predict [41].

Even though these analytic tasks are focusing on classical event sequences, most of these can be integrated in to our focus data type TSES. The unique data structure of TSES compared to classical event sequences limits the application of some tasks while allowing new set of tasks. Section 2.3 present unique characteristics of TSES such as metadata, metrics. Metadata creates the context for an event or event sequence or group of event sequences, which opens up a new task analysis space on identifying relations between different data granularities and metadata. Furthermore, many analysis tasks can be identified with metrics used by TSES for identifying interesting data features.

However, despite the new task analysis space, there is no task characterization for TSES. Generalized task taxonomies hides the new analytic tasks space. Due to this, most task abstractions in related design studies have been done manually [22, 60, 98].

Task Characterizations in InfoVis To build a task characterization, we review existing task taxonomies, task typologies, and task characterizations. Task abstraction is an important step in vis design process [66, 86]. Abstract tasks are domain agnostic & interface agnostic operations performed by users [66]. Therefore, abstract tasks permit systematic analysis and the re-use of existing visualization methods, transferability of a design study's specific result into another domain [67, 81, 5, 86]. Due to these benefits, there are many research proposing abstract task frameworks. Most of the proposed framework are generic taxonomies [100, 4, 85, 11]. However, these are too broad to applied into specific data types. There are set of task taxonomies which are domain-specific such as genomics [64] or network security [92]. There exist data-specific task taxonomies and more useful for specific data types since it consists with less generic task descriptions [12].

Examples on data-specific task taxonomies focus on dimensionally-reduced data [12], temporal network visualization [1], classical event sequence data [32], time series data [2], or graph visualization [54]. However, there is no data-specific task taxonomy for TSES.

Another important aspect is the granularity of the tasks. Rind et al. surveyed and categorized existing task taxonomies by considering the composition level of the tasks [81]. According to their survey, the most common composition level among existing task taxonomies is *'low level'*. However, there exist *'high-level'* [4, 83, 94], *'intermediate-level'* [100, 3, 5] taxonomies also. For our task characterization on TSES, we consider the low-level to mid-level granularity to be most useful, sufficiently fine-grained that each task is its own actionable operation but coarse enough to differ from atomic actions or interactions.

Methodological Approaches for Task Characterizations In the vis research community, there has been much work based on qualitative methods [11]. According to Kerracher et al. [47], even though there have been a wide variety of task taxonomies proposed, there has been very little attention paid to the methodological approaches. Additionally, they provided a summary of methods used for task generation, task categorization, task description, and associated risks with risk mitigation approaches based on previous work which we considered when deciding our methodology. McKenna et al. [63] also compiled a comprehensive list of methods that can be used to gather tasks. Several other previous works also have been focused on determining how to obtain recurring visualization tasks, and Schulz et al. [85] discuss the most influential methods found from these previous works, including surveying individuals who are proficient in visualization[3], observational studies [37], and inferring from prevailing knowledge [1, 106]. In addition, interview studies also have been widely used in several previous studies as a qualitative methodology [12, 42, 43]. On the other hand, the authors have proposed a taxonomy of analytical focusing strategies based on multi-strand approaches, combining their expertise in specific fields with knowledge of literature, email interviews [22]. The multi-strand approach is presented as a mitigation method for risk introduced by wrong or missing tasks when only one methodological approach is used in task elicitation [46] which is also the inspiration for our approach. Moreover, we were strongly influenced by the methodological approach followed by Lam et al. in [52] in which the previous design study papers were scanned and sections were opened-coded, following an affinity diagram in order to develop a task taxonomy. In contrast, we determined to focus only on the detailed descriptions of tasks listed in the relevant papers, rather than extracting information from sections where a target user performs the analysis [52].

Therefore, a great majority of qualitative approaches have been proposed that can be utilized in our study to build a valid, useful, and less risky task characterization for TSES.

Chapter 3

Problem Reflection

3.1 "Problem 1": Absence of awareness of affecting factors for problem characterizations for TSES

Problem characterization is a crucial step for a vis design [66, 86]. According to Munzner et al., the first and foremost step of any vis design process is to characterize the tasks and data in the vocabulary of the problem domain. There are a number of varied factors that help properly characterize TSES real-world problems. However, there is no generalized problem characterization for TSES (cf. Chapter 1). Various factors have been highlighted by researchers due to their identification of the significance in different studies [66, 10, 20, 33, 86]. It is necessary to identify those diverse factors in order to create a generalized problem characterization. Two previous bodies of knowledge (cf. Section 4.1.1) revealed four main aspects that influence problem characterization (cf. Chapter 1): domain context & user, data, task and metrics. We investigated each aspect in depth using divide and conquer principle to identify different influential factors. However, investigation of existing literature (cf. Section 2.2) revealed a number of factors that affect problem characterization under these aspects and in this section we concretely identify these different influential factors. The symbol **(F)** will be used every time a factor is found. Figure 3.1 illustrates the process we followed to identify different factors. We will present all the identified factors first and then present the details. Figure 3.2 shows the list of 26 factors we identified based on a reflection of related works, all of which affect the problem characterization.

3.1.1 Role of Domain Context & Users

Context: Understanding the domain context is really important to understand the needs of the target users. Pre-design methods were recommended by Brehmer et al. in order to identify the feasibility of a project based on context-related factors [10]. There are a number of factors to support the understanding of the domain. In some situations, domain experts use domain vocabulary to express their needs, which signifies the importance of understanding the domain vocabulary **(F)** in order to translate domain requirements in to the abstract format. Further, users often fail to

Figure 3.1: We followed divide & conquer principle to identify factors. The purpose of this is to illustrate the process we followed for identifying possible influencing factors.

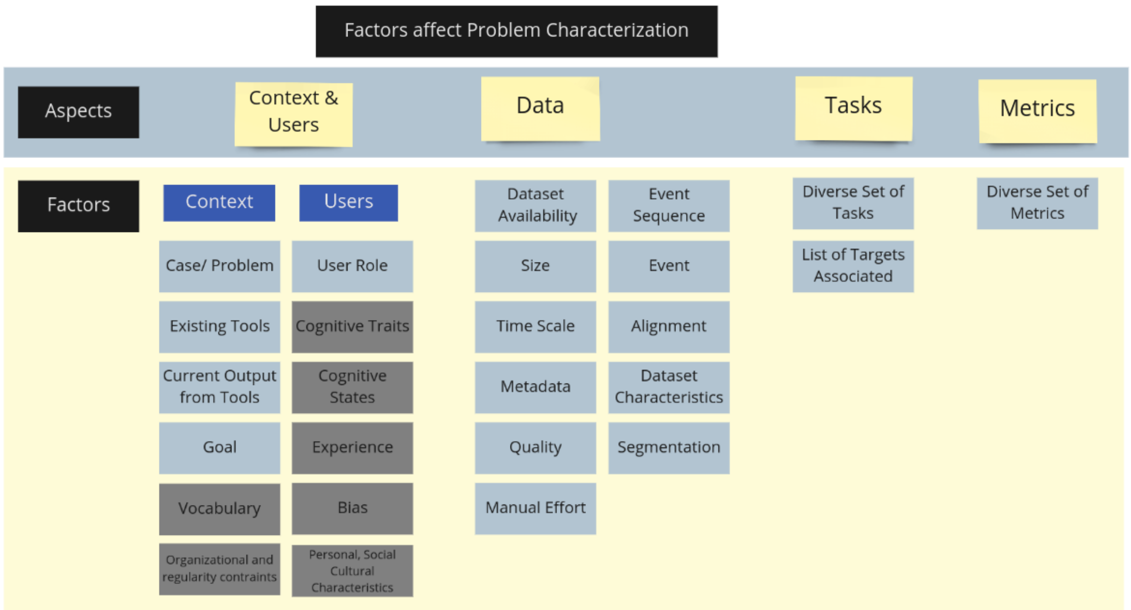


Figure 3.2: We used divide & conquer principle to identify influencing factors under each four main aspects: domain context & users, data, tasks and metrics which resulted in 26 factors.

accurately assess their own data analysis and visualization needs [66]. Therefore, standard practice is to do observational studies, interviews [38, 35] to understand the work flows and existing tools [86, 66]. Based on this, we identified existing tools **(***F***)** and results achieved from current tools **(***F***)** as possible factors, to understand the context. Crisan et al. argued that many studies focus only on user needs without giving attention to external constraints such as regulatory and organizational constraints **(***F***)** [20]. Furthermore, they have pointed out that ignoring these external factors can even lead to a premature termination of the project. Schulz et al. introduced domain goals as an aspect describing the tasks which specify the motive of a task. According to them, tasks and goals are independent, and the same motive can be accomplished with different actions, while the same action can be accomplished with different motives [85]. Therefore, domain goal **(**F**)** also identified as a main factor that affect the design of a visualization. Based on these findings, in our study we aim to extract understand the domain context using background, goals, tasks, and existing tools. Beside these factors, we discovered problem itself **(**F**)** is the most important factor to start the problem characterization.

Users: Section 2.1 reveals the importance of taking user characteristics into consideration when building a vis tool. In reviewing the existing literature, Peck et al. identified cognitive states, cognitive traits, and experience/bias as dominant cognitive dimensions affecting visualization performance [72]. (1) Cognitive Traits **(**F**)**: Individuals' perception, learning, and reasoning are significantly affected by cognitive traits such as spatial ability, verbal ability, and working memory capacity [72]. Liu et al. reviewed the existing work on the research perspectives, as well as the personality traits and cognitive abilities, visualizations, tasks, and measures [56] which uncovered the correlation between traits, visual design and tasks. The results revealed that there is a mismatch between visualization designs and individual differences of users [56]. However, there is growing interest in applying individual characteristics in to visualizations [110, 105, 6]. (2) Cognitive States **(**F**)**: The most studied cognitive state in visualization evaluation is cognitive load, as it often directly impacts performance [72]. In particular, working memory limits the visualization by both size and duration [72]. These factors already proposed as design guidelines to follow when building visualization. (3) Experiences **(**F**)** and Biases **(**F**)**: User experience is also affect the performance of the visualization reasoning [72]. It is common for both experience and bias to be formed from previous interactions with a given problem, and for these to be applied when a similar challenge is encountered later in life [72, 19, 96].

However, in visualization field there are existing works which considered user roles **(**F**)** when developing visualization. For example, Schulz et al. consider different user roles and proposed user-based interfaces by applying their design space to application problem in climate impact research [85] with the aim of relating to the context. Further, Irissary et al. proposed three different categories where audience can be grouped into based on their goal namely (1) own exploratory data analysis, (2) to convey a message to users, (3) to tell a story to audience [39]. In addition, Hackos et al. listed personal, social, and cultural characteristics of users **(**F**)** as factors affecting task analysis, in turn affecting problem characterization.

3.1.2 Role of Data Characteristics

Although classical event sequence data is not exactly similar to TSES, they have some characteristics in common, such as having three natural granularities: event, event sequences, and group of event sequences. For this reason, we consider recent research in either TSES or classical event data in order to identify possible factors in data aspects that may influence the tool's design which are specific to TSES. The review of TSES in Section 2.3 revealed the specific and even unique characteristics of this data type. Consequently, TSES also pose specific factors that need to be taken into account. As Du et al. described, one of the main challenges in the analysis of event sequence data is the volume of data available [23]. It was discovered that size **(**F**)** plays a major role in problem characterization, which is crucial during design decisions. However, size of the dataset depends on number of events and event sequences. It is possible to represent the same volume of data by more events and less event sequences or by fewer events and more event sequences, or by an equal number of events and sequences. Guo et al. have listed 'alignment' as a main analytic strategy that support comparison tasks [32]. According to them, multiple event sequences can be aligned based on either an event or time point. There are different alignment strategies have been used in different analytical tools. For example, Lifelines2 [99] supports the interactive alignment of event sequences based on a selected event, while Chen et al. allow both the alignment and change of the horizontal scale [17]. Based on the findings, an initial alignment **(**F**)** is also an important design decision that needs to be made early and should be considered during problem characterization. Another analytic strategy listed by Guo et al. is 'segmentation' which is useful when users are interested in analyzing a narrow scope than the whole event sequence [32]. Focusing on the requirements of tasks is significantly easier when you have a clear understanding of expert segmentation opinions. Segmentation **(**F**)** was therefore considered as a possible influencing factor of problem characterization.

Other than above factors, we identified some of the possible factors that affect the problem characterization based on the qualitative interviews we followed with six domain experts prior to this study. Clear understanding of data is crucial for proper problem characterization. Therefore, what is an event **(**F**)** and event sequence **(**F**)** relevant to the case also crucial. Section 2.3 provides details about metadata. Based on that number of metadata and type of the metadata, both can affect the problem characterization. Therefore, metadata is also considered as a possible factor. Often, real-world data contain noise or errors, resulting in the need for preprocessing. Depending on the dataset, different problems have different challenges and different processing efforts. Therefore, understanding the quality **(**F**)** and effort to processing **(**F**)** are also important factors that should be understood in advance. Lastly, it is easier for designers to work with data when they have a clear overview of it. To have an overview, we discovered some characteristics of dataset **(**F**)** including whether the data is historical, real, or both, whether the event sequences possess special qualities such as 'sparse', 'dense', 'regular', 'irregular', and whether collisions occur and how often.

3.1.3 Role of Tasks

There is a lot of discussion about tasks in the visualization community, since they play a crucial role in how we design and evaluate our work [81]. Therefore, we can identify task **(**F**)** as a main influencing factor for problem characterization. We listed tasks in the survey to identify the importance of different tasks for different cases. Since tasks is an ill-defined term, we use task cube, which is a conceptual space with three dimensions to describe tasks [81]. Since this is a domain agnostic survey, therefore these tasks should be in abstract form in the abstraction dimension, allowing all the users to understand. This questionnaire is to understand user needs therefore, all the tasks in the questionnaire were addressing why perspective with 'high-level to mid-level' composition. Previous bodies revealed a set of possible tasks that are relevant for TSES. Some examples of tasks are *'gain an overview'*, *'inspect individual events in detail'*, *'show metadata'*, *'sorting & ranking'* etc.

Tasks are described by Schulz et al. in five primary dimensions, with 'target' being one dimension we found most intriguing for our focus data type [85]. TSES consists of three natural granularities and set of metadata, and either granularity or metadata can be targeted by tasks, which determines which part it will perform on. Therefore, target associated with task **(**F**)** also identified as a possible factor that affect the problem characterization and further analyzed in section 4.

3.1.4 Role of Metrics

We also discovered metrics as a main aspect that influence the problem characterization. This is a special characteristic of TSES, as the related work on TSES in Section 2.3 reveals. In general, metrics emphasize important data characteristics by incorporating algorithmic support when it is difficult to capture interesting insights by using the human eye. Some examples for generic metrics are peaks, velocity, acceleration, density etc. Experts can select the relevant metrics for their analysis process, which can influence the tool design in the same way as tasks, so different types of metrics **(**F**)** also an important factor to consider.

Based on these discovered influential factors, we address the problem of "lack of awareness of factors affecting problem characterization".

3.2 “Problem 2: Lack of specific task characterizations for TSES

Tasks are crucial to assist the vis design. Therefore, a number of general task taxonomies have been proposed in order to assist the visualization design, such as Wehrend and Lewis's problem-oriented classification of visualization techniques [100], Amar and Stasko's prototypical analysis tasks [4]. As a result of the complexity of TSES due to the three natural granularities: event, event sequence, group of event sequence and varying number of metadata, these general task taxonomies are not suitable for the application in the context of the TSES. However, there are data-specific task characterizations in the literature which provides less generic description of tasks while considering specific set of data abstractions [12]. Examples of data-specific task taxonomies focus on dimensionally-reduced data [12], temporal network visualization [1], classical event sequence data [32], time series data [2], or graph visualization [54]. However, no taxonomy or characterization has been proposed so far for TSES. As a result, most of the task abstractions in TSES and related have been carried out manually [22, 60, 98]. Therefore, we identified the need for a task characterization for TSES.

Regarding the methodology, there are many approaches to build task characterizations such as survey studies, interviews with domain experts, observational studies. According to Kerrarcher et al. the most common method is to derive from existing literature [47]. Our focus data type TSES is a hardly investigated data type with very little design studies [104, 68]. Additionally, we identified some design studies focusing on data types which have similar characteristics to TSES. A design study provides a domain-specific task characterizations. However, TSES is a data type commonly observed across a wide range of domains. Therefore, it is nice to have domain-agnostic data-specific task characterization. However, generalization of existing design studies is not possible due to limited work on TSES or related since these design studies doesn't cover wide variety of domains. In order to achieve domain-agnostic task characterization, we had to find another source for data collection covering a wide range of domains. Various fields are interested in TSES, so we identified the possibility of data collection for users. In the literature, several methods have been proposed for the collection of data from users. The lack of existing tools makes observational studies unavailable. In comparison to personal interviews, survey research has become increasingly popular since sampling over a large area is less expensive and much less time-consuming. Based on the problem reflection, we address the problem of lack of task characterization in the chapter 5.

Chapter 4

Part 1: A Large-Scale User Survey for Time-Stamped Event Sequence Data

4.1 Methodology and Experiment Design

4.1.1 Introduction

There is evidence that different factors play an important role in the characterization of problems 3.1. These will ultimately influence the decisions to be made in the design process of vis tools. But details of these factors are incomplete, and examining existing studies for identifying different factors is inefficient and also not an option for TSES due to the very little work focusing on this data type. In our study, we investigate these factors in the context of TSES. Our study builds upon two previous bodies of knowledge that will be leveraged. For one, existing related work on building general visualization tools or event sequence-specific visualizations, proposed design methodologies, and proposed visualization principles revealed three main broad aspects including **domain context & users, data, and tasks**. Second, data gathered from a prior study in our lab served as a sample and led us to identify relevant, possible, and missing aspects identified from the previous studies. This led us to discover **metrics** also as an important aspect. In the end, four main aspects were identified.

In section 3.1 we investigated deeper into those discovered aspects to identify possible influencing factors which resulted in 25 factors under these four main aspects.

- Domain Context & Users - Learn about problem, target domain, goals, existing practices and user characteristics
- Data Characteristics - Learn about general discipline that dataset describes
- Tasks - Learn about user data analysis and visualization needs
- Metrics - Learn about user interested features

Due to the need for an overview of current real-world TSES problem 1, we build a problem characterization for TSES focusing on discovered four main aspects and associated factors.

4.1.2 Research Questions

The reflection on related work and the identification of four aspects of factors lead to five principal research questions that this study shall answer. Four of them directly relate to the individual aspects, whereas the fifth question asks for inter-connections.

- **RQ1 : How do domain-related factors impact the problem characterization?** - Investigate how do factors such as context of the problem, domain goals, existing use of tools, and the user groups involved impact the problem characterization.
- **RQ2: How do diverse data characteristics impact the problem characterization?** - Investigate whether factors such as number of events and sequences, complexity, quality and wrangling issues, alignment, segmentation opinions impact the problem characterization.
- **RQ3: How do tasks impact the problem characterization?** - Identify the significance of different types of tasks identified from previous work, distributed across diverse event sequence cases. Further, we are interested to examine whether the granularities associated with the tasks have an influence on the significance of different types of tasks.
- **RQ4: How do metrics impact the problem characterization?** - Analyze the distribution of interest in metrics across different TSES cases and identify metrics that are frequently requested by users.
- **RQ5: What effect do domain context, user characteristics, data characteristics, tasks and metrics of TSES have on each other?** Identify whether there are any relationships between different factors associated with different aspects.

4.1.3 Study Approach

Our goal is to build a problem characterization for TSES. Existing work doesn't provide many insights about real-world problems on TSES which inspired us to use another source of information to gather TSES problems. Therefore, we gathered data on current real-world problems of TSES spans across a wide range of domains using a qualitative method. Observational studies, questionnaires, and personal interviews are the most common approaches to gather data on real-world problems [46]. We had to ignore observational studies in this case, since there aren't many data visualization tools focusing on TSES. However, this limits us assessing some user specific factors. In comparison to personal interviews, survey research has become increasingly popular since sampling over a large area is less expensive and much less time-consuming. This led us to develop a questionnaire in order to interview larger groups of people. Further, our approach is data-first since we approach people who are interested in TSES. Therefore, we conducted a user-based survey a data-first study to collect real-world problems span across a wide range of domains without limiting to only few domains

4.1.4 Questionnaire Creation

Following the aspects of features discovered (cf. Section 3.1), we created the questionnaire to study those above discovered aspects namely (1) domain context and user groups, (2) data characteristics, (3) tasks, and (4) metrics. Based on the existing work, we identified the need of a proper questionnaire design. Therefore, we will first explain how we design the questionnaire based on existing work recommendations, and then move on to its content.

General Design Structure

A number of studies have examined how the survey design affects the responses of the participants, which revealed that poor design adversely affects answers [89, 40, 79]. Therefore, we followed the recommendations and guidelines proposed in the literature [7, 55] in order to receive valid, high quality responses. Initially, we defined the overall structure of the questionnaire and then focused on the specific questions within it (top-down). Following are some recommendations we applied to our questionnaire based on the literature.

- **Introduction:** Our study includes an introduction that gives users a quick overview of TSES, terms and definitions useful, as well as our main purpose.
- **Visual Cues:** To reduce the language gap between domains, as well as to help users perceive questions easier, we always included visual cues throughout the questionnaire.
- **Informal Language:** In order to motivate users to fill out the questionnaire, we used simple informal language with the aim of making it easier for them to comprehend more easily without having to put extra effort.
- **Close-Ended Questions:** Following Weisberg et al.'s advice, we used closed-ended questions whenever possible, and provided open-ended questions only when necessary [101].
- **Bias:** Couper et al. discovered that the visibility of the answer influence the answer than the order. Therefore, we used radio buttons or checkboxes while avoiding dropdowns/ scrollable dropdowns [18].

In the next section, we will present details on the content.

Content Structure

Section 3.1 provided a detailed explanation of the process of identifying different factors which resulted in 26 factors. We chose to ignore cognitive traits, cognitive abilities, experiences, biases, and personal, social, and cultural characteristics, vocabulary due to data collection difficulties which resulted in 19 factors. In this section, we discuss factors that will be addressed in our study, We designed the questionnaire structured on four main aspects and all factors associated with them.

- **Domain Context & User Group:** - We have identified 6 factors under this aspect, including vocabulary, problem, goal, existing tools, current output achieved using tools and user groups To understand diverse contexts of real-world problems, we added questions on asking users to explain the ***problem*** and the ***goal***. Further, we asked experts to provide some information on ***current use of tools***, and the ***output achieved through these tools*** to understand the current work flows and how data has been used to solve their problems. ***User group*** also identified as a main influence factor and questioned in the questionnaire.

- **Data Characteristics** - We identified 10 influencing factors including availability of the dataset, size, quality, segmentation, alignment, metadata, timescale, dataset qualities, event sequence, event. The experts were asked to provide information about what is an ***event*** and ***event sequence***, as well as about ***metadata*** relevant to their problem. Regarding the quality, we used open-ended question to ask about existing data challenges and Likert scale questions provided to rate based on manual effort to process data and quality of the data. In TSES dataset size depends on ***no of events*** and ***no of event sequences***. Initial alignment strategy is another factor identified as influencing factor as it alters the insights that can be obtained. Data characteristics and the users' needs help to determine the ***best alignment method***, so we decided to ask users for their opinion. We have identified four main types of alignment strategies such as alignment by event [107], alignment by first event [99], alignment by start and scale span [107] and align by global time axis [32] provided to the user to help them pick the best alignment strategy for their case. Another factor influences the problem characterization is ***segmentation,*** which narrows the exploration scope than the whole event sequence. Two possible ways identified, either segmenting it by time points or events [32], or grouping it to several sub-sequences using metadata and questioned user's opinion using an open-ended question. Having a clear overview on the dataset often help understand the problem better. We identified some attributes that provide an overview of the data to the user, such as data: live, historical, event sequence: dense, sparse, regular, and irregular. Further, collisions and the frequency of collisions also helpful to factors to understand the data better.

- **Tasks** - We followed a two-fold strategy to gain an input from users on tasks.
  - Open-ended Questions - Questions specific to domain and the context such as goals, high-level user intents which require downstream interpretation.
  - Close-ended Likert Scale Questions - All possible tasks that we discovered from existing studies and prior interviews since there is no task taxonomy proposed in the literature that is specific enough to support TSES. ***Tasks*** listed in the survey include tasks on overview, detail analysis, metadata, sorting, trends, comparison, grouping, filtering, motif analysis, prediction, recommendation, annotation, detection. As we already stated, there are three granularities in the event sequence data such as event, event sequence, collection of event sequences. Therefore, we structure these questions in such a way that we can analyze how different granularities affect task distribution. For instance, comparison tasks were formulated in three ways: *compare events, compare event sequences, and compare event sequence collections*. To avoid encouraging people to pick options which they are not comfortable with, we always provided four Likert scale options, such as "not important",

"nice to have", "highly important", and "no opinion". [7, 55, 26]. We further gave users the option to mention missing tasks.

- **Metrics** - Our aim is to explore interest in metrics among different event sequence scenarios and discover metrics that are most frequently requested by users. A set of possible metrics were given to the user in Likert scale format. ***Metrics*** listed in the survey includes size, subsequence length, regularity, distance, acceleration, outliers, gaps, skewness, density, entropy, peakiness. As with tasks, metrics listed in the questionnaire are not complete, and similar Likert scale options are offered. Similar to the tasks, we further, gave user an option to mention missing metrics.

- **Demographics**: Finally, we asked users to provide their email addresses if they would like us to contact them back if necessary.

A summary of the finalized aspects & factors addressed in the survey is on Figure 4.1.

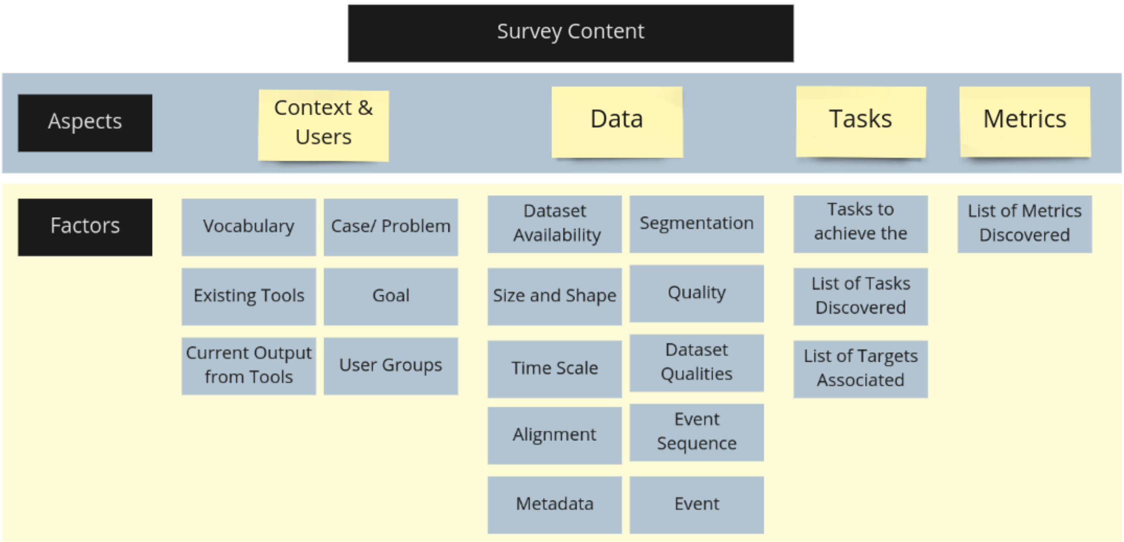


Figure 4.1: We identified 26 factors under four main aspects: domain context & users, data, tasks and metrics. This figure shows 19 factors we chose for the survey design.

4.1.5 Target Population & Data Collection

The target population includes people who are experts in different domains and either have TSES datasets or are interested in TSES datasets. Additionally, English proficiency was required.

A total of 70 survey responses were received spanning more than 25 domains, including sports, accidents, foods, technology, retail, health, and nature, generating a large analytic problem space in which 5 cases were excluded as invalid. The cases mentioned by participants were manifold, examples are *'eating doughnuts'*, *'food deliveries during lockdown'*, *'gaining Instagram followers'*, or *'purchasing airline tickets'*

4.2 Processing of Study Data

To conduct the analysis of the study data more effectively, we applied strategies to process open-ended and close-ended questions, initially. Since close-ended questions have categorical answers that do not require further processing, they were only restructured as necessary. However, open-ended answers needed further processing to translate these open-ended answers into categorical format for further analysis. Further, we derived new data considering relevant open-ended and close-ended questions. After encoding, we treated all the factors as closed ended questions. There are two main strategies that we can use for processing, such as open coding and focused coding [25].

4.2.1 Processing of Open-Ended Questions

Our goal was to encode open-ended answers into categorical formats to allow comparison between different answers as well as correlation with other factors. This is a multi-stage process which contains open-coding as the first step and then focused coding as the second step.

- **Step 1: Open Coding [25]**: After reading each answer, we encoded the data with one or many relevant codes that stand out. We followed the same process multiple times, and it is important to keep an open mind throughout the process. Once the commonalities between answers are noticeable, we grouped the open codes based on similarity to narrow down the categories. Group code should be always selected in a way that represent all the codes, and we needed to be careful to have consistent set of finalized codes.
- **Step 2: Focused Coding [25]**: The only requirement for this coding is to have a set of finalized codes that can be used for coding. After having the finalized set of codes, we re-coded the every answer using these finalized group codes. It is important to have a list of codes before starting the focused coding.

Figure 4.2 depicts the finalized codes for data centric challenges by grouping codes discovered using open coding.

4.2.2 Deriving New Data Factors

We generated new data with categorical values based on the answers provided by the user. Two main approaches followed. (1) We identified several existing characterizations proposed in the literature that can be used to encode the data and proceed with focused-coding, (2) Analysis of existing columns revealed some new characterizations that fit for the encoding which is similar to open-coding. Therefore, we aimed to use these strategies to derive new data.

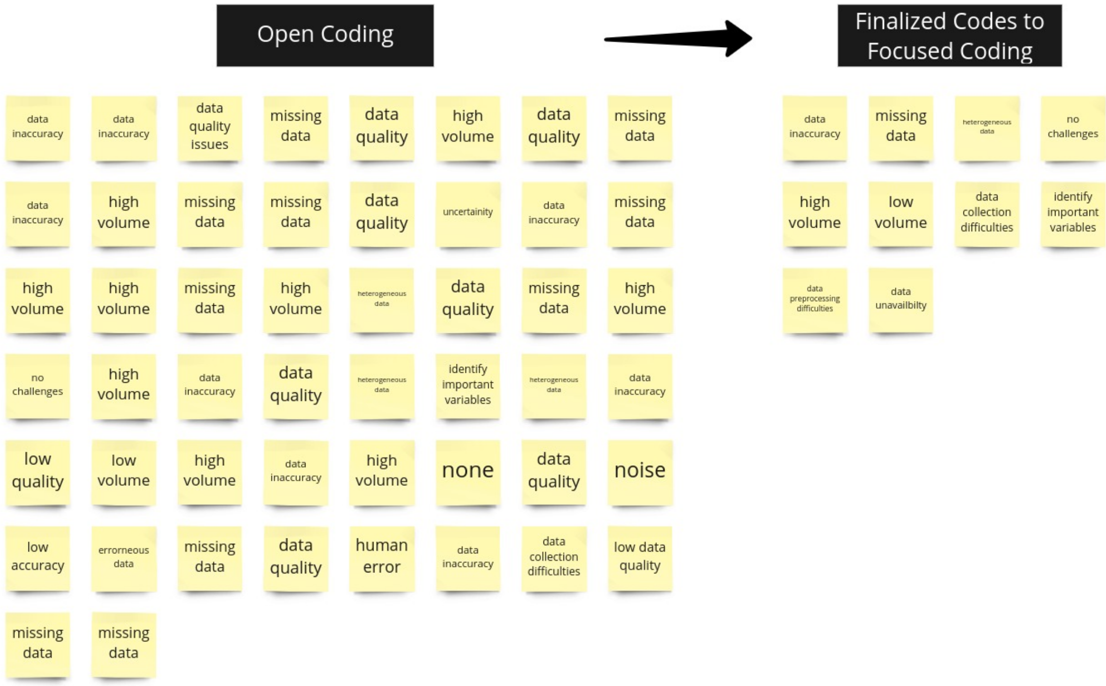


Figure 4.2: We used open coding on data gathered for open-ended questions, then we grouped similar codes together to find the finalize list of codes. These finalized list of codes used to do the focus coding on top of the same data.

New Data Factor	Existing Work Characterizations
Goal Type	Exploratory, Confirmatory, Presentation
User Type	Data Enthusiastic - Explorer, General Audience - Information Consumer, Researchers - Convey messages to experts
Problem Type	Myself, Many

Table 4.1: Existing characterizations from literature or revealed characterizations based on answers were used to derive new data factors. Goal type, user type, & problem type was generated.

4.3 Results Analysis

4.3.1 Analyze Distributions of Different Factors

Our strategy is to analyze the distribution of different factors in each aspect, and building upon this, answer the five research questions. Based on the gathered data, we generated charts which provide an overview of factors.

Domain Context & Users

We gathered data on domain and problem context and also on users in order to address the RQ1.

- **Domain**: Gathered data spanned across a wide range of areas. The result of the open coding and focused coding process revealed categories, as shown in Figure 4.3. The figure illustrates how different real-world problems spans across diverse areas and some example areas are *energy, transportation, health, sports, software/IT*. Based on the Figure 4.3 TSES are mostly observed in *leisure, health, sports, transportation, and software/IT* domains

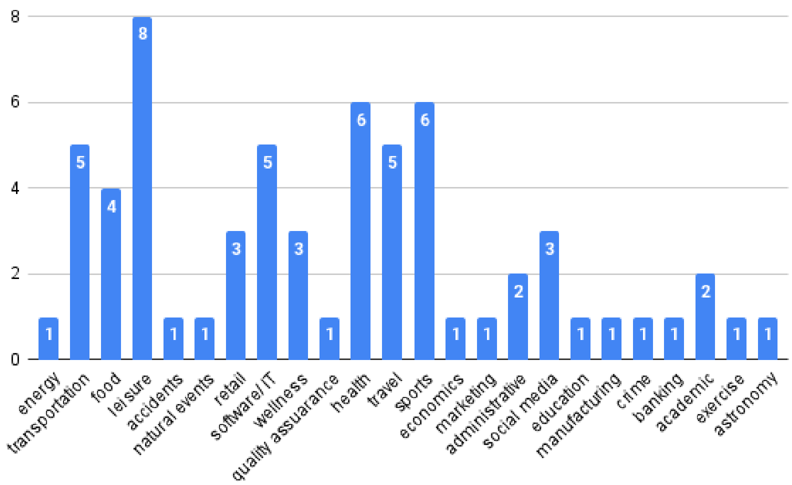


Figure 4.3: TSES spans across 24 domains. Based on the responses of the survey we extracted the domain from the problems to get a sense of the distribution of TSES data. Based on the figure TSES are mostly observed in *leisure*, *health*, *sports*, *transportation*, *and software/IT* domains

- **Goal Type**: Keim et al. have discussed 3 main goals of visualization including Exploratory Analysis (Aim to generate new hypothesis), Confirmatory Analysis (aim to test hypothesis), Presentation (aim to describe or present the results) [45] Schulz et al. used the same three goals to define the motive of tasks [85]. In the survey, we asked users to provide open-ended tasks for goals and tasks of this data analysis. Thus, based on the answers for both factors, we derived a new data category 'Goal Type' by assigning one or more of the above-mentioned goals. Figure 4.4 shows exploratory analysis is the goal of the majority of the problems, while there are few cases with the goal of confirmatory analysis and presentation

- **User Type**: Irissary et al. proposed three different categories where audience can be grouped into based on their goal, namely (1) own exploratory data analysis, (2) to convey a message to experts, (3) to tell a story to audience [39]. We used this characterization to derive a new factor "user type" by analyzing user, and the problem. Figure 4.5 illustrates the distribution of the user

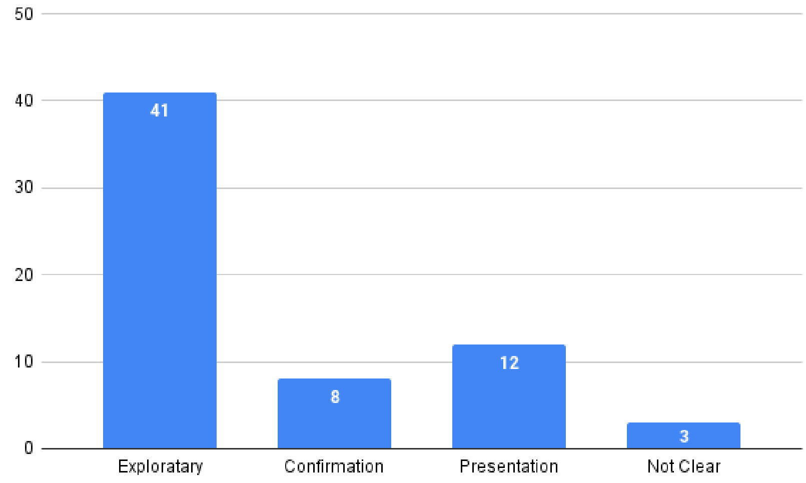


Figure 4.4: Distribution of goal types which we identified from existing literature. Majority of the TSES are exploratory goals while there are very few problems with presentation and confirmation goals.

types. The figure shows that a data enthusiastic is involved in 95% of the problems for exploration, which means more people are interested in data analysis, while considerable amount of users are interested in consumption of presented information. Based on these different cohorts, we identified the following characteristics of cohort.

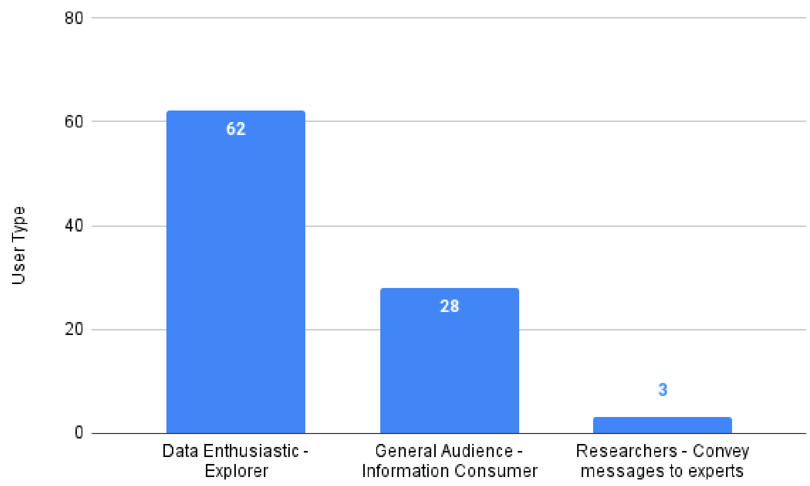


Figure 4.5: Distribution of different user types which we identified by focused coding *'user groups'* based on three main user types identified from the existing work [39]: (1) data enthusiastic - explorer, (2) researchers - to convey a message to experts, (3) general audience - consume information. The majority of users are engaged in exploratory data analysis, while a substantial portion is general audiences. It is very rare for researchers to use a tool to convey a message to experts.

- **Problem Type**: We identified user characteristics such as working memory, cognitive abilities are important factors that influence the vis design. We identified two problem types based on the target audience. 1) only the expert interacting with the system (2) Set of users interacting with the system. As shown in the Figure 4.6, the majority of problems' audiences fall under this category, while some problems are personal, and only experts will be using this tool.

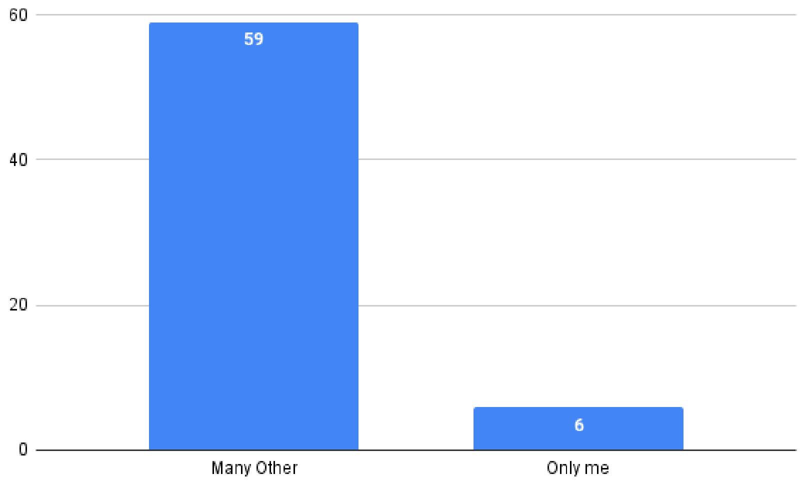


Figure 4.6: We identified two problem types based on the target audience. 1) only the expert interacting with the system (2) Set of users interacting with the system. Many users interact with the majority of the problems

- **Current Data Analysis Methods:** Figure 4.7 shows a summary of existing methods that are being used in the community to analyze TSES sequence data. As can be seen from the figure, most data analysis is done manually, which is both the simplest and the least optimal method. This reveals the interest towards TSES and also lack of existing work in TSES. Then there are cases where data is never analyzed, but is interesting. There's another considerable amount of cases which uses different existing vis tools for analysis. Finally, there are a set of cases which use other methods for analyzing data.

- **Output Achieved Through Current Method:** 4.8 shows the most common outputs of the existing tools: per category analysis, seasonal analysis, summary, statistical measures, and trend analysis. Further, there are some very specific results.

- **Dataset Availability**: In Figure 4.9, one third of those interested in TSES did not actually hold a dataset. This shows that there are people interested in TSES data problems even though they don't hold a dataset by hand, or they don't know how to collect the data.

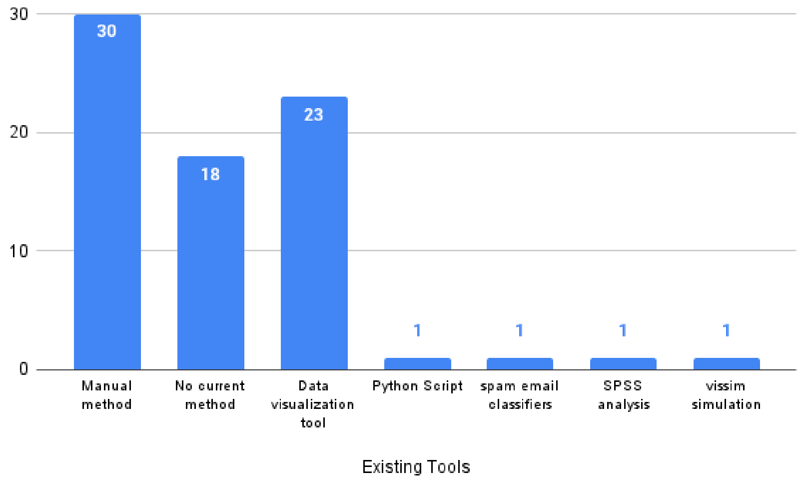


Figure 4.7: Different existing tools/ techniques are using in different problem context to solve their problems. While most cases are analyzed manually, a significant number are analyzed with existing visualization tools. There are a set of users who are interested in analyzing their data for the first time. There are also other methods that can be seen in some contexts besides the three main ones mentioned above.

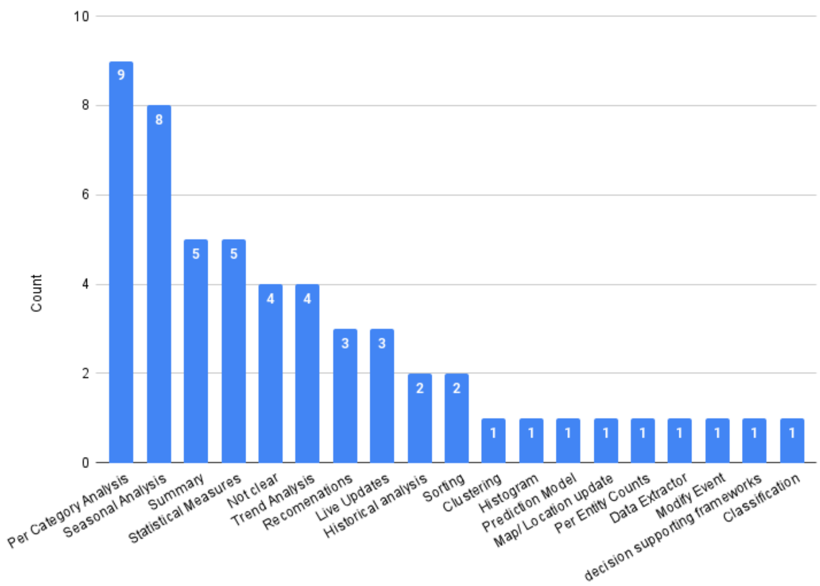


Figure 4.8: Output achieved from existing tools are mostly on per group analysis, seasonal analysis, summary, statistical measures, trend analysis. And also there are some other very specific results.

Data Characteristics

We gathered data on diverse characteristics of the dataset.

- **Size:** Figure 4.10 illustrates summary of the number of event sequences and events. Further,

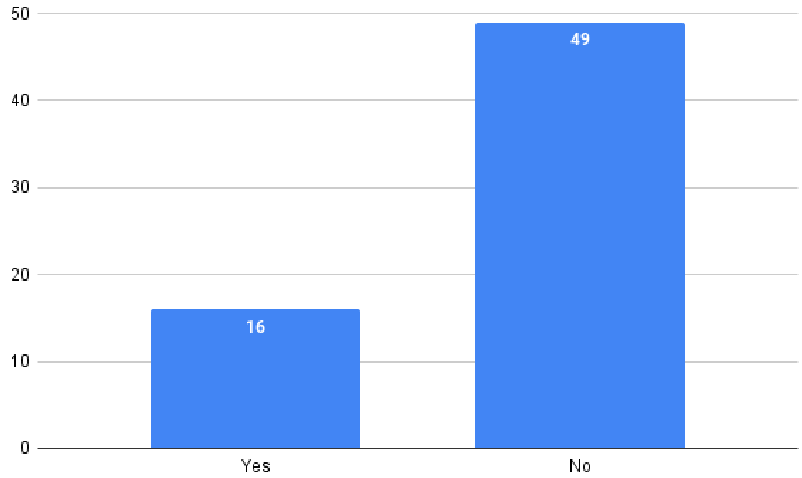
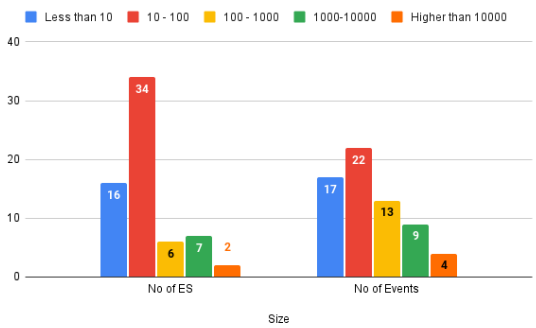


Figure 4.9: There are many participants who doesn't hold real data while there is a set of participants who are interested in analyzing TSES even they don't have a dataset by hand

Table 4.2 represents how the number of event sequences changes with the number of events. Most datasets number of event sequences and number of events fall within the 10-100 range. This shows most of the datasets contain fewer data point, which means sparse datasets, but there can be some rare cases where datasets contain very large volumes of data where event count and the event sequence count both have more than 10000 count. Further, mostly none of the described event sequences vary between 0-100 while event count can be varied within 0-10000.



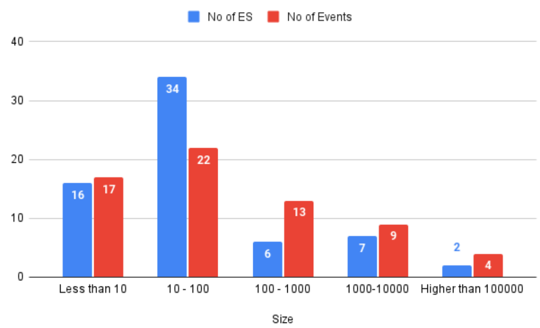


Figure 4.10: Size Distribution: **Left**: chart represents the distribution of no of event sequences and distribution of no of events. **Right**: Comparison of sizes of No of event sequences and No of events.

- **Dataset Characteristics:** Figure 4.11 shows the distribution of different characteristics. The fig-

No of Events \ No of Event Sequences	Less than 10	10 - 100	100 - 1000	1000-10000	Higher than 10000
Less than 10	7	9	0	0	1
10 - 100	7	10	4	1	0
100 - 1000	2	8	1	2	0
1000-10000	0	6	0	3	0
Higher than 10000	0	1	1	1	1

Table 4.2: Dataset sizes distribution of 65 survey responses using event count & event sequence count. Majority of the event counts and event sequence counts are within 10-100. It is possible to have datasets with more than 10000 events and event sequences, but it is extremely rare.

ure indicates that most analyses are based on historical data, but there are times when analysis of live data is also needed. In addition, there are a number of datasets that include both historical and live data. Regarding other characteristics, there are almost identical counts for each of the other four characteristics: sparse, dense, regular, and irregular. This reveals that it is equally likely to have a sparse, dense, regular, or irregular dataset.

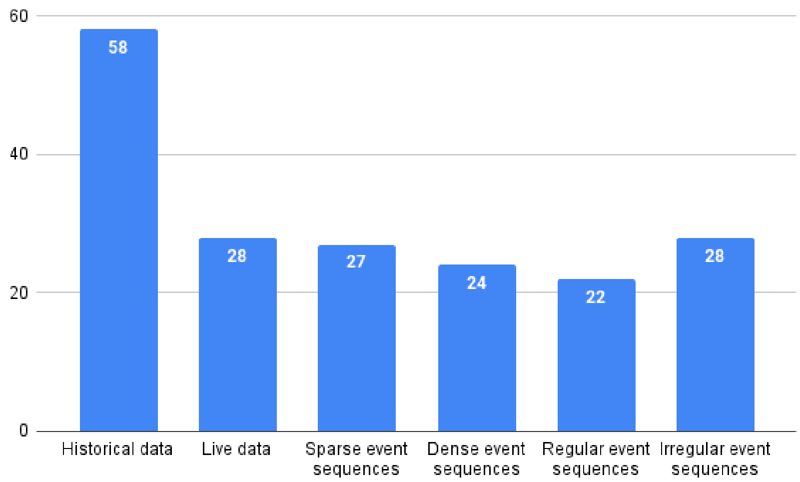


Figure 4.11: Different types of characteristics of the events and event sequence provides an overview of the data. There can be cases whether all these applies. Around 90% of the cases hold historical data while around 45% of the cases wants to analyze live data. It is both in some situations. Regarding other characteristics, there are almost identical counts for each of the other four characteristics: sparse, dense, regular, and irregular.

- **Data-Centric Challenges:** Figure 4.12 shows different data-centric challenges when dealing with TSES data. We used open coding and focused coding 4.1 to extract these challenges. Issues due to data inaccuracy is the mostly seen data-centric challenge for many participants. Significant portion of challenges are due to missing data and high volume. Further, very few sets of participants encounter with problems due to heterogeneous data, data collection difficulties, data processing difficulties.

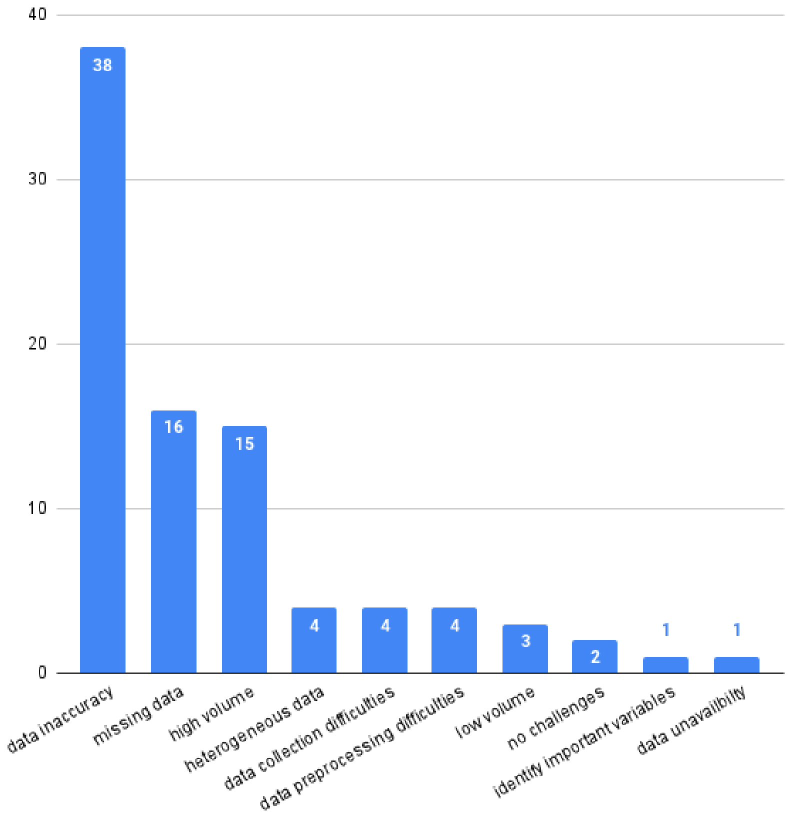


Figure 4.12: Majority of the data-centric challenges when dealing with data are due to data inaccuracy. In other words uncertainty, noise, human error. Significant portions of the data are missing, and the volume of the data is high. Participants also faced challenges related to heterogeneity, data collection, and data preprocessing. There are only two instances where data handling is not a challenge.

- **Alignment:** Figure 4.13 confirms the opinions for the alignment of the timescale. More than half of the study participants wanted the event sequences to be aligned with the global time axis, so that they can compare different timestamps of events across multiple event sequences. Very few study participants were not clear about this alignment option. Another important findings of this graph is some domain experts want to have support for multiple alignment strategies for different type of tasks. Therefore, this reveals an additional user need *"support for multiple alignment strategies"* which was not revealed in the user tasks or anywhere.

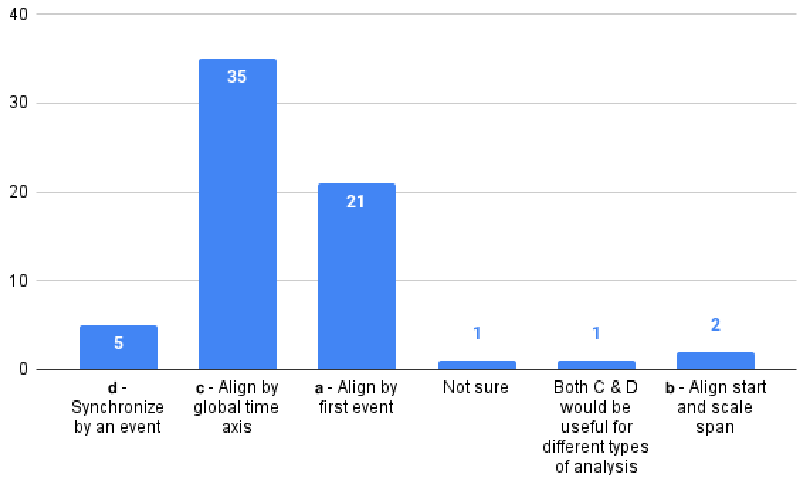


Figure 4.13: Majority of the participants are interested in aligning by global time axis. There are relatively high portion of participants who are interested in aligning by the first event. Some participants have shown interest on align by start and scale span option. Additionally some participants are interested in multiple alignment options for different types of analysis.

- **Segmentation:** This is an open-ended question and we used open coding initially and then focused coding to identify different segmentation types. Figure 4.14 shows a summary of open and focused coding. Based on the graph, only few people are not interested in segmentation. Further, we identified two types of segmentation which are partitioning using metadata [23] and segment using an event or a specific date, or season [23]. Moreover, we found that people are interested in partitioning within partitions, a concept we call "hierarchical partitioning". A majority of people are equally interested in partitioning or segmenting, while another group is interested in both.

- **Data Processing:** Figure 4.15 illustrates the rating for quality of the data and the manual effort that need to process the data. Table 4.3 shows the contingency table to identify the correlation between these two columns. We can observe an inverse relationship between data quality and manual effort. The table and figure illustrate that there are only a limited number of datasets that need no processing, and there are also very few datasets that demand significant processing effort. There are, however, very few exceptions, such as situations where data quality and manual effort are both high.
- **Collisions and Frequency of Collisions:** Table 4.4 shows that in around half of the cases multiple events can occur at the same time, and can occur frequently. Therefore, when picking techniques, it is important to ask from the expert whether they are interested in collisions or whether it is okay to ignore.

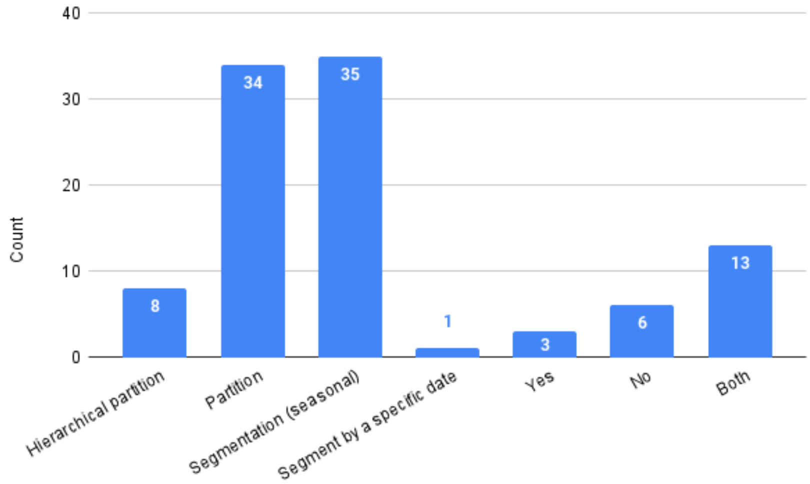


Figure 4.14: A similar number of participants are interested in segmentation or partitioning. Segmentation refers to splitting the event sequence or dataset by time segments while partition refers to partitioning event sequence or dataset into disjoint subsets. Small set of participants just mentioned whether they were interested in analyzing subsequences using yes and no without specifying the method. Only few people are not interested in analyzing subsequences. There's around 10% of the participants interested in both. Partitioning within partitioning, or hierarchical partitioning is of interest to some people.

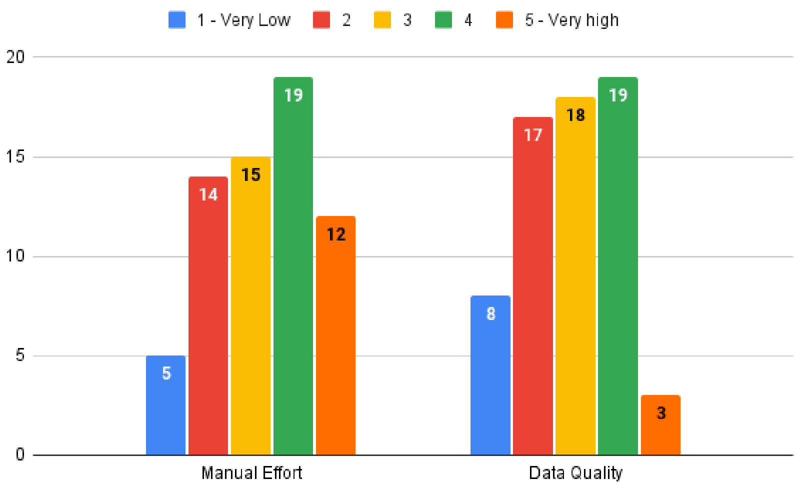


Figure 4.15: In a scale of 1-5, the majority of respondents rated manual data processing effort between 2-4 and data quality between 2-4. Only a few datasets require too little manual effort. There are also very few datasets with high data quality.

Tasks

We followed two main task collection steps. (1) Open-ended questions and (2) Likert scale questions. Analysis on Likert scale questions will be the focus in this part of the study, however, in

	1 - Very Low	2	3	4	5 - Very high
1 - Very Low	0	2	3	1	2
2	1	2	5	7	2
3	3	5	4	4	2
4	1	5	3	5	5
5 - Very high	0	0	0	2	1

Table 4.3: Contingency table of manual effort and data quality based on distribution of 65 survey responses

Collisions Happen Often \ Collisions	Yes	No
Yes	29	3
No	12	21

Table 4.4: Contingency table of collision and frequency of collsion based on distribution of 65 survey responses. In most cases, collisions don't occur unless they are frequently occurring.

depth analysis on open-ended questions related to tasks will be done in the part two. To avoid bias in answers for open-ended questions of tasks, we introduced Likert scale questions after open-ended questions.

- **Task Distribution**: We provided users with a set of possible tasks using a Likert scale and Figure 4.16 shows a summary of the tasks distribution based on 65 survey responses. More than 75% of the problems rated *'identifying trends'* as the most important task, indicating it is the most important task among diverse domains. Further, around 13 tasks out of lists 23 tasks were rated as 'highly important' by more than 50%. There is only one task rated below 25% as 'highly important'. However, it is also rated as a 'nice to have' by many people.
  Further, we examined whether there are tasks that are not important at all for some case while it is highly important for some other cases. Figure 4.17 depicts the summary of percentages for each task based on ratings. It was interesting to see, that no case think that *'extracting features'* as an unimportant task. 'Modify ES' and 'Detect Duplicates' stays at the bottom, being the least important tasks for many people. Only the task 'Modify ES' is rated more unimportant than highly important.

- **Influence of Associated Targets**: Figure 4.18 depicts how different targets influence the main action in separate figures for easy comparison within one action. Figure 4.19 shows the same data in one scale for easy comparison across both actions. Observations reveal that some comparison tasks are more important than all filtering tasks, while some comparison tasks are less important than all filtering tasks. This shows that importance is determined by the target, not just the action. It is not possible to rate the importance of the filtering or comparison action alone, but it can be rated along with the target granularity.

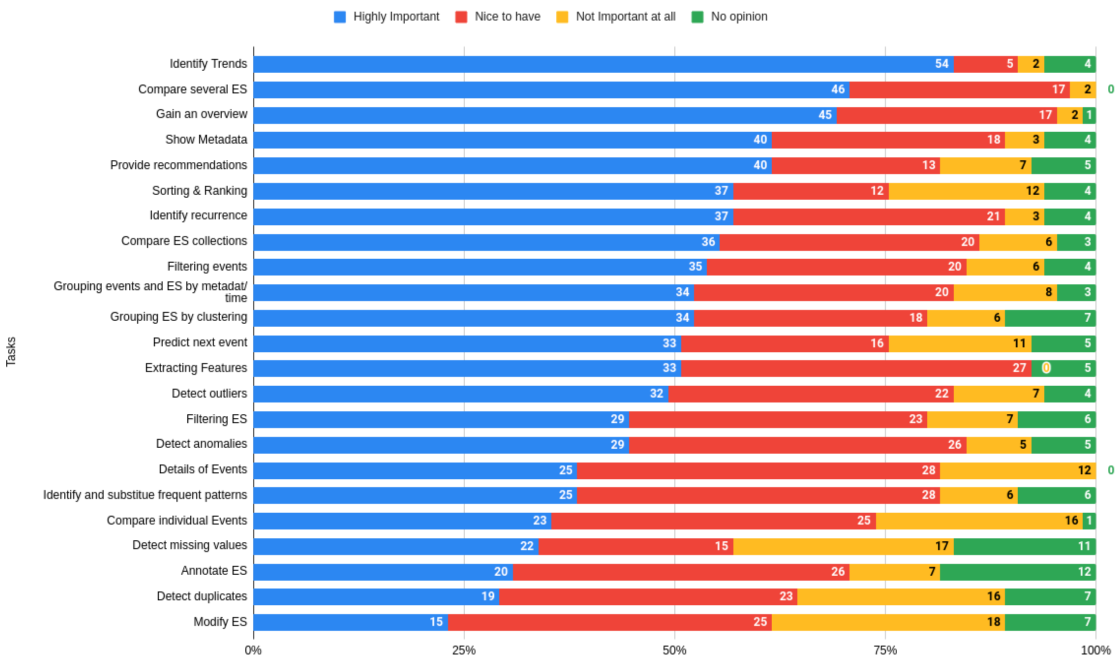


Figure 4.16: Y axis shows anticipated tasks for TSES we provided in the survey as a Likert scale. These horizontal color bars indicate what percentage of participants rated each task as highly important, nice to have, not important at all, or not important at all. Most of the tasks have more than 25% high important rate while only 'Modify ES' task have a low high important rating compared to others.

Metrics

- **Metrics Distribution**: Similar to the tasks, we provided users with a set of possible metrics using a Likert scale and Figure 4.20 shows a summary of metrics distribution based on 65 survey responses. Only the metrics 'Size' and 'Regularity' were rated as highly important by more than 50% of the responses. In the previous section, tasks were rated as highly important by many people, which is not the case with the metrics. However, many metrics were rated as at least nice to have in many responses. Figure 4.21 depicts that all the metrics have a higher highly important rate than not important at all rate. Therefore, it is evident that the people don't have a clear idea on what metrics are and the benefits of them to decide whether this is really important. However, they still think metrics will give some important insights, therefore many have selected 'nice to have' for most of the metrics.

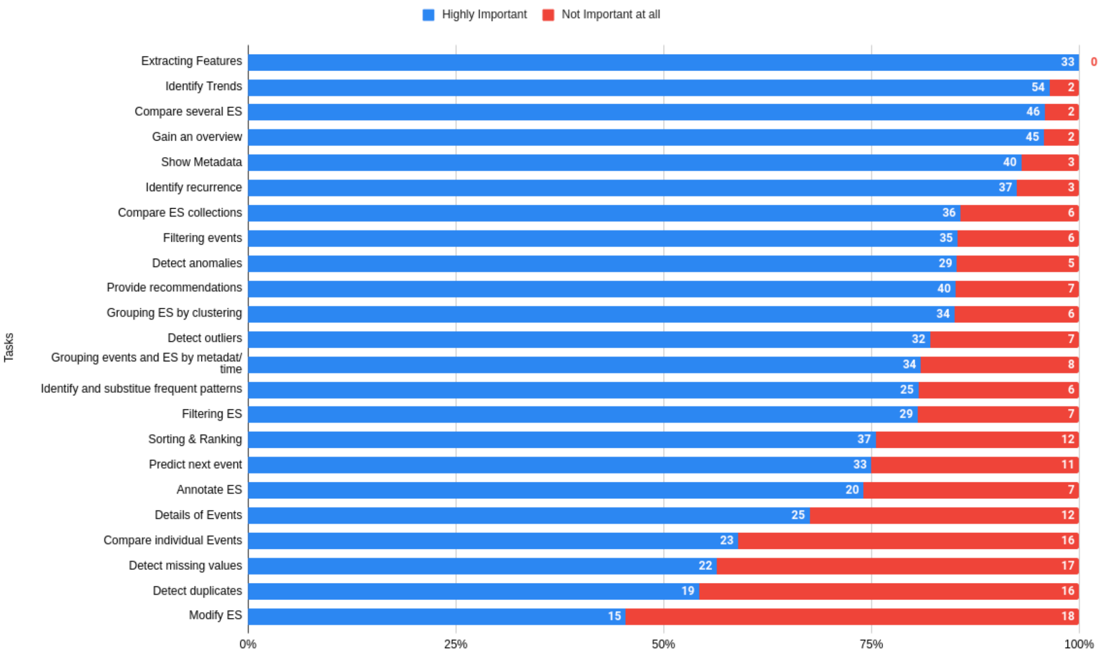


Figure 4.17: Y axis represents anticipated tasks for TSES and horizontal color bars represent 'high important' rate vs 'not important at all' rate of tasks according to study participants responses. Only one task, 'Modify ES,' has a lower importance rate, revealing that most tasks have higher important rates.

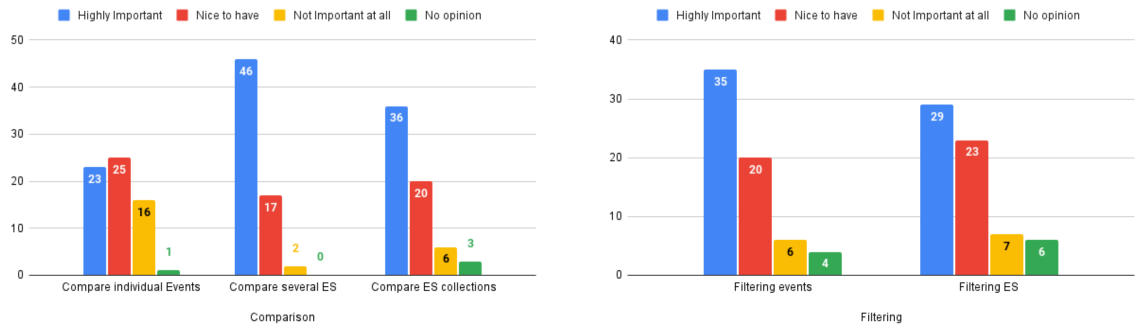


Figure 4.18: Different granularities associated with each task influence the importance of the tasks. Two main tasks as examples **Left**: Comparison tasks, compared to other comparisons, comparing several event sequences is highly important **Right**: Filtering tasks, In terms of filtering, events are more important than event sequences but does not show much difference between the two compared to comparison

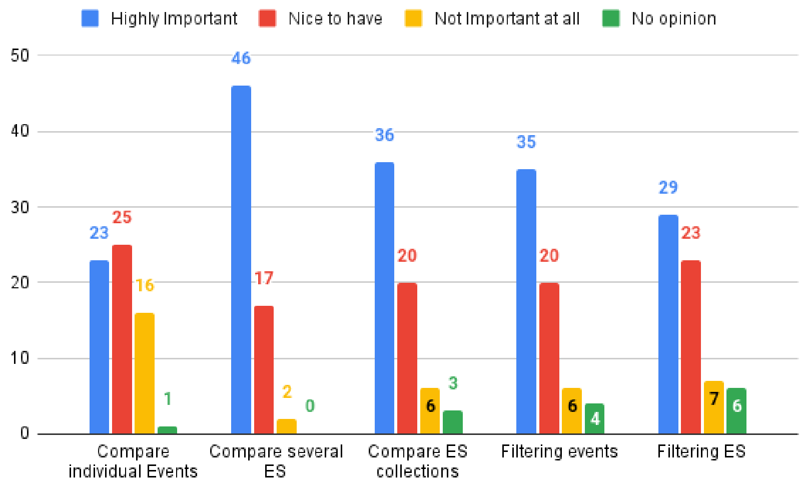


Figure 4.19: The filtering and comparing of all granularities on the same scale are for easy comparison. A comparison of event sequences is more important than both filtering tasks, while a comparison of individual events is less important than both filtering tasks.

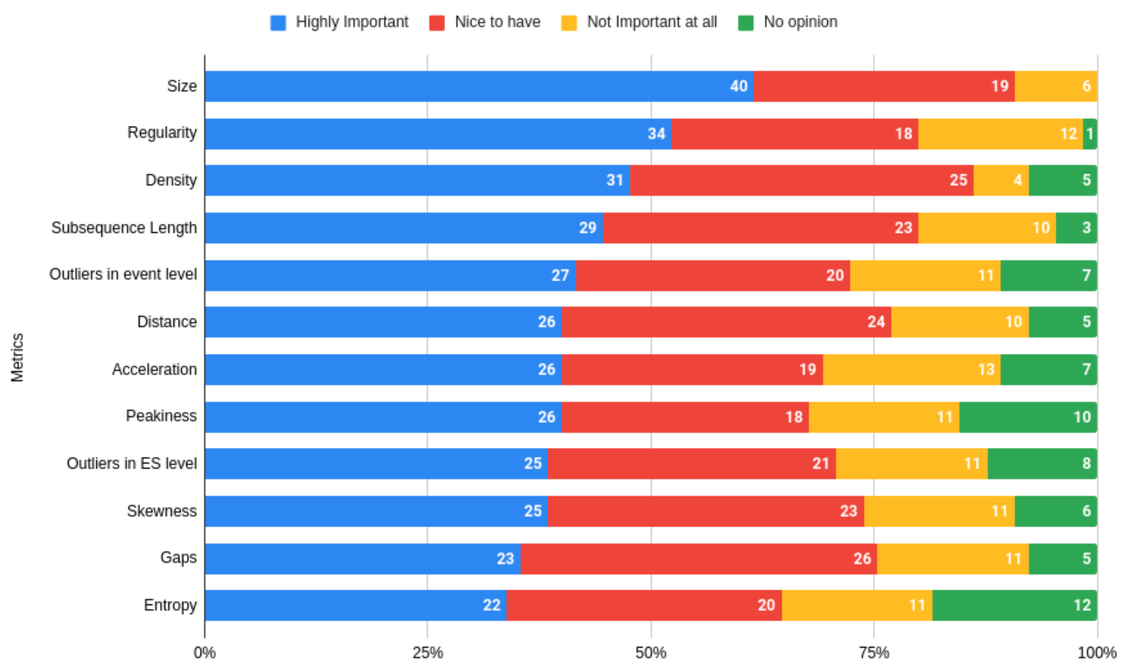


Figure 4.20: Y axis shows anticipated metrics for TSES we provided in the survey as a Likert scale. These horizontal color bars indicate what percentage of participants rated each metrics as highly important, nice to have, not important at all, or not important at all. Most of the metrics have more than 25% 'high important' rate. A significant number of participants have rated 'nice to have' as well.

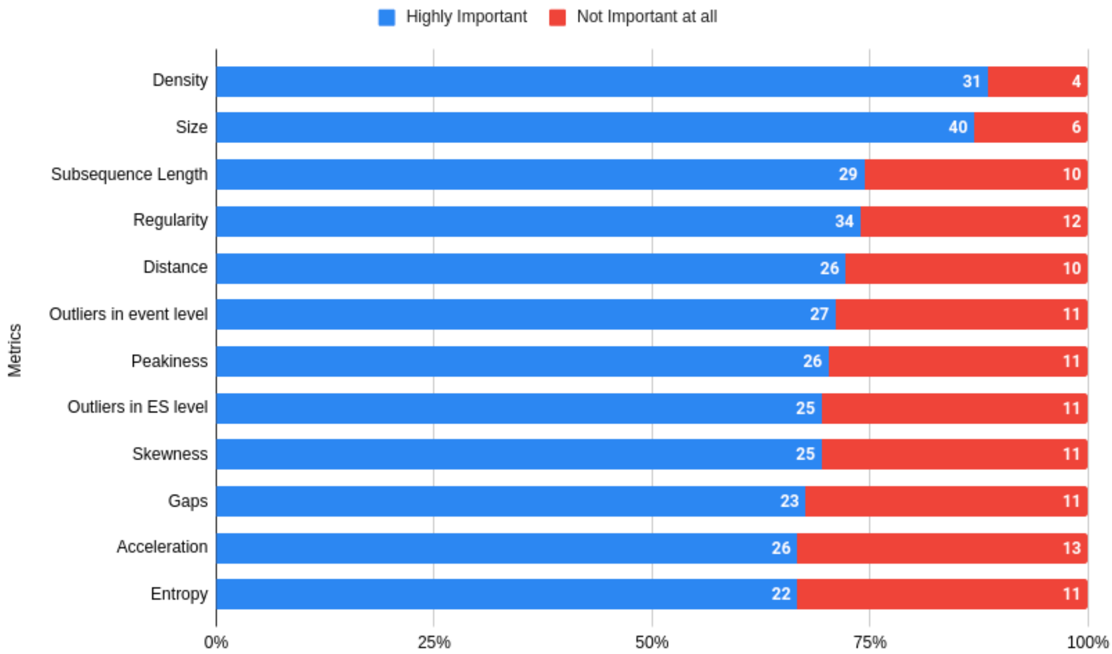


Figure 4.21: Y axis represents anticipated metrics for TSES and horizontal color bars represent 'high important' rate vs 'not important at all' rate of metrics according to study participants responses. All the metrics have a high important rate compared to not important at all.

4.3.2 Correlation Analysis

Interesting Insights

We have discovered some interesting insights based on analyzing relation between different factors. We followed two strategies, (1) An exploratory relation-seeking tool that was previously published by the lab members [9], (2) Correlation analysis using python notebook. In part 1 and part 2 we will present the results of these strategies.

Part 1 :

The tool bins all answers of the questionnaire according to all studied factors, numerical attributes are binned with a domain-preserving binning strategy. Based on statistical testing, the graph-based visualization uses colors and connections to reveal relations between bins that are most striking, in the sense that distributions differ most from a balanced expectation values. Figure 4.22 revealed the following findings.

- There is a weak relation between no of events per event sequence and importance of tasks.
- More events mean large data volume, which means difficult to analyze each data point separately.

- There is a high degree of correlation between grouping tasks and identify recurrences tasks when the number of events is higher. This reveals that in order to cope with high data volume challenges, these experts are more likely to be interested in tasks that sharpen their analytical focus.

- Another interesting finding is that grouping is an important analysis task, regardless of which technique was used to group.

- Moreover, the extraction of features task is also seen as being of high significance.

- Comparing individual events is considered as highly important when the event count is relatively middle, not too high or too low, indicating individual events are more significant with a moderate volume of data.

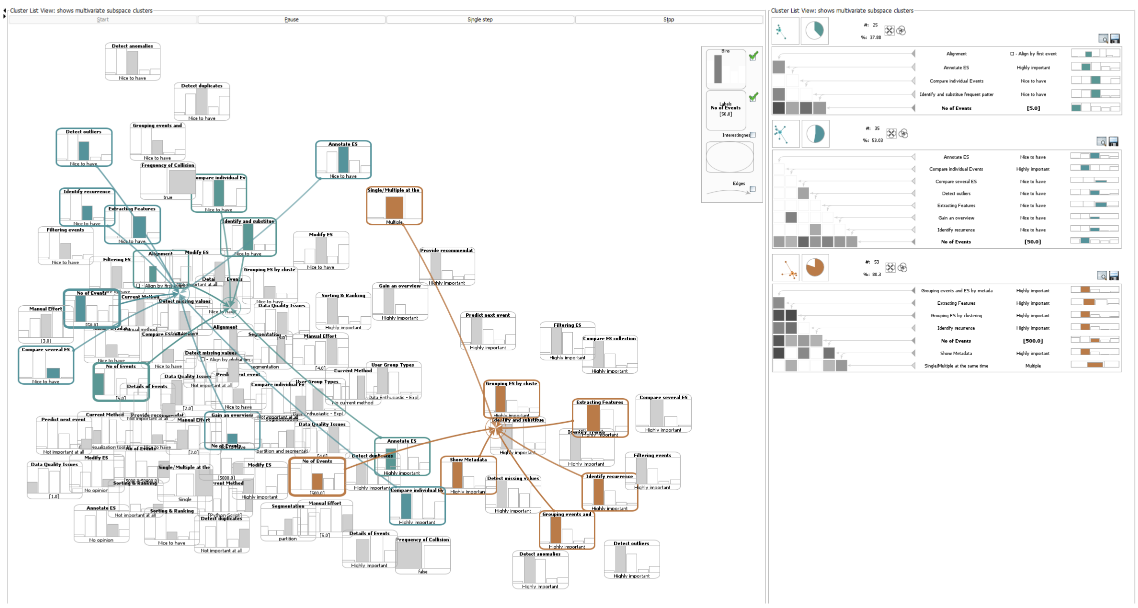


Figure 4.22: Correlation between number of events per event sequence and other tasks. There is a weak relation between event count and some tasks. Grouping tasks & Identify recurrence tasks are highly important when the event count is higher. (This figure was generated using exploratory relation-seeking tool 4.3.2)

Figure 4.23 indicates the following.

- we can observe of more task importance, when the manual effort is higher.

- and also have a more interest towards identifying outliers, missing data tasks with the increase of manual effort for data processing.

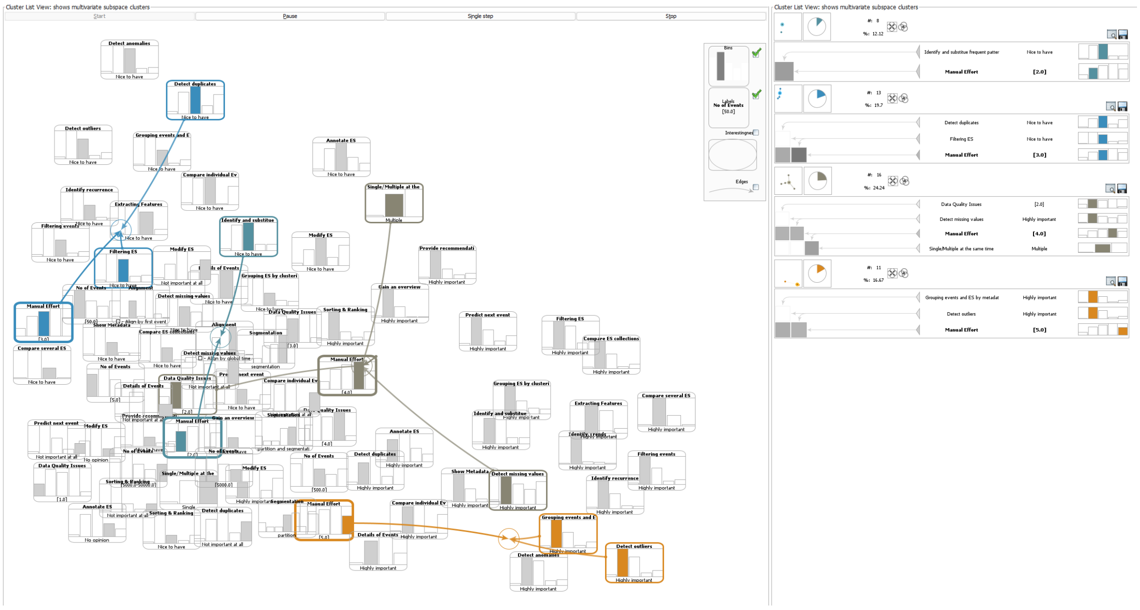


Figure 4.23: Correlation between manual effort and other tasks. It can be seen with the increase of manual effort tasks have a higher importance. (This figure was generated using exploratory relation-seeking tool 4.3.2)

Part 2 : The diagram was generated in two steps. After calculating the correlation between each column, we selected only those columns with at least one correlation greater than 0.6. Once more, this chart was generated based on extracted factors. Only the factors with more than 0.6 have a higher correlation. Figure 4.24 depicts correlations between sets of factors. We observed the following findings.

- There is a higher correlation between 'details of events' task and 'compare individual events'
- 'Identify & substitute frequent patterns' have a strong correlation to 'identify recurrence' task
- Detect anomalies and detect outliers also have a strong correlation.
- Detecting outliers task have a strong correlation to both event level and event sequence level outlier metrics
- Further, gaps and skewness also have a high correlation.

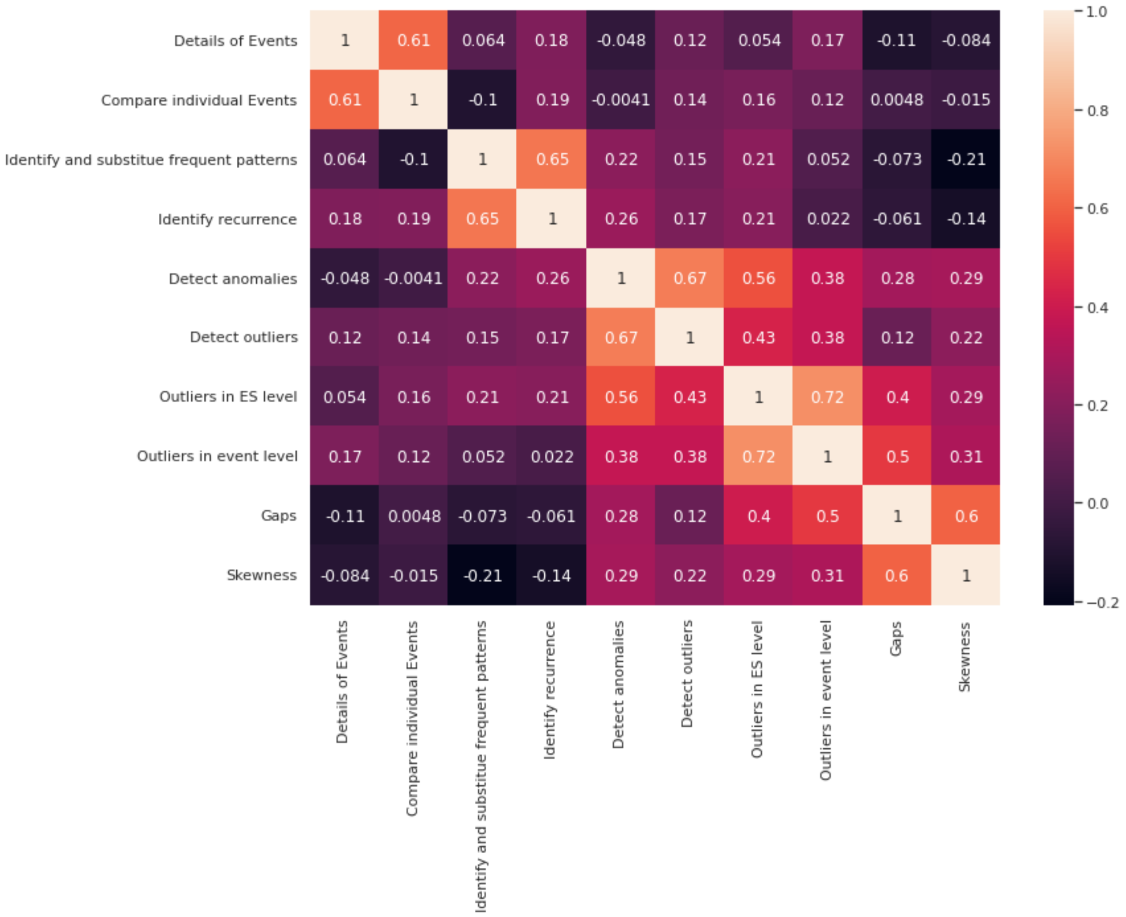


Figure 4.24: It shows different factors with at least one correlation greater than 0.6. Higher correlation can be observed between following: (1) 'details of events' and compare individual events' tasks, (2) 'Identify & substitute frequent patterns' and 'identify recurrence' tasks, (3) 'Detect anomalies' and 'detect outliers' tasks, (4) 'Detecting outliers' task and event sequence level outlier and event level outlier metrics, and (5) gaps and skewness metrics.

4.3.3 Correspondence Analysis

4.4 Overall Analysis

Answering RQ1: Based on the analysis of factors, we addressed RQ1 (cf. Section 4.1.2) by characterizing the real-world problems using *domain*, *goal type*, *user type*, *problem type*, *current data analysis methods*, *output achieved through the results*. Analysis of domain context and users showed that,

- answers of questionnaires were very heterogeneous since there was no single pair of answers that could be considered similar or even equal with some degree of abstraction

- 24 domains which are interested in TSES & this represents high diversity of domain interested in TSES.
- data enthusiast users are the most commonly encountered in TSES issues.
- the most common data analysis goal of TSES problems is exploration.
- two different problem types, including *'only the expert interacting with the system'* and *'set of users interacting with the system'* based on the number of users who interact with the tool.
- most of the TSES are currently analyzing manually

In summary, the context and users of the domain influence the characterization of the problem.

Answering RQ2: Based on the analysis of factors, we addressed RQ2 (cf. Section 4.1.2 by characterizing the real-world problems using *dataset size*, *dataset characteristics*, *data-centric challenges*, *alignment*, *segmentation*, *data processing*, *Collisions and Frequency of Collisions*. Analysis of data characteristics showed that,

- most datasets' number of event sequences and number of events fall within the 10-100 range.
- majority of the data for TSES are historical data. Further, there are problems which needs analysis of both historical and live data.
- the most common data-centric challenge of the analysis of TSES is data accuracy.
- majority of the participants are interested in aligning by global time axis.
- majority is interested in analyzing subsequences
- many datasets contains events with collisions and this collisions occur frequently.

In summary, data characteristics influence the characterization of the problem.

Answering RQ3: Based on the analysis of factors, we addressed RQ3 (cf. Section 4.1.2) by characterizing the real-world problems using *tasks*, *targets associated with tasks*. Analysis of tasks showed that,

- *'Identify Trends'* is the most important task among the majority of the cases.
- *'Modify ES'* is the least important task among the majority of the cases.
- Different granularities associated with the task affect the importance of the action.
- No one rated *'Identify Trends'* as not important.

Answering RQ4: Based on the analysis of factors, we addressed RQ4 (cf. Section 4.1.2) by characterizing the real-world problems using *metrics*. Analysis of metrics showed that

- *'Size'* is the most important metrics among the majority of the cases.
- Metrics have higher 'nice to have' rating, while task have more 'highly important' ratings.

In summary, metrics influence the characterization of the problem.

Answering RQ5: Based on the correlation analysis of factors (cf. Section 4.3.2) we addressed RQ5 (cf. Section 4.1.2).

4.5 Interesting Insights based on Survey Answers

In the questionnaire, we left out some points that could have been addressed, such as not only focusing on user groups, but also asking about their expertise in related domains, not only ask about whether there are collisions(multiple events occur at the same time), but also ask whether they want to emphasize it or ignore it. Lastly, we identified that one open-ended question was not clear due to too vague to the user. Different people have provided answers by thinking in different perspectives.

Chapter 5

Part 2: In-Depth Task Characterization for Time-Stamped Event Sequences

5.1 Methodology and Experiment Design

5.1.1 Introduction

There is no task characterization for TSES 3.2. We identified that domain-agnostic task characterization is more beneficial due to the diversity of domains in which TSES exists. Design studies provide domain specific task characterizations. However, generalization of existing design studies is not possible due to the limited number of design studies. This inspired us to extract tasks from users spans across a wide range of domains using a user-based survey. Further, we decided to compare these extracted tasks with the tasks already supported in the visualization literature by using design studies as another source of information. We aim to identify the tasks that have not yet been supported by the literature, and also the ones users are not aware of, but existing work already addressed. Our goal is to unify these two sources to build a domain agnostic task characterization. This part was collaboratively done with Clara Maria Barth, a student member of the lab, to stick into the basic principles of the field of human-computer interaction.

5.1.2 Research Questions

- **RQ1: Can tasks be derived from surveys be described in a common language?**
  Investigate whether we can have a refined list of tasks extracted from each source by grouping similar tasks by describing using a common language.

- **RQ2: Do common tasks exist?**
  Discover whether survey responses indicate that certain tasks are common while others are

more specific to certain scenarios, which will allow us to prioritize accordingly.

- **RQ3: How well do user study tasks match those from related design studies?**
  Examine if there is a gap between the user needs and the tasks addressed in the related literature by comparing tasks extracted from two sources. We wanted to identify tasks that are unique to user survey study and tasks that are unique to survey of design studies.

- **RQ4: Can two sources be unified?**
  Find out if it is feasible to unify these two sources. As there is no task characterization for TSES, we can contribute a task characterization to future research.

5.1.3 Study Approach

To address the research questions, we incorporated two complementary sources of information to reduce the risk of misclassifying and overlooking tasks [47].

- **User-Based Survey** - Generalization of data-first design study survey.
- **Survey on Design Studies** - Generalization of problem-driven design study survey

Next, we will present the process we followed, including data collection and processing of two study data.

5.1.4 Data Collection

Data gathered from two complimentary sources including user-based survey and survey on design studies. The process followed at each method will be presented under each study.

User-Based Survey

We already collected real-world problems using a user-based survey in order to develop a problem characterization. We used the same source to collect data for this study also. In the questionnaire, two type of questions were used for the task extraction. (1) Open-ended questions on goals and tasks that require downstream processing (2) Likert scale list of possible tasks to select how important each task. Likert scale question responses were already analyzed and presented the results in Part 1 4. Therefore, data gathered for the open-ended questions were used as one main source of data for this study.

Survey on Design Studies

This study is inspired by Lam et al.'s qualitative analysis of 20 IEEE InfoVis design studies [52], to use open coding for tasks related to TSES that have been extensively studied in literature in a variety of domains. In contrast to the task abstraction approach taken by Lam et al., we directly used the task abstractions provided with the design studies and survey papers. As only a very limited number of design studies on TSES exist, we chose papers in the realm of other time-oriented data as our primary source, while maximizing the variety across involved application domains and user groups.

The paper selection process involved several iterations:

1. **Preliminary Selection**: we searched for literature using keywords such as 'event-based data', 'temporal event data', 'time-stamped event sequences', 'temporal events' and 'event sequence' to locate papers related to TSES, leading to 55 papers.
2. **Secondary Selection**: We selected the subset of design studies and survey papers with clear summaries of task abstractions, finalized requirements, or state-of-the-art analytical tasks.
3. **Final Selection**: We examined each paper in depth and excluded papers with tasks or requirements that hide data-specific information due to a high degree of abstraction (cf. Section 2.4), resulting in 14 design studies[107, 104, 68, 16, 69, 57, 22, 60, 59, 36, 109, 93, 30, 98] and 2 survey papers [32, 75].

5.2 Processing of Two Study Data Sources

The two studies were conducted concurrently but independently until task synthesis phase in processing study data. To extract the tasks in each study, we used **coding**, a well-known qualitative research technique [27] to extract tasks from each source, followed by an **affinity diagramming** phase to group the extracted tasks to identify the most common tasks among them. To comply with the methodology of coding and affinity diagramming as proposed in HCI literature, a second person was involved with these two phases. This person was Clara-Maria Barth, a fellow student with expertise in HCI and VIS.

5.2.1 Coding

In this phase, RQ1 is primarily addressed. Both persons were responsible for creating new codes, reviewing existing codes, and commenting on any codes that were incorrect or not agreeable. This was an iterative process, and several discussions resolved many of the conflicting opinions. The third party opinion was used to resolve conflicts when it was impossible to resolve it ourselves. Ultimately, every code was refined iteratively until no conflicts existed. In spite of the fact that the user-based survey study and the survey on design studies seemed to be isolated, we coded them interchangeably in order to mitigate bias between the two studies.

User-based Survey

Each of the 70 user survey responses described a specific time-stamped real-world scenario, along with details about domain context, user goals, the data analysis tasks necessary to reach those goals, current data analysis methods, and tools available etc. More details on the survey is described in section 4.1.4. Further, participants provided information on event and event sequence also, useful for data characterization purposes. We extracted goals and analysis tasks provided by the user and applied hierarchical task abstraction [107] to each survey response to decompose domain-specific tasks into intermediate or lower level tasks [81]. Further, to find similarities and differences across domains, we translated all domain-specific tasks into abstracted tasks. We will refer to these abstracted tasks as codes. Figure 5.1 shows the coding process followed, Figure 5.2 presents the results.

Survey on Design Studies

In this study, we used the open coding [27] approach, which was more straight-forward than the previous coding strategy. We used abstract task descriptions summarized in the papers for coding. In the cases where task descriptions were too domain-specific, we converted into abstracted form when coding.

5.2.2 Task Categorization

Figure 5.3 shows the coding process followed, Figure 5.4 presents the results.

5.2.3 Initial Affinity Diagramming Phase

In the task categorization, we used a similar approach to open coding. Two authors involved with the iterative process of grouping the codes, reviewing the groups and also commenting in disagreements. Similar to coding, in the cases of opposing opinions that cannot be resolved, discussions opened to a third party opinion. In a later iteration, we split or merged those groups as necessary and further ignored the groups with only one task if they did not fall within the existing categories. For a group to be considered, there have to be at least two sticky notes, either from different survey responses in the user survey study or from different papers in the survey of design studies. Following several iterations, 33 unique tasks were identified from the survey responses, while 36 tasks were derived from the design studies by addressing RQ2.

User-Based Survey Study

1. The initial affinity diagram related to the user-based survey study is depicted in Figure 5.5.

2. Figure 5.2.3 summarizes the tasks extracted from the user-based survey study following the initial phase of the task categorization.

Survey of Design Studies

1. Initial affinity related to the survey of design studies is depicted in Figure 5.6.

2. Figure 5.7 summarizes the tasks extracted from the survey of design studies following the initial phase of the task categorization.

Refinement Phase

We also conducted third-party validation on all the extracted tasks. This reduced the number of tasks to 24 and 26 respectively by merging possible tasks and ignoring least relevant tasks.

5.2.4 Task Synthesis

The task synthesis unifies the previous two-sourced approaches. This is accomplished by joining the two spaces of abstracted tasks, represented by the previously created task categorizations of user-based survey and survey of design studies. We treated both sources as equally important and decided to create a unified list of tasks by translating all the tasks in to a common language that allows unification. Our criteria for unification was to focus on low-level and mid-level granularity of tasks, that each is not too coarse, and sufficiently fine-grained to be executed independently. We first started with the common tasks from both sources, and then moved to the unique tasks of each source that need further processing to create the unified task space. We followed an iterative process of merging, splitting and rewording to have a set of finalized groups that satisfies our criteria. By addressing RQ3, we identified 23 similar tasks, while 3 tasks only occur in design studies and 2 tasks only occur in user survey study which cannot be merged anymore as those tasks are distinct enough to represent each own action. Figure 5.8 shows the merged affinity diagram which we built from all the tasks that we identified from both studies. We provide a summary of the finalized tasks list of each source together with the information on task frequency in the table 5.1. Figure 5.9 compares the results of both task categorizations, emphasizing the relative frequency of every task per study. *'Merge'* and *'Analyze Fluctuations'* are the tasks that users have found that have not yet been addressed by the literature, but may be worth exploring in the future. There were several tasks in the existing work that no user addressed, such as *'Discover Causality'*, *'Save Subset of Interest'*, and *'Identify Distinct Entities'*.

5.3 Results Analysis

We consider both sources of information equally important because both are well-known We also discovered any task that only appear in one source cannot be represented by common tasks due to the fact that they are distinct from each other, making them more significant in the task analysis space. Table 5.1 shows that even the unique tasks also appears to be more important to more than one subject, while there are some tasks that appear in both sources at lower frequencies. Consequently, we identified the significance of each and every task and consider union of both sources as finalized task list resulted in 28 tasks.

The following section describes the tasks in detail.

Action / Objective	User Survey	Design Survey
Derive Metrics	39	12
Summarize	32	14
Group	29	11
Compare	26	12
Relate	23	6
Identify Extremes	22	7
Analyze Trends	21	3
Emphasize	16	8
Annotate	14	4
Show Details	13	14
Segment	12	5
Sort	12	5
Find Similar	11	9
Predict	11	2
Identify Motifs	10	10
Add/Modify	8	3
Filter	7	9
Analyze State Transition	7	4
Compare Threshold	7	4
Recommend	7	4
Gain Overview	6	7
Align	6	3
Detect Outliers/Anomalies	3	12
Analyze Fluctuations	6	-
Merge	3	-
Discover The Causality	-	7
Save Subset Of Interest	-	3
Identify Distinct Entities	-	3

Table 5.1: Comparison between task occurrences in the two studies: the 65 user surveys (US) and the 16 design studies (DS). Most of the tasks for TSES are common in both studies.

5.3.1 Overview of Tasks

In this section, we describe all the 28 tasks we found in greater detail in the order of frequency of the user survey.

1. **Derive:**
   General Description: By calculating numerical outputs based on data content & metadata characteristics, new insights about the data can be gained that can't be seen visually. At different granularities, metrics can be derived using content, metadata, or both, for events , event sequences , and group of event sequences . The derived results can be analyzed by the user, as well as can be used as input to perform other tasks, such as comparing and summarizing. The most commonly requested metrics by users include regularity, density, gaps, frequency, and event count. Additionally, many other metrics exist that are valuable but unrecognized by users.
   Examples: Frequency of events for each es, Calculate density for events in es.

2. **Summarize:**
   General Description : The majority of the event sequence dataset contains around millions of datapoints, which makes it impossible to discover interesting information. In order to provide users with an overall view of the respective event, sequence, or group, metadata, metrics and content can be used to summarize events , event sequences , group of event sequences . In addition, users are interested in statistical measures such as mean, median, variance, etc to better understand the underlying distribution of metadata and metrics of each entity.
   Examples : Show summary of events for each group (event count, regularity trend...), Show summary statistics for event sequences

3. **Group:**
   General Description: A user can reduce the number of events and event sequences by assigning them to groups based on content similarity, metadata, metrics, annotations and recommendations. Cluster algorithms will apply to group events based on similarity by selecting appropriate metrics. Analyzing group characteristics gives a clearer understanding of individual characteristics within each group, since each group represents its whole population.
   Examples: Group es based on derive metrics for each TSES, Group events by year

4. **Compare:**
   General Description:In order to understand the differences and similarities between different entities, it is important to compare them. Comparisons can be made on an event , event sequences , or group of event sequences level based on content, metadata, or metrics. Also, there were some tasks relating to comparing metadata at the event level to metadata at the dataset level which discussed comparisons at all granularities. The comparison task is useful to identify similarities and differences across the compared items.
   Examples: Compare patterns of events event sequences , Compare derived metrics for different event sequences

5. **Relate:**
   General Description: Identifying the relationship between two things enables users to explain the reasons behind certain observations in the dataset. A relation can exist between any of the following: metadata, content, metrics, patterns, time, trends. It is supported either at event sequence level granularity or group of event sequences level.
   Examples: Identify relation between two metadata for each event sequence Identify which metadata mostly affect the event frequency of the event sequence.

6. **Identify Extremes:**
   General Description: Metrics and metadata are the only areas where extreme values can be identified. In addition to identifying most frequent entities, users were interested in identifying least frequent entities as well.

Examples: Identify time periods of high frequency of event segments for each TSES, Identify cluster patterns which are more frequent.

7. **Analyze Trends:**
   General Description: A common task identified among users is identifying trends of event sequences or group of event sequences based on numerical metadata or derived metrics. Further, users are particularly interested in identifying daily, monthly, and seasonal trends.
   Examples: Find the trend of event frequency over the years for each event sequence , Analyze trends of metadata with 2 week time gaps for each.

8. **Emphasize:**
   General Description: In large data volumes, finding interesting entities is often hard to do visually. By emphasizing interesting entities, it is possible to grab the attention of users at any time. The task can be only performed at events , event sequences , group of event sequences based on some criterion. Examples for criterion from the user studies are metadata, metrics, trend turning points, user defined threshold.
   Examples : Highlight event segments with high density, Highlight events with specific metadata values.

9. **Annotate :**
   General Description : An annotation refers to adding additional information to specific event , event sequence or group of event sequences . It is also supported for metadata. We identified two types of annotation, one where users are manually adding annotations, and the other is when annotations are automatically generated based on content, metadata and metrics and also there were cases requested trend change pattern.
   Examples : Label segments using density, Label events based on metadata values

10. **Show details:**
    General Description: Users are interested in analyzing specific event sequence in detail, especially if the number of the sequences and/or events is large. They are interested in more details on either content or the metadata.
    Examples : Show metadata of events, Inspect details of event sequence.

11. **Segment:**
    General Description: Users were interested in analyzing different segments of the same event sequence so they could compare them. While some users are interested in splitting the sequence based on time segments such as day, week, month, etc, some users are interested in splitting the sequence based on feature values, such as density, frequency etc. Additionally, there are situations where users want to split the whole dataset into two segments using a specific date so that they can analyze each segment independently and compare it to the other.

Examples : Segment in to two groups using a certain date, Segment event sequence by gaps

12. **Sort:**
General Description: The users wanted event sequences to be sorted based on metadata values as well as metric values. There were also requests for sorting on based on a combination of metadata and metrics values. It allows users to identify an interesting event sequences for detailed analysis while discovering new horizontal and vertical patterns at the same time. However, this is only supported for group of event sequences .
Examples : Sort event sequences . by metadata attribute, Sort by metrics

13. **Find Similar:**
General Description: Find similar events, event sequences, groups of event sequences based on selecting or formulating a query using event sequences, metadata, and predominantly: features.
Examples : Find similar event sequences based on metadata

14. **Predict:**
General Description: Most users want to predict the next event or the metadata based on previous events of the same type or on similar events in order to have a prior understanding of upcoming events. In addition, some users are interested in predicting the likelihood of the next event.
Examples: Predict next event metadata based on derived metrics of each event sequence .

15. **Identify/Simplify Motifs:**
General Description: Identify and simplify motif patterns in event sequences or groups of sequences based on event sequences or group of event sequences. This task includes the term pattern, always when referring to sequences or subsequences. As pattern is a high-level concept that can be applied to any finding, we resolved this ambiguity by stringently using the term motif for (sub-) sequences, as it is often done in the sequence data mining community. Simplification refers to the idea to substitute a motive by a (visual) placeholder whenever the motif occurs and is useful when complex datasets are to be simplified.
Examples: Select and identify similar patterns in different segments.

16. **Add/Modify:**
General Description: Most of the tasks identified under this category are related to modifying or adding new metadata for existing events or TSES or group of event sequences . In addition, if anyone is interested, they can add events , event sequences , or group of event sequences , although the last two are not feasible.
Examples: Let user assign metadata to events.

17. **Filter:**
General Description: Filtering supports users to filter out irrelevant information by keeping only relevant information. Users can focus only on a few events or event sequences based on their interests without digging into larger datasets. USers can filter events , event sequences , group of event sequences based on metadata, metrics or time period.
Examples: Filter event sequence by time period.

18. **Analyze State Transition:**
General Description: Analyze the transitions of states changing over time, e.g., to understand the reasons for the behavior of sequences.
Examples: Identify metadata values transitions of event sequences with reducing trends.

19. **Compare Threshold:**
General Description: It was a common task identified among users to filter query data against a threshold based on a comparison between metrics and metadata. Additionally, the user can compare expectations with metrics such as event count and filter TSES those that meet them.
Examples: Filter events by metric threshold.

20. **Recommend:**
General Description: Users want to get recommendations on the best time for the next event based on a selected set of metadata and metrics. Further, some people are interested in metadata recommendations. This task is supported at the event level only.
Examples: Provide recommendations based on metadata in each event sequence .

21. **Gain Overview:**
General Description: An overview of all the data points was requested by the users so that they could identify high level horizontal and vertical patterns, sparse dense regions, and compare all the event sequences in the overview while the summarize task provides a more compact overview of the data. The whole overview with metadata distribution is also requested by some users. This is only supported at the group of event sequence level.
Examples: Detailed event overview with metadata.

22. **Align:**
General Description: Aligning event sequences can be done based on several techniques. This is an important task that allows easy comparison between event sequences. Especially, when the event sequences are not of the same length and the user wants to compare them, aligning is very important. This task is only supported at group of event sequences .
Examples: Align event sequences by global time axis.

23. **Detect Outliers/Anomalies:**

General Description: Outlier is an ill defined term. Users, however, are interested in detecting events, ESs, or groups of ESs that are far apart. Outlier detection can be done visually or based on derived metrics. Anomaly detection is more identification of patterns of events that deviate from the significant majority of the events. This task is identified only at the es granularity
Examples: Identify outlier events using irregularities, Identify irregularities since event is periodic.

24. **Analyze Fluctuations:**
    General Description: Identify changes in trends and metrics over time. This task was only identified by users.
    Examples: Identify change of trend of events over whole dataset

25. **Merge:**
    General Description: Merging several segments/event sequences to calculate metrics for the merged event sequence, find the trend line with absolute time. This task was discovered by users only.
    Examples: Merge all event sequences

26. **Discover the Causality:**
    General Description: The effects of one event or metadata change on the next event(s)/metadata. This was a task discovered from previous literature.
    Examples: Identify causes of specific metadata values

27. **Save Subset of Interest:**
    General Description: Users are interested in saving intermediate results from previous analysis tasks such as clustering, filtering. Discovered only from the previous literature.
    Examples: Save clusters as metadata

28. **Identify Distinct Entities:**
    General Description: Users are interested in finding unique events, event sequences, groups of events, and metadata. This was also only identified from the previous literature.
    Examples: Show unique state/cluster transition patterns

5.4 Application of Task Overview: Comparison with ”Likert Scale Tasks” in the User-Based Survey

Three sources of tasks have been identified, two based on user feedback and one based on existing literature: (1) Open-ended questions of user-based survey, (2) Survey of design studies, and (3) Anticipated tasks provided as Likert scale. Using the first two sources, we systematically extracted 28 tasks and presented the results in the previous section. In the third approach, we provided a list of anticipated task as Likert scale questions in the user-based survey to determine the most important tasks. These tasks list was generated based on six qualitative interviews with domain experts focusing on TSES in a prior study and findings from existing literature without following any systematic process. We already compared results of first two sources to come up with 28 final task list. Our goal here is to compare these two sources with the third source 'Likert scale' questions to gain some new insights. We used Likert scale with four different options to let user pick either opinion without being bias. However, for the analysis purposes, we are considering only the 'highly important' answers. We use the following three approaches to comparison.

- Set comparison: Examine if task exist in both or one.
- Frequency Comparison: Compare frequencies of each task using three sources.
- Rank Comparison: Investigate the order of tasks in the two Rankings

5.4.1 Set Comparison between Different Sources

There is already a common vocabulary for tasks extracted from the first two sources, unlike Likert scale questions. As a result, we attempted to map all Likert scale questions to these identified tasks to make the comparison easier. Table 5.2 shows the summary of the results, and it reveals that our anticipated task list is not complete. Moreover, it seems that only 14 of the discovered tasks have been anticipated before this study, which ultimately reveals 28. It was a $n : 1$ mapping, revealing that Likert scale tasks are in a low granularity than the tasks in the task characterization. Further, we identified two tasks which were not in the final task characterization, including 'Detect missing values' and 'Detect duplicates'. One possible reason could be these task has been ignored due to low frequency of occurrence.

5.4.2 Frequency Comparison between Different Sources

Earlier, we presented a summary of frequency of the first two sources, and here we are mainly focusing on Likert scale questions. There is a 1:n relationship with the discovered tasks and the Likert scale. Due to this, when we calculated the frequency, we always counted as 1 if either task is rated high importance. As an example, both grouping by metadata and grouping by clustering can be considered grouping. In table 5.3, we show a summary of percentage occurrences of each task. Figure 5.10 compares the relative frequency of three sources per each task.

We provide a summary of the finalized tasks list of each source together with the information on task frequency in the table 5.1. Figure 5.9 compares the results of both task categorizations, emphasizing the relative frequency of every task per study. On average, the percentage of task rated with 'highly important' are higher in Likert scale compared to the other two studies. Two main reasons could be user unawareness of what is possible, while the other could be user tend to go all possibilities without focusing their actual needs.

All Tasks	User based Survey	Survey of Design Studies	Likert Scale Tasks (Anticipated Tasks)	Likert Scale Tasks
Derive Metrics	•	•	•	Extracting Features
Summarize	•	•		-
Group	•	•	•	Grouping Events and ES by Metadat/ Time, Grouping ES by Clustering
Compare	•	•	•	Compare Several ES, Compare ES Collections, Compare Individual Events
Relate	•	•		-
Identify Extremes	•	•		-
Analyze Trends	•	•	•	Identify Trends
Emphasize	•	•		-
Annotate	•	•	•	Annotate ES
Show Details	•	•	•	Details of Events, show metadata
Segment	•	•		-
Sort	•	•	•	Sorting & Ranking
Query	•	•		-
Identify/Simplify Motifs	•	•	•	Identify and Substitue Frequent Patterns, Identify Recurrence
Predict	•	•	•	Predict Next Event
Filter	•	•	•	Filtering ES, Filtering Events
State Transition	•	•		-
Compare Threshold	•	•		-
Recommend	•	•	•	Provide Recommendations
Add/Modify	•	•	•	Modify ES
Align	•	•		-
Gain An Overview	•	•	•	Gain an Overview
Detect Outliers/ Anomalies	•	•	•	Detect Outliers, Detect Anomalies
Analyze Fluctuations	•			-
Merge	•			-
Discover The Causality		•		-
Save Subset Of Interest		•		-
Identify Distinct Entities		•		-
Detect Missing Values			•	
Detect Duplicates			•	

Table 5.2: Comparison between task existence in the three studies: the 65 user surveys, the 16 design studies, anticipated tasks Likert scale. Most of the anticipated task are found from the both desgin studies and user survey study except two tasks. More than anticipated tasks are found from the user survey study and the survey on design studies

Most of the Likert scale have rated as 'highly important' by at least 25% of the users. *Compare and group* tasks have rate more than 80% of the users. The results of the Likert scale indicate that tasks such as *recommend, gain overview, predict and analyze trends* occur very frequently, whereas these tasks show low frequency in other sources.

Action / Objective	User Survey Final Tasks	Design Study Final Tasks	Anticipated Tasks
Derive Metrics	60	75	51
Summarize	50	88	-
Group	45	69	68
Compare	40	75	84
Relate	36	38	-
Identify Extremes	34	44	-
Analyze Trends	33	19	84
Emphasize	25	50	-
Annotate	22	25	31
Show Details	20	88	74
Sort	19	32	57
Segment	19	32	-
Find Similar	17	57	-
Predict	17	13	51
Identify/Simplify Motifs	16	63	62
Add/Modify	13	19	24
Filter	11	57	68
Recommend	11	25	62
Analyze State Transition	11	25	-
Compare Threshold	11	25	-
Gain Overview	10	44	70
Align	10	19	-
Analyze Fluctuations	10	0	-
Detect Outliers/Anomalies	5	75	54
Merge	5	0	-
Discover The Causality	0	44	-
Save Subset Of Interest	0	19	-
Identify Distinct Entities	0	19	-
Detect Missing Values	0	0	34
Detect Duplicates	0	0	30

Table 5.3: Comparison between task occurrences in the three studies: the 65 user surveys, the 16 design studies, anticipated tasks Likert scale. On average, the percentage of task rated with 'highly important' are higher in Likert scale compared to the other two studies. The results of the Likert scale indicate that tasks such as recommend, gain overview, predict and analyze trends occur very frequently, whereas these tasks show low frequency in other sources

5.4.3 Ranking Comparison between Different Sources

Finally, we assessed the ranking of the tasks based on frequency of each source. In the table 5.4, tasks have been ranked descending, by the frequency of occurrence for all three sources. The most common tasks among all three sources are *'Group', 'Compare', and 'show details'* and results are presented in the Figure 5.11. *Derive metrics and Summarize* also highly relevant based on first two main sources.

User Survey (Extracted Tasks)	Design Study Survey (Extracted Tasks)	User Survey (Anticipated Tasks)
Derive Metrics	Summarize	Compare
Summarize	Show Details	Analyze Trends
Group	Derive Metrics	Show Details
Compare	Compare	Gain Overview
Relate	Detect Outliers/Anomalies	Group
Identify Extremes	Group	Filter
Analyze Trends	Identify/Simplify Motifs	Identify/Simplify Motifs
Emphasize	Find Similar	Recommend
Annotate	Filter	Sort
Show Details	Emphasize	Detect Outliers/Anomalies
Sort	Identify Extremes	Derive Metrics
Segment	Gain Overview	Predict
Find Similar	Discover The Causality	Detect Missing Values
Predict	Relate	Annotate
Identify/Simplify Motifs	Sort	Detect Duplicates
Add/Modify	Segment	Add/Modify
Filter	Annotate	
Recommend	Recommend	
Analyze State Transition	Analyze State Transition	
Compare Threshold	Compare Threshold	
Gain Overview	Analyze Trends	
Align	Add/Modify	
Analyze Fluctuations	Align	
Detect Outliers/Anomalies	Save Subset Of Interest	
Merge	Identify Distinct Entities	
	Predict	

Table 5.4: Comparison between task ranking in the three studies: the 65 user surveys, the 16 design studies, anticipated tasks Likert scale. The most common tasks among all three sources are 'Group', 'Compare', and 'show details'. 'Derive metrics' and 'Summarize' also highly relevant based on first two main sources

Figure 5.1: Two people iteratively worked on coding 70 user-based survey responses. Five responses were excluded because they were invalid. Using hierarchical abstraction, codes were extracted based on domain context, goals, tasks, and definitions for events and event sequences. To ease the use of codes throughout the process, every code is labeled with the survey response ID.

Figure 5.2: Result of the coding process of the 70 survey responses. On average, 6-10 codes (green) have been identified per response (purple).

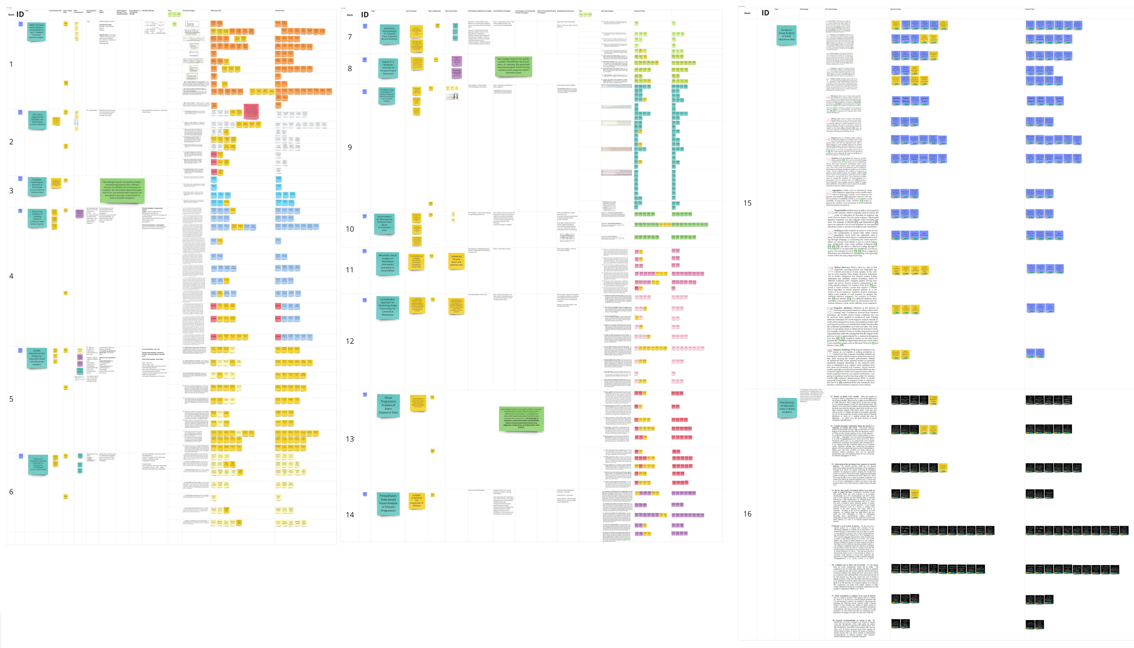


Figure 5.3: The coding was done iteratively by two people based on tasks extracted from 14 design studies and 2 survey papers. The codes were extracted from tasks listed in design study papers either at a domain-specific or abstract level, and open coding was used since the extracted tasks were already mid- or low-level. We followed a similar coding process for each design study, and we used distinct colors for each to assist us in the next step.

Figure 5.4: Codes extracted from each of the design studies, color-coded per study. The number of codes per design study is considerably higher than the number of codes per user survey (cf. Figure 5.2)

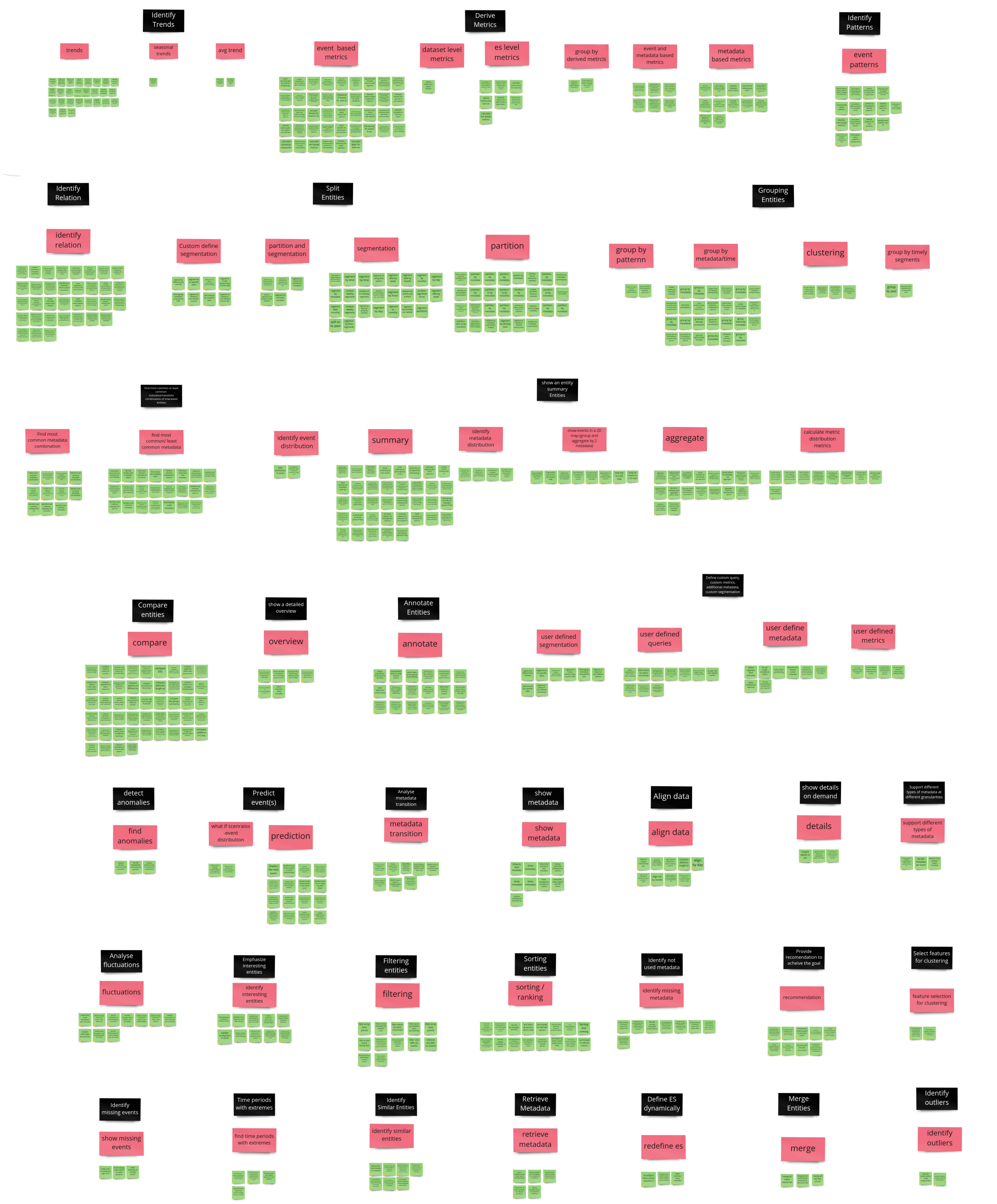


Figure 5.5: Based on the coding of the user-based survey study, we grouped the codes according to their content. Two authors are involved in the iterative process. Additionally, we assigned meaningful task names based on the set of codes. This affinity diagram represents 33 tasks derived from the initial task categorization phase.

Action/ Task	Derive Metrics	Show an entity summary (show overview)	Compare entities	Identify most common/ least used metadata combination	Split ESS	Group Entities	Identify Relation	Identify Trends	Emphasizing interesting entities	Annotate entities	Identify Patterns	Sorting ESS	Show metadata	Predict event(s)	Define expectation	Identify similar entities	Search by query (query entities)	Filtering entities	Custom define metadata	Analyse metadata transition	Provide recomendations	Custom define segmentation	Align ESS	Analyse fluctuations	Merge ESS	Define custom metrics	Support different types of metadata at different granularities	Show details on demand	Abstract entities	Identify Outliers	Detect annomalies	Define ES dynamically (reconfigure overview)	Select metrics for clustering
Occurence (number of surveys find with each tasks)	46	33	26	25	22	22	22	21	15	14	12	12	11	11	10	9	9	7	7	7	6	6	5	5	3	3	3	2	2	2	2	2	2

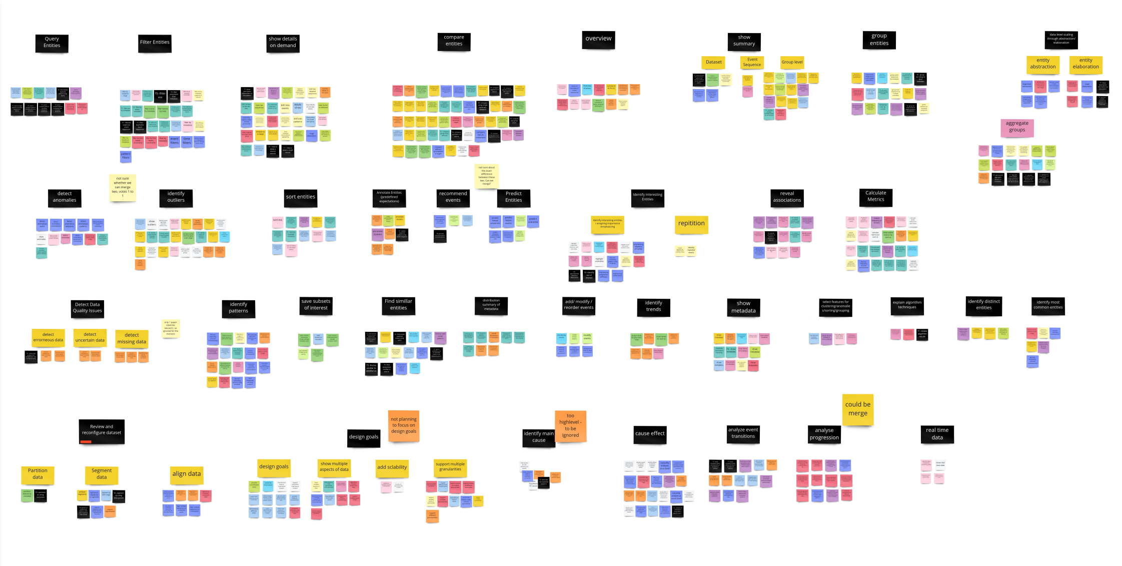


Figure 5.6: Based on the coding of the survey on design studies, we grouped the codes according to their content. Two authors are involved in the iterative process. Additionally, we assigned meaningful task names based on the set of codes. This affinity diagram represents 36 tasks derived from the initial task categorization phase.

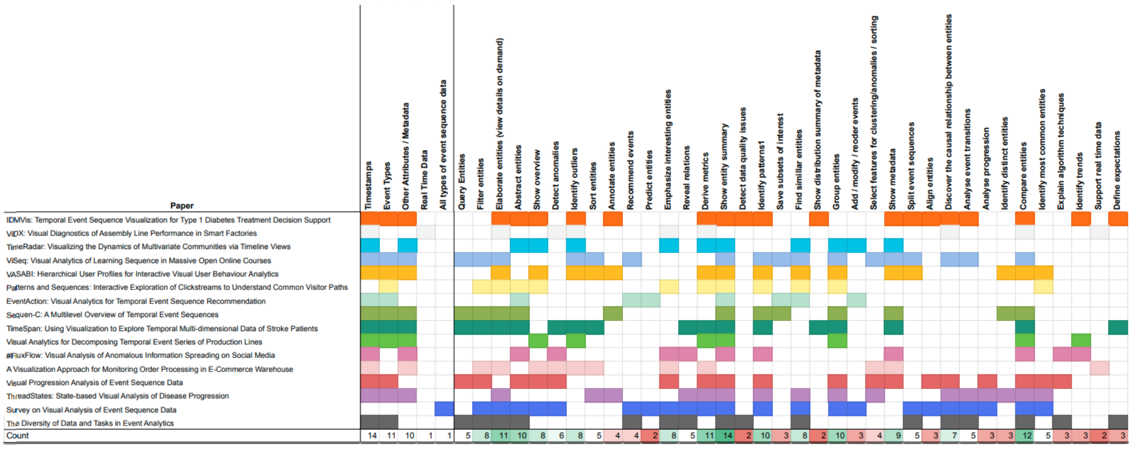


Figure 5.7: As a result of the study of design studies affinity, 36 tasks (code groups) were identified, along with how many unique papers have already discussed each task.

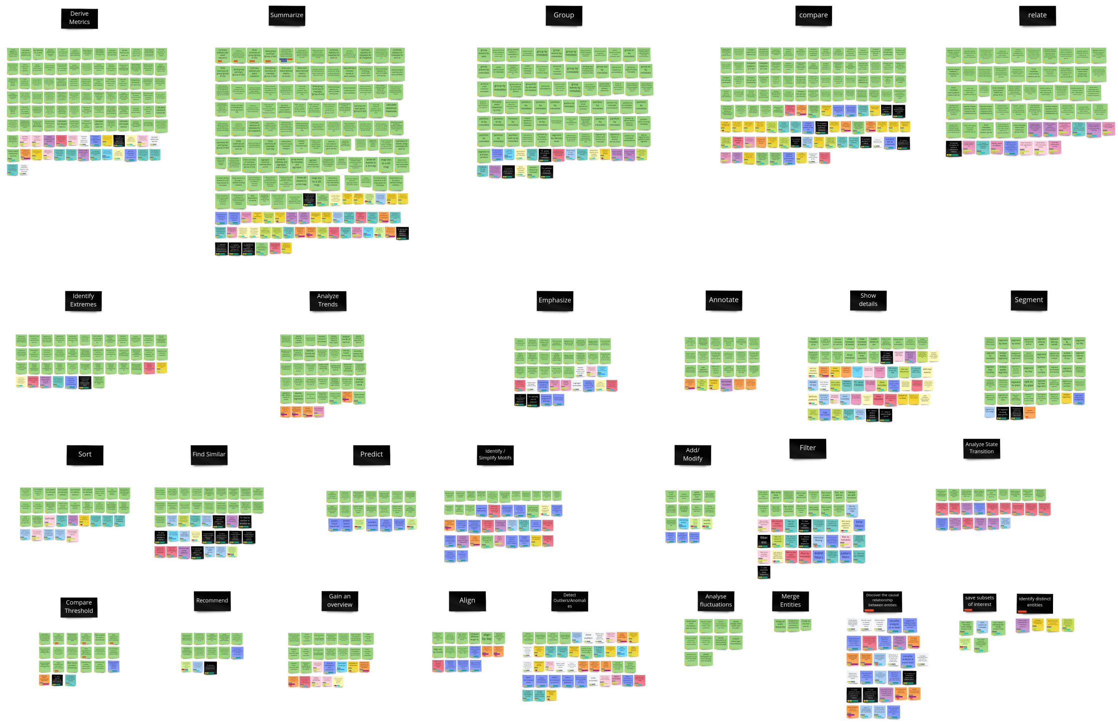


Figure 5.8: In the task synthesis phase, the two affinity diagrams from both studies were merged into the merged affinity diagram as shown. In order to support merging, we split or merged tasks of each affinity as necessary, and if neither of the task names for each affinity was comprehensive enough, we assigned meaningful names to each group. Overall, 23 tasks matched in both studies, while two and three unique tasks were identified in user-based survey studies and survey of designs studies, respectively.

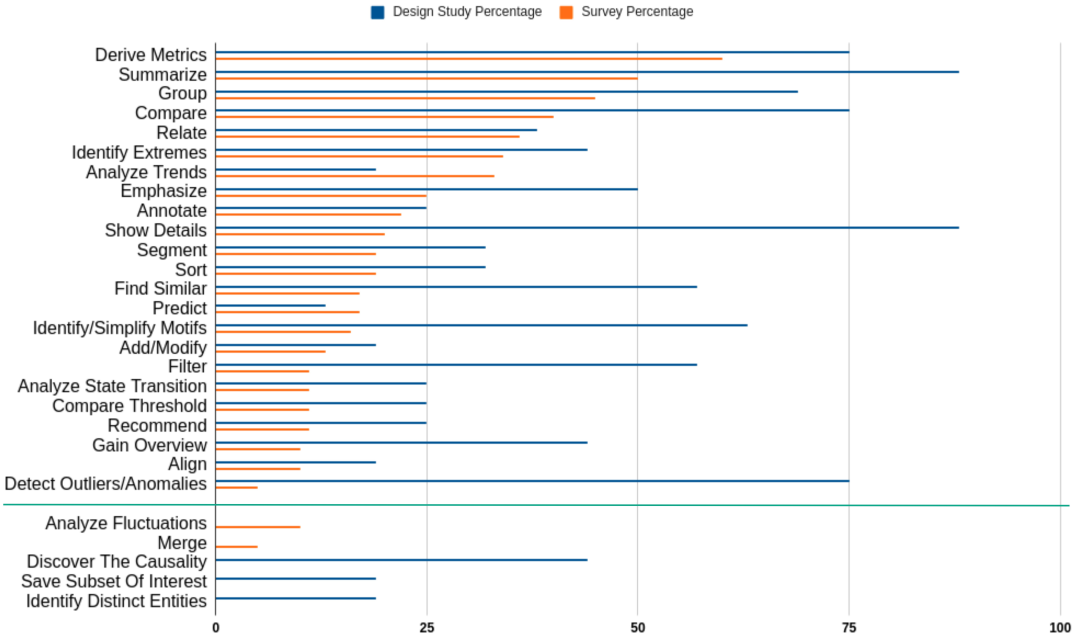


Figure 5.9: Comparison of tasks identified in the two studies based on the percentage of occurrence. Tasks identified from the user based survey study (orange) is underrepresented when compared to tasks identified from the survey of design studies (blue). One plausible reason maybe that design studies are a systematic way conducted by researchers with output that is thought through, compared to user surveys which are just anecdotal observations.

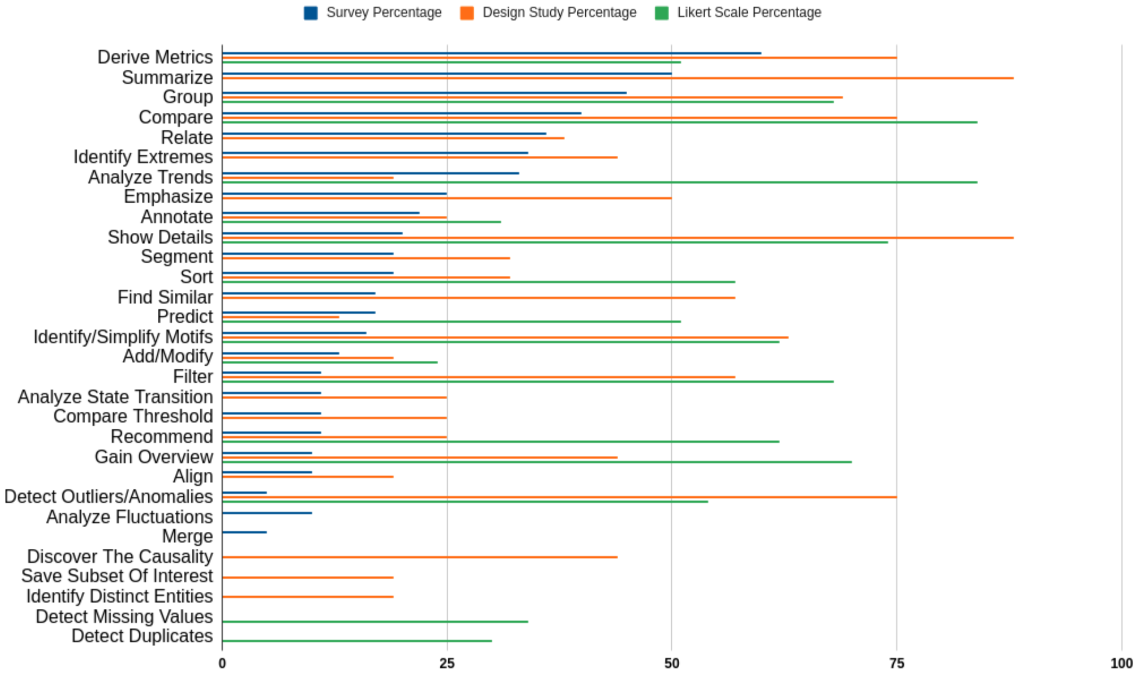


Figure 5.10: This figure shows comparison of the frequency of occuring different tasks in all 3 sources.

User Survey (Extracted Tasks)	Design Study Survey (Extracted Tasks)	User Survey (Anticipated Tasks)
Derive Metrics	Summarize	Compare
Summarize	Show Details	Analyze Trends
Group	Derive Metrics	Show Details
Compare	Compare	Gain Overview
Relate	Detect Outliers/Anomalies	Group
Identify Extremes	Group	Filter
Analyze Trends	Identify/Simplify Motifs	Identify/Simplify Motifs
Emphasize	Find Similar	Recommend
Annotate	Filter	Sort
Show Details	Emphasize	Detect Outliers/Anomalies
Sort	Identify Extremes	Derive Metrics
Segment	Gain Overview	Predict
Find Similar	Discover The Causality	Detect Missing Values
Predict	Relate	Annotate
Identify/Simplify Motifs	Sort	Detect Duplicates
Add/Modify	Segment	Add/Modify
Filter	Annotate	
Recommend	Recommend	
Analyze State Transition	Analyze State Transition	
Compare Threshold	Compare Threshold	
Gain Overview	Analyze Trends	
Align	Add/Modify	
Analyze Fluctuations	Align	
Detect Outliers/Anomalies	Save Subset Of Interest	
Merge	Identify Distinct Entities	
	Predict	

Figure 5.11: An overview of the most common tasks across all three sources is shown in this figure. Using arrows we tried to mark the tasks across all three cases. As long as the Likert scale did not list those tasks, we considered only the first two sources.

Chapter 6

Limitations, Discussion, and Future Work

6.1 Limitations of User Surveys

Our goal in this study is to build a problem characterization for TSES and present an overview of the current real-world problems on TSES. To extract the details on real-world problems, we decided to go for a user-based survey study since we can have a large problem space spanning towards a large number of domains. However, we identified some limitations of this approach such as, if someone needs some help in the middle of the questionnaire, it cannot be provided as with an interview. As a result, one misunderstanding can propagate through the questionnaire, affecting the quality of the responses. In order to mitigate the risk of poor quality responses, a second round of interviews could be conducted with people who provided answers that needed clarification

6.2 Limitations of Questionnaire Design

We followed the guidelines on questionnaire design recommended by existing research. However, based on the insights we identified, there were malformed questions and missing questions. We can remedy this by sending the questionnaire to a small sample covering a wide range of domains and adjusting it according to the feedback and responses before distributing it to the target population. Furthermore, there was no free questionnaire design tool available, which limited the design of the questionnaire.

6.3 Questionnaire Target Audience

We distributed the survey among people who have a TSES case with an actual dataset who we call domain experts and people who don't have a case with TSES with no dataset but have shown interest in TSES. We asked participants to either take a real-life scenario or imagine a case study they want to analyze and answer the questions describing the case. This further revealed with the question dataset availability. The majority of the participants don't hold actual real data while a considerable portion of participants holds real data. If our target population is all domain experts, then our results could be different a bit than this.

6.4 Influential Factor Identification and Analysis

We discovered a set of factors that would influence the problem characterization and built the questionnaire based on that. These factors set are not complete, and future research can be conducted to identify missing factors. Further, for some factors we discovered, the questionnaire is not a probable method to collect data. Examples include user characteristics such as cognitive load, working memory, context vocabulary etc. In addition, assessing these discovered but unanalyzed factors would be also a nice aspect of future research.

6.5 Missing Factors for Analysis

When analyzing the answers, we identified some missing questions. Generally, I was asking about user roles, but I identified that when coding, it would be nice if we asked them to rate from 1 to 5 their knowledge in two aspects: data analysis expert and domain expert besides the role.

6.6 Limitations of Data Analysis

In the analysis process, there can be lots of crosscuts between different factors which resulted in a lot of undiscovered relations. This would be a nice future work aspect. We tried to use existing libraries for Multiple Correspondence Analysis to detect the underlying structures of the data. Unfortunately, due to the large number of columns, generated diagrams were cluttered and were not usable to gain insights. In addition, we had only 65 responses, which is too little to discover relationships between around 60-70 columns with a high accuracy.

6.7 Methodological Contribution

Further, we identified the absence of methodological support to create a task characterization for hardly investigated data types. Most of the task characterizations are based on reviewing existing literature, which is not an option for TSES. Beyond the scope of this thesis, we will address this aspect by proposing a developed methodology for the BELIV workshop at the IEEE VIS conference, together with Clara-Maria Barth, who has also been involved in the coding process (cf. Section 4 and Chapter 5).

Chapter 7

Conclusion

In our research, we followed two main studies. Previous bodies of knowledge revealed four main aspects that influence problem characterization including domain context & users, data, tasks, and metrics. We systematically identified 25 factors under those four aspects by following the divide and conquer principle. Based on this, in our first study, we presented a systematic characterization of TSES-oriented real-world problems structured by those four aspects and 20 selected factors using a user-based survey study. A total of 65 responses shaped our proposed problem characterization for TSES. We further systematically analyzed these factors and identified the relationships between those factors. Future design studies can benefit from the knowledge gained from this study results, as both design and evaluation decisions can be informed by it. This also serves as a guide for domain experts to better understand their data and problems, as many experts often fail to assess their needs and data. Researchers who are interested in TSES can also benefit from this problem characterization.

In the second part of the research, we proposed a task characterization for TSES based on a user-based survey study and a survey literature study. A total of 65 survey responses and 16 design studies related to TSES formed the basis for this study, resulting in 28 clearly separable tasks. Building upon tasks, we compared the two sources used to build the tasks with a set of anticipated tasks using three main approaches, (1) set comparison, (2) frequency comparison, and (3) ranking comparison.

With the TSES problem characterization and task characterization, we aim to inspire future design studies on TSES and domain experts who are dealing with TSES.

Bibliography

[1] J.-w. Ahn, C. Plaisant, and B. Shneiderman. A task taxonomy for network evolution analysis. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 20(3):365–376, 2013.

[2] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Human-Computer Interaction Series. Springer, 2011.

[3] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization (InfoVis)*, pages 111–117. IEEE, 2005.

[4] R. A. Amar and J. T. Stasko. Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 11(4):432–442, 2005.

[5] N. V. Andrienko and G. L. Andrienko. *Exploratory analysis of spatial and temporal data - a systematic approach*. Springer, 2006.

[6] C. Arbesser, T. Mühlbacher, S. Komornyik, and H. Piringer. Visual analytics for domain experts: Challenges and lessons learned. In *Proceedings of the second international symposium on Virtual Reality & Visual Computing*, 2017.

[7] H. Arsham. Questionnaire design and surveys sampling.

[8] M. Baas, S. Bernonville, N. Bricon-Souf, S. Hassler, C. Kolski, and G. A. Boy. Towards a context model for human-centered design of contextual data entry systems in healthcare domain. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 223–233. Springer, 2014.

[9] J. Bernard, M. Steiger, S. Widmer, H. Lücke-Tieke, T. May, and J. Kohlhammer. Visual-interactive Exploration of Interesting Multivariate Relations in Mixed Research Data Sets. *Computer Graphics Forum (CGF)*, 33(3):291–300, 2014.

[10] M. Brehmer, S. Carpendale, B. Lee, and M. Tory. Pre-design empiricism for information visualization: Scenarios, methods, and challenges. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 147–151, 2014.

[11] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19(12):2376–2385, 2013.

[12] M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *Workshop on BEyond time and errors: novel evaluation methods for information visualization (BELIV)*, pages 1–8, 2014.

[13] N. Cao, D. Gotz, J. Sun, and H. Qu. Dicon: Interactive visual analysis of multidimensional clusters. *IEEE transactions on visualization and computer graphics*, 17(12):2581–2590, 2011.

[14] N. Cao, C. Shi, S. Lin, J. Lu, Y.-R. Lin, and C.-Y. Lin. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. *IEEE transactions on visualization and computer graphics*, 22(1):280–289, 2015.

[15] B. C. Cappers and J. J. van Wijk. Exploring multivariate event sequences using rules, aggregations, and selections. *IEEE transactions on visualization and computer graphics*, 24(1):532–541, 2017.

[16] Q. Chen, X. Yue, X. Plantaz, Y. Chen, C. Shi, T.-C. Pong, and H. Qu. Viseq: Visual analytics of learning sequence in massive open online courses. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 26(3):1622–1636, 2018.

[17] Y. Chen, P. Xu, and L. Ren. Sequence synopsis: Optimize visual summary of temporal event data. *IEEE transactions on visualization and computer graphics*, 24(1):45–55, 2017.

[18] M. P. Couper, R. Tourangeau, F. G. Conrad, and S. D. Crawford. What they see is what we get: Response options for web surveys. *Social science computer review*, 22(1):111–127, 2004.

[19] J. R. Cox and R. A. Griggs. The effects of experience on performance in wason's selection task. *Memory & Cognition*, 10(5):496–502, 1982.

[20] A. Crisan, J. L. Gardy, and T. Munzner. On regulatory and organizational constraints in visualization design and evaluation. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 1–9, 2016.

[21] W. Cui. Visual analytics: A comprehensive overview. *IEEE access*, 7:81555–81573, 2019.

[22] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. Eventaction: Visual analytics for temporal event sequence recommendation. In *IEEE Visual Analytics Science and Technology (VAST)*, pages 61–70. IEEE, 2016.

[23] F. Du, B. Shneiderman, C. Plaisant, S. Malik, and A. Perer. Coping with volume and variety in temporal event sequences: Strategies for sharpening analytic focus. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 23(6):1636–1649, 2017.

[24] E. Ergonomics. International encyclopedia of ergonomics and human factors. 2001.

[25] K. G. Esterberg. *Qualitative methods in social research*. Number 300.18 E8. 2002.

[26] S. Ganassali. The influence of the design of web survey questionnaires on the quality of responses. In *Survey research methods*, volume 2, pages 21–32, 2008.

[27] B. G. Glaser and A. L. Strauss. *The discovery of grounded theory: Strategies for qualitative research*. Routledge, 2017.

[28] M. C. Goulden, E. Gronda, Y. Yang, Z. Zhang, J. Tao, C. Wang, X. Duan, G. A. Ambrose, K. Abbott, and P. Miller. Ccvis: Visual analytics of student online learning behaviors using course clickstream data. *Electronic Imaging*, 2019(1):681–1, 2019.

[29] S. Guo, Z. Jin, Q. Chen, D. Gotz, H. Zha, and N. Cao. Visual anomaly detection in event sequence data. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 1125–1130. IEEE, 2019.

[30] S. Guo, Z. Jin, D. Gotz, F. Du, H. Zha, and N. Cao. Visual progression analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 25(1):417–426, 2018.

[31] S. Guo, K. Xu, R. Zhao, D. Gotz, H. Zha, and N. Cao. Eventthread: Visual summarization and stage analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 24(1):56–65, 2018.

[32] Y. Guo, S. Guo, Z. Jin, S. Kaul, D. Gotz, and N. Cao. A survey on visual analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[33] J. T. Hackos and J. Redish. User and task analysis for interface design. 1998.

[34] Y. Han, A. Rozga, N. Dimitrova, G. D. Abowd, and J. Stasko. Visual analysis of proximal temporal relationships of social and communicative behaviors. In *Computer Graphics Forum*, volume 34, pages 51–60. Wiley Online Library, 2015.

[35] A. Henderson. Interaction design: beyond human-computer interaction. *Ubiquity*, 2002(March):2, 2002.

[36] D. Herr, F. Beck, and T. Ertl. Visual analytics for decomposing temporal event series of production lines. In *International Conference on Information Visualisation (IV)*, pages 251–259. IEEE, 2018.

[37] S. L. Hibino. Task analysis for information visualization. In *International Conference on Advances in Visual Information Systems*, pages 139–146. Springer, 1999.

[38] K. Holtzblatt and H. Beyer. *Contextual design: defining customer-centered systems*. Elsevier, 1997.

[39] R. A. Irizarry. *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press, 2019.

[40] C. R. Jenkins and D. A. Dillman. *Towards a theory of self-administered questionnaire design*. Citeseer, 1995.

[41] Z. Jin, S. Cui, S. Guo, D. Gotz, J. Sun, and N. Cao. Carepre: An intelligent clinical decision assistance system. *ACM Transactions on Computing for Healthcare*, 1(1):1–20, 2020.

[42] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(12):2917–2926, 2012.

[43] Y.-a. Kang and J. Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *IEEE Visual Analytics Science and Technology (VAST)*, pages 21–30. IEEE, 2011.

[44] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual analytics: Scope and challenges. In *Visual data mining*, pages 76–90. Springer, 2008.

[45] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler. Challenges in visual data analysis. In *Tenth International Conference on Information Visualisation (IV'06)*, pages 9–16. IEEE, 2006.

[46] N. Kerracher and J. Kennedy. Constructing and evaluating visualisation task classifications: Process and considerations. In *Computer Graphics Forum (CGF)*, volume 36, pages 47–59. Wiley Online Library, 2017.

[47] N. Kerracher, J. Kennedy, and K. Chalmers. A task taxonomy for temporal graph visualisation. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 21(10):1160–1172, 2015.

[48] L. C. Koh, A. Slingsby, J. Dykes, and T. S. Kam. Developing and applying a user-centered model for the design and implementation of information visualization tools. In *2011 15th International Conference on Information Visualisation*, pages 90–95. IEEE, 2011.

[49] J. Krause, A. Perer, and H. Stavropoulos. Supporting iterative cohort construction with visual temporal queries. *IEEE transactions on visualization and computer graphics*, 22(1):91–100, 2015.

[50] B. C. Kwon, V. Anand, K. A. Severson, S. Ghosh, Z. Sun, B. I. Frohnert, M. Lundgren, and K. Ng. Dpvis: Visual analytics with hidden markov models for disease progression pathways. *IEEE transactions on visualization and computer graphics*, 27(9):3685–3700, 2020.

[51] B. C. Kwon, M.-J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, and J. Choo. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE transactions on visualization and computer graphics*, 25(1):299–309, 2018.

[52] H. Lam, M. Tory, and T. Munzner. Bridging from goals to tasks with design study analysis reports. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 24(1):435–445, 2017.

[53] P.-M. Law, Z. Liu, S. Malik, and R. C. Basole. Maqui: Interweaving queries and pattern mining for recursive event sequence exploration. *IEEE transactions on visualization and computer graphics*, 25(1):396–406, 2018.

[54] B. Lee, C. Plaisant, C. S. Parr, J.-D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Workshop on BEyond time and errors: novel evaluation methods for information visualization (BELIV)*, pages 1–5, 2006.

[55] P. Lietz. Research into questionnaire design: A summary of the literature. *International journal of market research*, 52(2):249–272, 2010.

[56] Z. Liu, R. J. Crouser, and A. Ottley. Survey on individual differences in visualization. In *Computer Graphics Forum*, volume 39, pages 693–712. Wiley Online Library, 2020.

[57] Z. Liu, Y. Wang, M. Dontcheva, M. Hoffman, S. Walker, and A. Wilson. Patterns and sequences: Interactive exploration of clickstreams to understand common visitor paths. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 23(1):321–330, 2017.

[58] D. Lloyd and J. Dykes. Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE transactions on visualization and computer graphics*, 17(12):2498–2507, 2011.

[59] M. H. Loorak, C. Perin, N. Kamal, M. Hill, and S. Carpendale. Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 22(1):409–418, 2015.

[60] J. Magallanes, T. Stone, P. D. Morris, S. Mason, S. Wood, and M.-C. Villa-Uriol. Sequen-c: A multilevel overview of temporal event sequences. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2021.

[61] S. Malik, F. Du, M. Monroe, E. Onukwugha, C. Plaisant, and B. Shneiderman. Cohort comparison of event sequences with balanced integration of visual analytics and statistics. In *Conference on Intelligent User Interfaces*, page 38–49, New York, NY, USA, 2015. ACM.

[62] N. McCurdy, J. Dykes, and M. Meyer. Action design research and visualization design. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 10–18, 2016.

[63] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200, 2014.

[64] M. D. Meyer, T. Munzner, and H. Pfister. Mizbee: A multiscale synteny browser. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 15(6):897–904, 2009.

[65] X. Mu, K. Xu, Q. Chen, F. Du, Y. Wang, and H. Qu. Moocad: Visual analysis of anomalous learning activities in massive open online courses. In *EuroVis (Short Papers)*, pages 91–95, 2019.

[66] T. Munzner. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6):921–928, 2009.

[67] T. Munzner. *Visualization Analysis and Design*. A.K. Peters visualization series. A K Peters, 2014.

[68] N. V. Nguyen, J. Hass, and T. Dang. Timeradar: Visualizing the dynamics of multivariate communities via timeline views. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 350–356. IEEE, 2021.

[69] P. H. Nguyen, R. Henkin, S. Chen, N. Andrienko, G. Andrienko, O. Thonnard, and C. Turkay. Vasabi: Hierarchical user profiles for interactive visual user behaviour analytics. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 26(1):77–86, 2019.

[70] C. B. Nielsen, S. D. Jackman, I. Birol, and S. J. Jones. Abyss-explorer: visualizing genome sequence assemblies. *IEEE transactions on visualization and computer graphics*, 15(6):881–888, 2009.

[71] M. Oppermann and T. Munzner. Data-first visualization design studies. In *2020 IEEE Workshop on Evaluation and Beyond-Methodological Approaches to Visualization (BELIV)*, pages 74–80. IEEE, 2020.

[72] E. M. Peck, B. F. Yuksel, L. Harrison, A. Ottley, and R. Chang. Position paper: towards a 3-dimensional model of individual cognitive differences. *Beyond Time & Errors–Novel Evaluation Methods for Visualization (Proceedings of BELIV 2012), ACM*, pages 1–6, 2012.

[73] A. Perer and F. Wang. Frequence: Interactive mining and visualization of temporal frequent event sequences. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, pages 153–162, 2014.

[74] C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, pages 109–116, 2004.

[75] C. Plaisant and B. Shneiderman. The diversity of data and tasks in event analytics. In *IEEE VIS Workshop on Temporal & Sequential Event Analysis*, 2016.

[76] P. J. Polack, S.-T. Chen, M. Kahng, M. Sharmin, and D. H. Chau. Timestitch: Interactive multi-focus cohort discovery and comparison. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 209–210. IEEE, 2015.

[77] P. J. Polack Jr, S.-T. Chen, M. Kahng, K. D. Barbaro, R. Basole, M. Sharmin, and D. H. Chau. Chronodes: Interactive multifocus exploration of event sequences. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(1):1–21, 2018.

[78] A. J. Pretorius and J. J. Van Wijk. What does the user want to see? what do the data want to be? *Information Visualization*, 8(3):153–166, 2009.

[79] C. D. Redline and D. A. Dillman. *The Influence of Alternative Visual Designs on Respondent'sperformance with Branching Instructions in Self-administered Questionnaires*. US Bureau of the Census, 2000.

[80] P. Rheingans. Are we there yet? exploring with dynamic visualization. *IEEE Computer Graphics and Applications*, 22(1):6–10, 2002.

[81] A. Rind, W. Aigner, M. Wagner, S. Miksch, and T. Lammarsch. Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization*, 15(4):288–300, 2016.

[82] Y. Rogers, H. Sharp, and J. Preece. Interaction design: Beyond human-computer interaction, jon wiley & sons. *Inc*, 2002.

[83] R. E. Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 19(12):2356–2365, 2013.

[84] B. Saket, A. Endert, and J. Stasko. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 133–142, 2016.

[85] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[86] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE transactions on visualization and computer graphics*, 18(12):2431–2440, 2012.

[87] P. Senin, J. Lin, X. Wang, T. Oates, S. Gandhi, A. P. Boedihardjo, C. Chen, and S. Frankenstein. Time series anomaly discovery with grammar-based compression. In *Edbt*, pages 481–492, 2015.

[88] T. A. Slocum, D. C. Cliburn, J. J. Feddema, and J. R. Miller. Evaluating the usability of a tool for visualizing the uncertainty of the future global water balance. *Cartography and Geographic Information Science*, 30(4):299–317, 2003.

[89] T. W. Smith. Little things matter: A sampler of how differences in questionnaire format can affect survey responses. In *Proceedings of the American Statistical Association, Survey Research Methods Section*, pages 1046–1051. American Statistical Association Alexandria, VA, 1995.

[90] Sreejith. Data visualization vs. data analytics - what's the difference?, Oct 2021.

[91] G.-D. Sun, Y.-C. Wu, R.-H. Liang, and S.-X. Liu. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 28(5):852–867, 2013.

[92] X. Suo. *A task-centered visualization design environment and a method for measuring the complexity of visualization designs*. Georgia State University, 2009.

[93] J. Tang, Y. Zhou, T. Tang, D. Weng, B. Xie, L. Yu, H. Zhang, and Y. Wu. A visualization approach for monitoring order processing in e-commerce warehouse. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 28(1):857–867, 2021.

[94] J. J. Thomas. *Illuminating the path: the research and development agenda for visual analytics*. IEEE, 2005.

[95] J. J. Thomas and K. A. Cook. A visual analytics agenda. *IEEE computer graphics and applications*, 26(1):10–13, 2006.

[96] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. 1986.

[97] F. Viégas, M. Wattenberg, J. Hebert, G. Borggaard, A. Cichowlas, J. Feinberg, J. Orwant, and C. Wren. Google+ ripples: A native visualization of information flow. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1389–1398, 2013.

[98] Q. Wang, T. Mazor, T. A. Harbig, E. Cerami, and N. Gehlenborg. Threadstates: State-based visual analysis of disease progression. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 28(1):238–247, 2021.

[99] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 457–466, 2008.

[100] S. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *IEEE Conference on Visualization*, pages 139–143. IEEE, 1990.

[101] H. Weisberg, H. F. Weisberg, J. A. Krosnick, and B. D. Bowen. *An introduction to survey research, polling, and data analysis*. Sage, 1996.

[102] K. Xu, Y. Wang, L. Yang, Y. Wang, B. Qiao, S. Qin, Y. Xu, H. Zhang, and H. Qu. Clouddet: Interactive visual analysis of anomalous performances in cloud computing systems. *IEEE transactions on visualization and computer graphics*, 26(1):1107–1117, 2019.

[103] P. Xu, H. Mei, L. Ren, and W. Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *IEEE transactions on visualization and computer graphics*, 23(1):291–300, 2016.

[104] P. Xu, H. Mei, L. Ren, and W. Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 23(1):291–300, 2017.

[105] J. S. Yi. Implications of individual differences on evaluating information visualization techniques. *International Journal of Human-Computer Studies*, 45(6):619–637, 2012.

[106] J. S. Yi, Y. ah Kang, J. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(6):1224–1231, 2007.

[107] Y. Zhang, K. Chanana, and C. Dunne. Idmvis: Temporal event sequence visualization for type 1 diabetes treatment decision support. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 25(1):512–522, 2018.

[108] Z. Zhang, D. Gotz, and A. Perer. Iterative cohort analysis and exploration. *Information Visualization*, 14(4):289–307, 2015.

[109] J. Zhao, N. Cao, Z. Wen, Y. Song, Y. Lin, and C. Collins. Fluxflow: Visual analysis of anomalous information spreading on social media. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 20(12):1773–1782, 2014.

[110] C. Ziemkiewicz, A. Ottley, R. J. Crouser, K. Chauncey, S. L. Su, and R. Chang. Understanding visualization by understanding individual users. *IEEE computer graphics and applications*, 32(6):88–94, 2012.