EXPLORATION SPACE
OF
HUMAN-DATA INTERACTION

by

ERDEM KAYA

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Sabancı University
December 2017

EXPLORATION SPACE OF HUMAN-DATA INTERACTION

APPROVED BY

Assoc.Prof. Selim Saffet Balcısoy ....
(Thesis Supervisor)

Prof. Burçin Bozkaya ....

Asst.Prof. Çağatay Türkay ....

Asst.Prof. Kamer Kaya ....

Prof. Uğur Güdükbay ....

DATE OF APPROVAL: 30/11/2017

ABSTRACT

EXPLORATION SPACE OF HUMAN-DATA INTERACTION

ERDEM KAYA

Ph.D. Dissertation, December 2017

Supervisor: Assoc.Prof. Selim Saffet Balcısoy

Keywords: Human-data interaction, data analytics, information visualization.

Data is everywhere. Starting with the invention of writing, representation artifacts brought the data to observable state which led to natural establishment of an interaction form between human and data. In the human-data interaction (HDI) environment, data representations and analytic systems act as an intermediary role. I suggest a new definition for HDI in which this interaction is conceptualized as a communication model over a set of media. The interaction occurs with the exchange of messages originated from both human and data. Timing and content of the messages are employed to facilitate objective evaluation of properties of analytic system in question.

To systematically investigate the complex nature of HDI, my methodology postulates the phenomenon as a high-dimensional space in which data analytic systems could be positioned based on their properties. Evaluation of the properties are performed based on solid definitions of the dimensions. I define five properties for data analytic systems, namely, responsiveness, communication media level, unit task diversity, closeness factor, and progressiveness level, and demonstrate how these properties could be objectively calculated.

I visually explore the HDI space in which data analytic systems reported in my thesis are plotted on a two-dimensional Cartesian system whose axes are responsiveness and communication media level. Visually identifiable patterns in this plot, which I call realms, are characterized by quantitative and qualitative analysis of objective, behavioral, and subjective data collected during the user interaction with the corresponding analytic system.

ÖZET

İNSAN-VERİ ETKİLEŞİMİNİN KEŞİF UZAYI

ERDEM KAYA

Doktora Tezi, Aralık 2017

Tez Danışmanı: Doç.Dr. Selim Saffet Balcısoy

Anahtar Kelimeler: İnsan-veri etkileşimi, veri analitiği, bilgi görselleştirmesi.

Veri her yerdedir. Yazının icadından itibaren, gösterim araçları veriyi insan ve veri arasında bir etkileşim şeklinin doğal olarak kurulmasına neden olan gözlemlenebilir bir duruma getirdi. İnsan-veri etkileşim (İVE) ortamında, veri gösterimleri ve analitik sistemler insan ve veri arasında aracı olarak rol oynarlar. İnsan ve veri arasında bir ortam üzerinden gerçekleşen bir haberleşme modeli olarak etkileşimin kavramsallaştırıldığı yeni bir İVE tanımı öneriyorum. Böyle bir etkileşim, insan ve veriden kaynaklanan mesajların karşılıklı gönderimi ile gerçekleştirilir. Bu mesajların zamanlaması ve içerikleri ise bahse konu analitik sistemin özelliklerinin nesnel bir şekilde hesaplanmasını mümkün kılar.

İVE'nin karmaşık yapısını sistematik bir şekilde incelemek için kullandığım metodoloji, İVE'yi içerisinde analitik sistemlerin kendi özelliklerine göre konumlandırılabileceği çok boyutlu bir uzay olarak ele alır. Analitik sistemlerin özellikleri ise bahse konu uzay boyutlarının kesin tanımlarına bağlı olarak yapılır. İsimleri hızlı cevap verebilirlik, haberleşme ortam seviyesi, birim görev çeşitliliği, yakınlık etkeni ve devamlılık seviyesi olan beş veri analitik sistem özelliği tanımlıyor ve bu özelliklerin nesnel olarak nasıl hesaplandığını gösteriyorum.

Eksenleri hızlı cevap verebilirlik ve haberleşme ortam seviyesi özelliklerinden oluşan iki boyutlu bir Kartezyen sistem olan ve üzerine tezimde detayları açıklanan veri analitik sistemlerinin yerleştirildiği İVE uzayını görsel olarak keşfediyorum. Bu yerleşimde görsel olarak tespit edilebilen ve alt alanlar olarak adlandırdığım paternler, bahse konu analitik sistemlerle ilgili kullanıcıların yapmış olduğu etkileşimlerden toplanan nesnel, davranışsal ve öznel verilerin nitel ve nicel analizleriyle karakterize edilir.

to my little dearest, Aren Adem.

ACKNOWLEDGMENTS

My doctoral research would not be possible without support of my family, advisor, friends, and colleagues. I am grateful for their support and assistance.

First and foremost, I should express my appreciation to my little dearest two-years-old son, Aren Adem, and my dear wife Didem, from whose time I spared much, and will never be able to tolerate. I am also thankful to my mother Fatma, mother- and father-in-law Melihat and İrfan, brother Emre, and brother-in-law Rahmi for filling in the social gap that I caused during my never ending studies. For those who plan to start a doctoral education while, at the same time, being a parent with a full-time position, I strongly suggest reconsidering it since such a goal necessitates great support which could only be provided by a team of incredibly strong family members.

None of the advisors can claim that they are as much helpful to their students as my advisor Assoc.Prof. Selim Balcısoy has been to me. He not only advised and directed my research, but also spent much effort for my development as a researcher, and supported me as a friend during the time I followed the tough path of rigorous Ph.D. student life. If I will be able to call myself a "data scientist" in the future, no doubt that I will owe it to the great support of Prof. Burçin Bozkaya. I would like to express my sincere appreciation to him for suggesting research on data analytic topics, and opening pavements to the league of data scientists.

Turkish Navy supports and encourages academic development of its personnel by tasking them at foreign or national Universities with completing an academic program, and, apparently, this policy has proven to be a successful effort bearing flourishing naval industry. I am grateful for the time and support that I was provided with by Turkish Navy.

I am also thankful to the Scientific and Technological Research Council of Turkey (TUBITAK) for supporting part of my research under both BIDEB 2214-A fellowship program and project number 114E516.

Part of my research would not be realized without industrious support of Asst.Prof. Çağatay Türkay, Dr. Yoshihiko Suhara, Dr. Xiaowen Dong, Yasin Fındık, and Mert Toka. I am very thankful for their efforts and times during and after the research conduct. I am also thankful to Prof. Alex "Sandy" Pentland for advising and hosting part of my research in his research group, namely Human Dynamics in MIT Media Lab.

Several other people spent much effort in order to bring my research to perfection. I am grateful to Asst.Prof. Sema Alaçam and Ceren Kayalar for participating in my studies, co-authoring, and participating in brainstorming sessions, to Gökhan Göktürk for all his support with his tremendous technical knowledge, to Alicem Batmansuyu and Doğacan Bilgili for their efforts in designing and realizing our physicality studies, and preparing professional video representations for my studies. I also thank to all the participants of my case studies and laboratory experiments. In particular, I would like to thank to the members of Akbank CRM Department for their contribution to our research with their valuable know-how.

And finally, I would like to stress the importance of within-lab collaboration for the success of academic research. I believe most of the creativity aspect of my research work stemmed from my discussions with the intelligible and skillful members of Sabancı University Behavioral Analytics and Visualization Lab (BAVLAB). I am unable to explain how confident a researcher can feel when any of his or her ideas could be realized and demoed in almost no time in the hands of the members of this great group.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The press agency of a university tasked two young journalists with preparing a report pertaining to the advancements of the recently established data analytics laboratory. They were quite interested in the laboratory's research activities as they were dealing with the data representations in order to support some of the news published by the agency. The researchers demonstrated various kinds of information visualization demos and several interesting results from their recent studies while the journalists were taking notes and video recording the demonstrations. It was not until they were shown a table on which a vibrant data visualization is projected via a projector hanged down from the ceiling. An animation of geographically plotted hourly credit card transaction data of a group of anonymized customers was being played. They were impressed with what had been shown, started to ask questions, and expressed their desire to interact with the table. It was being a different experience for them as they were able to monitor the spending behavior of the customers by just waving hands over the table.

After around an hour, the journalists were still at the table (Figure 1.1) reasoning about the visualizations, and deepening the discussion. Finally, the researchers had to kindly notify them that it was already lunch time, and they had to leave the laboratory.

The relatively high interest on the tabletop information visualization was neither by chance nor due to personal taste. A particular combination of several factors such as the purpose of the visualization, needs of the users, data visualization medium, and the interaction type led to an immense level of *engagement* with the data. For this particular case, if the researchers were to represent the same visual analytic system to, for example, the bank officials from the CRM[1] department looking for customer segmentation profiles, it would have been rather a disappointing experience as the task would not be a proper fit for their needs. Customer segmentation is a computationally expensive problem involving processing of various kinds of data from different sources and advanced data mining approaches. Due to this reason, such tasks requiring high throughput would be better supported with applications

[1] Customer Relationships Management

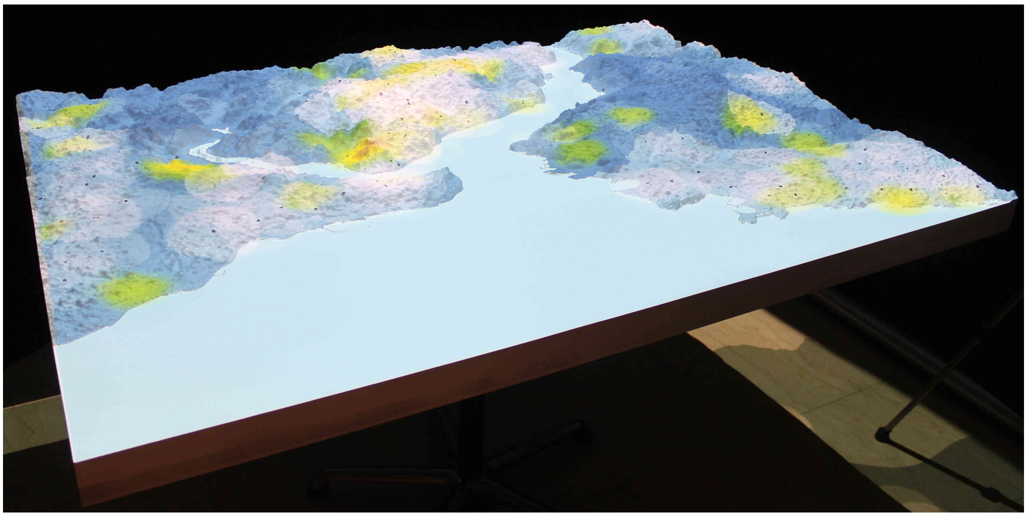

Figure 1.1: A tabletop hybrid visual analytic system for representing high dimensional spatio-temporal data in both physical and digital form. Presented in the picture is a frame from an animation of credit card transaction density change over time.

adopting efficient data management and representation capabilities, and hence, the interactivity aspect would be less important in that regard.

High level of engagement with the analytic tools is at the heart of the success of making sense of the data [1], [2]. The more we have interest in the data and the tool, higher the chances that we will be able to locate *the important* and *the valuable* laying under the data. With that being said, investigating the relationship between the engagement and sensemaking in data analytics terms could be considered as an important attempt. Development of knowledge on this relation can lead us to develop or build data analytic systems that can prove to be better fits for the needs of their users.

The terms *engagement* and *sensemaking* seem to have broad definitions according to the current body of literature [3], [4]. In this thesis, I redefine these terms in rather simpler form for the sake of narrowing down the scope of my thesis work. Engagement corresponds to the extent of temporal continuity of the user interaction with the data analytics system in question. For example, a user manipulating an online data visualization in almost real time regardless of whether for short or long periods of time would be considered as highly engaged whilst a scientist running a batch analytical task whose duration is measured in terms of hours or days would be considered as lowly-engaged.

On the other hand, I define sensemaking as the continuous update of analyst's mental model during the interaction between the *human* and the *data*, or its representation. The modality of the interaction medium can be considered as a continuous spectrum between the fully physical and fully virtual representation of the data. The more the physical components has the data analytic system, the more it is considered as physical, and similarly, systems involving more virtual components are evaluated to be highly digital.

By taking the new definitions of the engagement and sensemaking as two axes, we can build a two-dimensional *exploration space* into which we can position any data analytics system. In such a space, one can characterize the analytics systems with respect to their projections onto each of these axes. Such characterization could lead to a better understanding of why we can carry out productive analysis sessions with some tools or systems while we fail to do so with other systems. Furthermore, we can also have a more mechanical framework in order to decide the data analytic system type for the given tasks of the users.

Visual and data analytics systems closely positioned in the exploration space are expected to have a set of common properties. Such properties can involve various concepts pertaining to the analytic system. For example, the analytic task type, visualization kind, and form of data could be considered as characteristics differentiating a data analytic system from the other. In this thesis, I name specific subregions of exploration space involving closely positioned analytic systems in the exploration space as a *realm*s. The extent and the number of the realms residing in the exploration space is a matter of scope specification and observation data resolution. In order to more accurately specify the common properties of a given realm, more observation data from data analytic studies should be collected and analyzed.

I define three regions in the exploration space, and endeavor to characterize those regions with five analytic studies which my colleagues and I conducted during the last five years (2012-2017). Please see the plot of analytic systems reported in those studies in the exploration space in Figure 1.2. Depending on the clusters of those studies, I identified *hybrid systems*, *responsive digital systems*, and *unresponsive digital systems* realms (Figure 1.3). In the following chapters, I explain the characteristics of those realms with the associated studies that are shown in Figure 1.2. For the other regions of the exploration space for which I could not collect observation data, I try to make reflections on the possible characteristics of those realms in Chapter 6.

It could be proposed that the whole exploration space could be characterized with the observations available in the literature. Such a brute force effort would involve a systematic extraction of the observation data pertinent to engagement and sensemaking, and be followed by transcription, coding, and analysis processes. Nevertheless, success of that effort would be rather impossible than tedious, due to the fact that very few data analytics research effort seem to measure or report such observation data. Still, it is possible to draw reflections and suggestions based on the few studies reporting adequate amount of observation data.

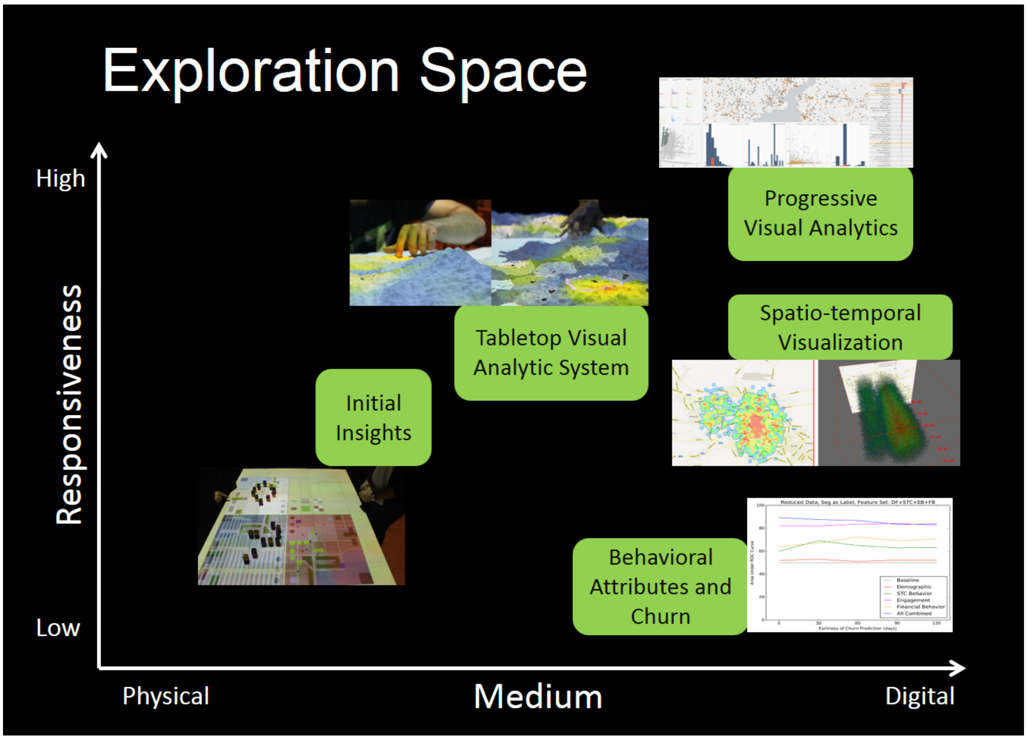

Figure 1.2: Exploration space for human-data interaction. The conceptual human-data interaction space has been depicted by employing the responsiveness and physicality levels as the axes.

With the term *observation data*, I mean all objective, behavioral, and subjective data that could be collected during the interaction sessions with a given analytic system. Unfortunately, collection and custodianship of such data seems not to be publicly established yet. Due to the lack of observation data, I would consider the reported studies of my colleagues' and mine as *probing* attempts into this big exploration space. By understanding the characteristics of the *probing locations*, I will try to make characterization of the realm surrounding the location.

1.1 Thesis Statement

Since the very early ages of civilization, as human beings we always had been in interaction with some form of data in order to manage our short- and long-term activities. However, the recent improvements in the technology, more specifically in the information visualization and visual analytic research, conveyed this interaction to a more observable state. I state that this opportunity can be employed to enhance the way we interact with and make sense of the data by exploring the conceptual *space* of human-data interaction (HDI). Such exploration along the *responsiveness* and *medium* dimensions can aid us in communicating the visual analytic models in terms of a finite set of criteria. Furthermore, having a structured view of HDI can facilitate a framework for making more accurate task-tool associations.

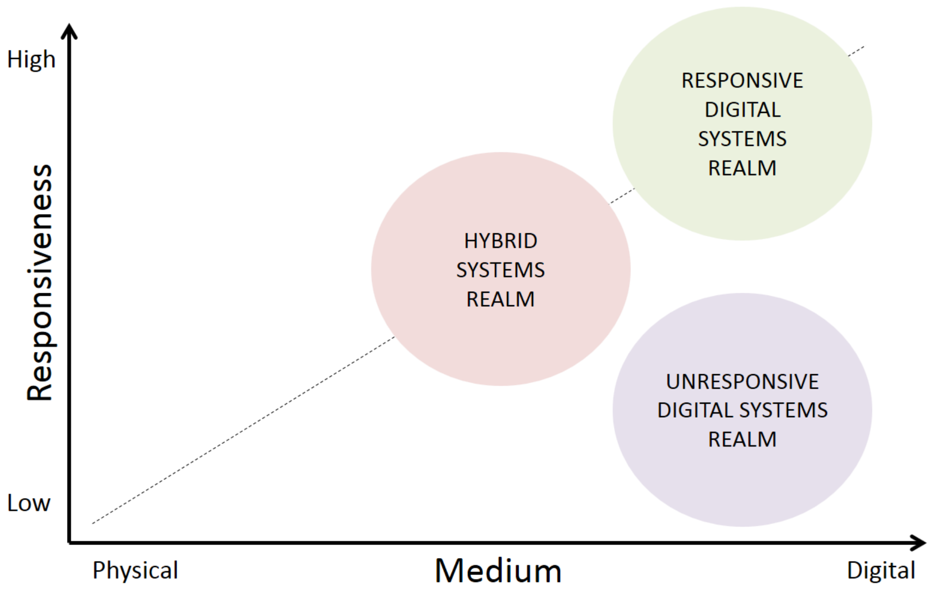

Figure 1.3: Realms of the exploration space. Exploration space is explained based on the subregions (i.e., realms) determined based on the characteristics of the analytic studies reported in the thesis. Studies located in the same realms are expected to have a particular set of shared properties.

HDI is a very broad concept such that reducing all of its aspects to only two dimensions would be far from a useful simplification. I state that, however, responsiveness and medium dimensions alone can adequately enable the establishment of the communication framework. I assert that such framework can at least guide the analytic system designers by suggesting the appropriate *criteria* for the tasks that they have to support in their solutions.

Exploration of the HDI space not only requires the investigation of infinitely many previous studies, but also necessitates the availability of the observation data for an objective evaluation. In this sense, the exploration can be considered as a reverse engineering approach where we would be collecting the data points to be able to have a bird-eye view of the exploration space. With the studies reported in this thesis, I will be *probing* into the exploration space at distinct points which could be representatives of the *realms* in their vicinity. From this perspective, my doctoral research work could be considered as the first step towards understanding the exploration space. A systematic exploration could be performed by collecting and analyzing organized and shared observation data from future data analytics work.

1.2 Thesis Overview

Chapter 2: Exploration Space of Human-data Interaction. In this chapter, I explain how the terms *human-data interaction*, *engagement*, and *sensemaking* are perceived in this thesis. Then, I introduce the *exploration space* and explain the possible ways to use it as a model followed by the characterization criteria that can be used to define various parts of the exploration space. The methodology and organization employed in the investigation of the previous studies are explained.

Chapter 3: Hybrid Systems Realm. Hybrid systems realm represents the set of analytic systems with average interactivity and tangibility. Average values for these two dimensions are considered to be in the middle of the respective axes. For this realm, I represent two studies my colleagues and I conducted on tangible interfaces and hybrid visual analytic systems.

Chapter 4: Responsive Digital Systems Realm. Responsive digital systems realm comprises the systems that are both highly interactive and highly digital. I characterize this realm by providing observation data from two infovis studies involving fully computer-based visual analytic systems.

Chapter 5: Unresponsive Digital Systems Realm. This realm encompasses a particular type of human-data interaction where the interaction takes place in the digital side, however, the interaction is relatively low. To better explain this realm, I investigate a data analytic study on bank customer churn prediction.

Chapter 6: Discussion & Conclusion. In this chapter, I make projections on what could be possible examples of other realms that could not be explained with studies reported in this thesis, and envision possible new analytic systems that could be representatives of other realms. Moreover, I explain how significant previous works can fit in the exploration space. Furthermore, I explain why the HDI exploration space is important, and how we can utilize this framework. Finally, I explain what other possible dimensions could be for the exploration space, and conclude the thesis by highlighting important remarks.

1.3 Clarification on Pronouns

I conducted all analytical studies reported in this thesis in collaboration with my colleagues. Due to this reason, the pronouns "we", "our", "ours", and "us" are used in the statements explaining the actions performed during the execution of those studies. Likewise, I say "I", "my", "mine", and "me" for explaining my own actions or ideas that are not related to the work done in collaboration.

CHAPTER 2

EXPLORATION SPACE OF HUMAN-DATA INTERACTION

Along with the improvement in the computing and data storage capabilities, the field of computer science in academia have been embracing various research interest groups. Particularly the multidisciplinary fields having intersections with social sciences such as human-computer interaction (HCI) have attracted much attention from wide range of communities. This is mainly due to the fact that computing and processing tools and devices have widely become ubiquitous, and problem at hand is more related to how we use the technology rather than how to build more powerful devices, and, improvements in the technology tend to follow the user experience and needs.

User experience is defined as "a person's emotions and attitudes about using a particular product, system or service" by the standard on ergonomics of human system interaction (ISO 9241-210). In addition to the usability aspects, human-centered design necessitates the consideration of needs and tasks of users as suggested by the product design field [5]. For that matter, interaction between human and the relevant system is a phenomenon that has been well studied for the last four decades [6], and understanding of user interaction can reveal invaluable hints on the success of a given system [7].

In the visual analytic and information visualization domains, *interaction* has been studied as a means to communicate the user commands to visual analytic system [2], [8], [9]. Visual analytic system, in turn, is expected to respond to user with desired data processing and visualization manipulation. Due to its human-centric nature, the validation of visualization system interaction requires employment of methodologies with human participants such as controlled experiments,insight-based methodologies, and case studies [10]. However, I suggest that user interaction with data analytic systems has other aspects deserving investigation. For example, the *continuity* of interaction could explain the extent of the engagement with the analytic system. The *medium* on which the interaction occurs (e.g., tangible or graphical) can serve as bottleneck or catalyst for the success of a proposed system. In this thesis work, I will be investigating these aspects of user interaction with data rather than the technical side of the interaction with a particular analytic system. Therefore, it

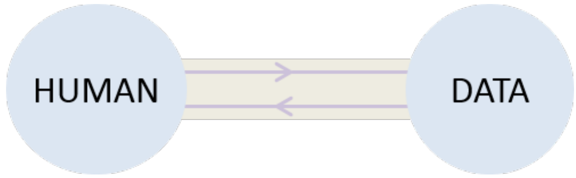

Figure 2.1: An alternative human-data interaction conceptualization.

seems to be a more appropriate approach to consider the investigations reported in this work in the context of a broader field, namely human-data interaction (HDI).

As will be explained in the following sections, HDI is a broad field involving various subjective projections. For example, Haddadi et al. considers the problem as a social issue on the interaction with the personal data [11].

I will be approaching HDI as a communication paradigm between user and data that might occur in *any* circumstance (Figure 2.1). For example, a stack of wooden cubes can communicate summary information of a given distribution data, or a scatter plot on computer display can demonstrate broader view of the distribution, however, omitting the tangible side of the interaction. Human-to-data direction of the communication involves *intelligent* commands sent by user to data over a *medium*, which stands as the base for the *representations* of the data.

HDI, in this perspective, embraces numerous characterization criteria such as engagement, sensemaking, representation modality, and data consumption schemes (stream or static). Please refer to Section 2.2.1 for all the criteria that I propose. Conceptually, HDI can be conceived as a multidimensional space whose dimensions are comprised of these criteria. In such a space, any analytic system could be positioned based on their *levels* (values) on each of these axes (criteria). Exploration of this space would be, if not impossible, a quite challenging problem due to high-dimensionality. In order to narrow down the scope of the problem, I investigate the exploration space from the responsiveness and the medium perspectives, and employ other criteria as the evaluation features of the investigated studies. In the following sections, I explain how I conceptualized the exploration space, and how a given analytic study or system in this space could be positioned. Examples of the positioning of data analytic research have been presented in Chapters 3, 4, and 5. The methodology used for mapping a given research work to the exploration space is an important and tedious task, and will be explained in Section 2.6.2.

2.1 Human-data Interaction

HDI is a nascent field involving a wide range of problems varying from the technical to social issues [11]. Haddadi et al. pose the HDI concept as the engagement of a subject with a flow of data and its consumers attributed as stakeholders. From

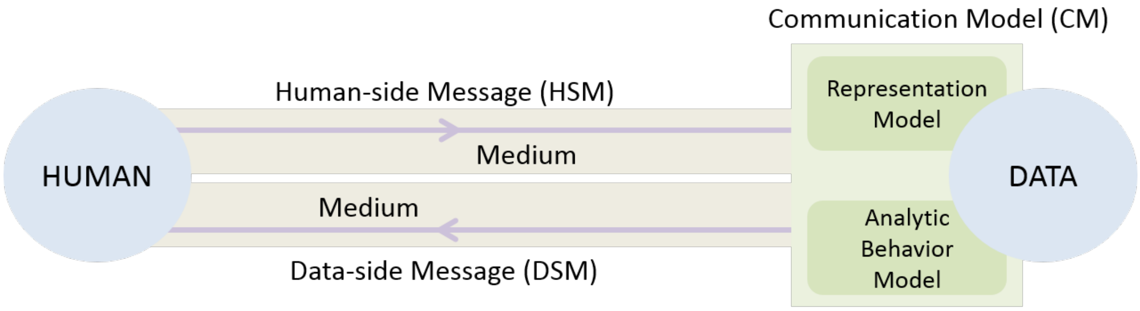

Figure 2.2: HDI as a communication concept between human and data. HDI is defined as the naturally-established communication between the human agent(s) and the data.

this vantage point, the problem is basically accepted to be in the vicinity of personal data formed around three core principles, namely, legibility, agency, and negotiability [12]. *Legibility* principle is related to the transparency of both data and the processing mechanisms to the users of the data. Being able to understand what the data is and how and what kind of benefit could be acquired by processing it is accepted as a prerequisite for the *agency* principle. With the agency principle, capabilities of users are emphasized as preferences on the data collection and processing, engagement in the data collection and storage, usage, making inferences, and modification on the data. And finally, *negotiability* is related to the freedom of users to re-evaluate their decisions in case of context change. In this regard, alongside the *current* data, future data should also be considered in order to preserve properties such as privacy. For example, users that are de-identified in a past dataset could become identifiable with additional available data [13].

In this thesis, I define HDI as the naturally-established two-way communication between human(s) and data. This communication has two parties: Human side and data side (Figure 2.2). On the human side, there has to be at least one *human agent*. Surrogates or similar other non-human intelligent entities are beyond the scope of this definition. At the other end of the communication is the *data* in which the human agent is interested. Data is rendered as a conceptual entity in this definition, and cannot be of use without its *communication model*. Data communication model comprises representation and analytic behavior models. Form of the *messages* that will be originated from data side are determined based on the *representation model*, whilst the *analytic behavior model* arranges the message traffic (when or whether to send messages). Various forms of human-data interaction could be established with any combination of one or more human and data agents. In Figure 2.3, a few examples of different human-data agent combinations are represented. Please note that not all possible combinations are shown in the figure.

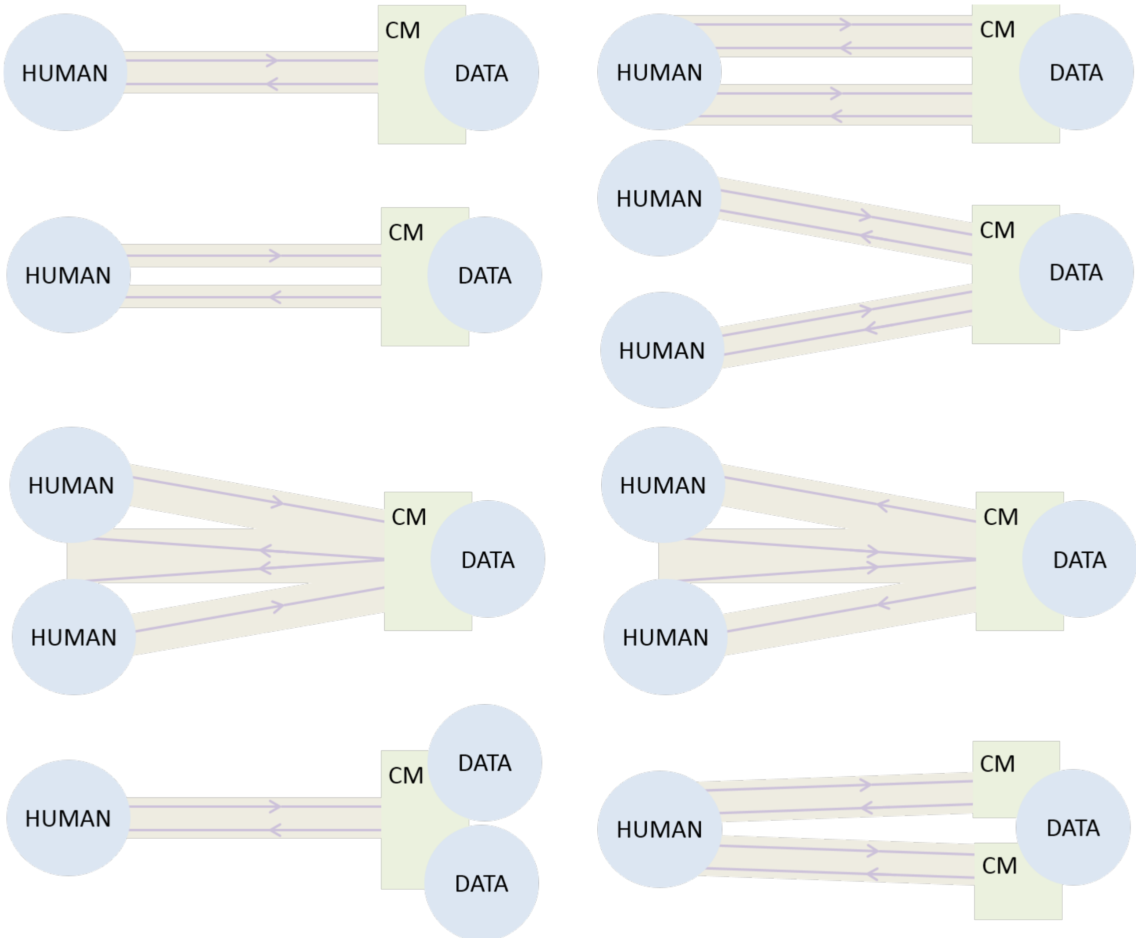

Figure 2.3: Possible forms of human-data interaction. With different combinations of one or more human and data agents, various forms of HDI could be established. Note that only a small subset of all possible combinations is presented in the figure.

Medium is the base (or platform) on which data *representations* are formed and carried. For example, for a scatter plot on a computer display, medium is a digital platform, and the scatter plot is the representation of the data.

Similarly, in a physical histogram made of wooden bricks, medium and representation would be physical (wood) and histogram, respectively. Representations could be human-readable (e.g., numbers on digital platform or written on paper), or other observable forms such as stream charts, box plots, or radar charts (Figure 2.4). Not all representations have to readily be associated with a particular meaning.

Messages carried in this communication scheme are different in each of the direction. The messages originated at the data side are the representations generated and managed by the communication models of the data. Human-side messages carry the commands that initiate data manipulation, representation generation, or analytic model processing. In this communication scheme, sensemaking of the data is a process occurring on the human side, and it is fed and directly effected by the messages originated from the data side. Due to this reason, content of the human-

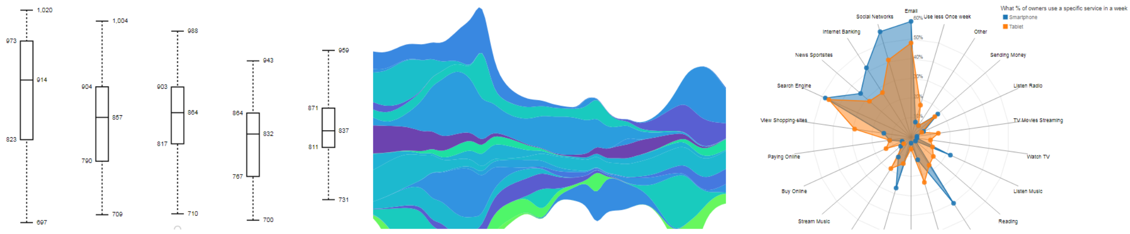

Figure 2.4: Data representation examples using digital medium.

side messages (representations), medium, and timing of the messages are the main arguments controlling the sensemaking. Sensemaking in the context of my thesis is explained in Section 2.2.3.

Various forms of HDI with respect to organizations of the communication parties are depicted in Figure 2.3. Possible cases of communications based on enumeration of human agent and medium count combinations are listed in Table 2.3. Number of media in a given analytic system should be considered as distinct channels for communication between user and data, therefore, for example, if a user interacts with the system via a graphical user interface and an additional tangible interaction mechanism, she is considered to be interacting over two distinct media.

Similarly, factors on the data side can also cause human-data interaction to diverge. For example, data might have one or multiple communication models, and each communication model might generate different representations of data with various responsiveness levels. In some cases, responsiveness cannot be enumerated with discrete values. For example, a set of credit card transaction data could be summarized with statistical analysis tools and instantly presented in number forms, and, at the same time, data could be batch processed with an analytic task (e.g., prediction, segmentation) and the respond could be retrieved after several hours. In this example, we can state that the data has two different communication models and the messages generated by them have different characteristics in terms of representation and responsiveness.

2.2 Exploration Space

Investigation of all proposed or realized analytic systems in hopes of understanding why they fail or success would be a fruitful, interesting, yet quite challenging task. An attempt to alleviate the problem would require a systematic approach which could map each data analytic system to a vector of properties, and then search for patterns or clusters of analytic systems with respect to property vectors. Such pattern search would necessitate conceptualization of a multidimensional exploration space whose dimensions would be the properties based on which the data analytic systems are characterized. For instance, responsiveness, data representation medium, and support for multi-user interaction could be accepted as properties of a statistical analysis software, and these three characterizations could be dimensions for the exploration space.

Nevertheless, enumeration of all characterization could be a subjective and never ending task bearing infinitely many properties. Moreover, exploration of patterns in such a high-dimensional space would be, if not impossible, prohibitively challenging. As a remedy to this problem, I propose that reducing the number of dimensions could facilitate an environment in which we could characterize the analytic systems to make problem more tractable. For example, as in this thesis work, response patterns and data representation medium could be used as the dimensions, and a two-dimensional space would enable visual recognition of possible patterns.

2.2.1 Dimensions

In my thesis work, I selected responsiveness and data representation medium as the dimensions for the exploration space. The selection has been made based on the currently available observation data acquired during my research work. As shown in Figure 1.2, observation data reported in this thesis leads to distribution of research in disparate location in the exploration space. Highly distributed view of the research work in the exploration space assists us in investigation of subregions which I name *realms*.

Responsiveness is defined as the temporal pattern of the time elapsed between the messages of the human agent and data. For example, a web-based information visualization application with with almost real-time responses to user interactions is assumed to be highly responsive. On the other hand, a physicalization system building a wooden line chart in an hour as a response to the user's initiation is considered as much less responsive.

The data medium dimension indicates the physicality level of the medium on which the messages carried between user and data. As the amount of physical components in the data medium increases, the medium is considered to be more physical.

2.2.2 Engagement

With respect to my definition, engagement has a direct relationship with the overall responsiveness of the system in question *and* the task at hand. Engagement is a naturally established case where the analytic task necessitates responsive mechanisms. For example, a visual exploratory analysis tool would be attributed as useful when it supports primitive tasks [14] with a high response rate. In particular,

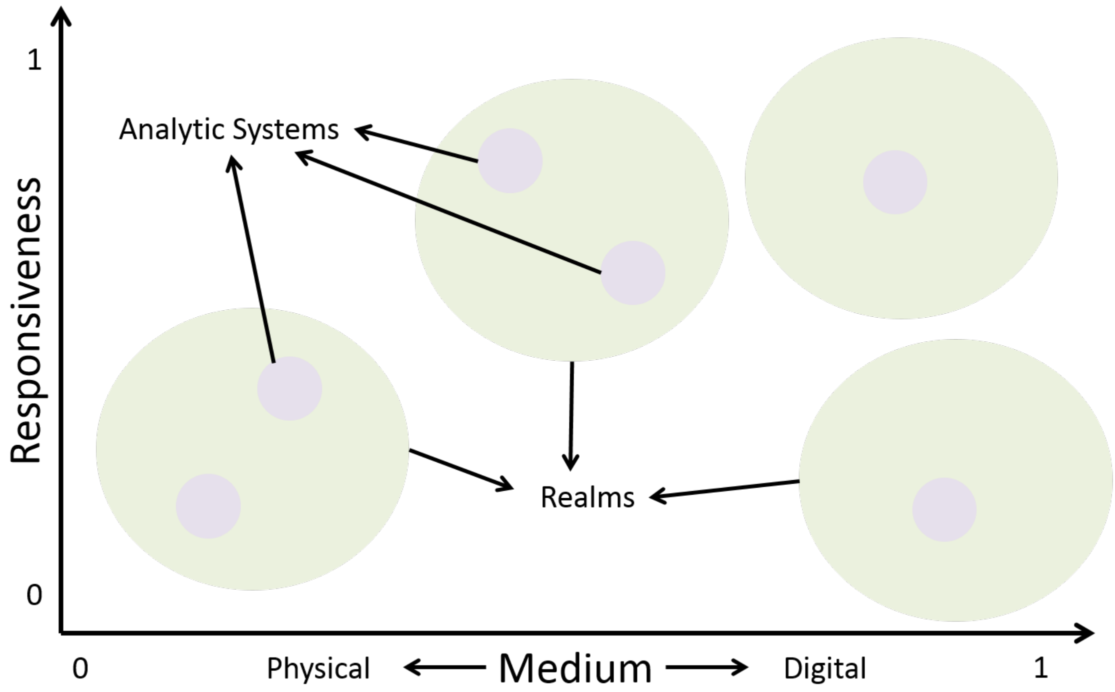

Figure 2.5: Analytic systems and realm concept. Each analytic system and its associated research effort (observation data) can be considered as *probes* into the exploration space. As the number of such probes increases, groupings of closely positioned analytic systems could form *realms* in the exploration space.

for a healthy interaction between the user and an analytic system, responses should be generated in less than a second [15].

High engagement with analytic systems may not be desired under all circumstances. In some cases, high throughput might be the priority during the interaction with data. For example, a clustering operation might be of necessity on large amounts of data. Such batch data operations (i.e., human-side message) are not expected to produce swift representations (i.e., data-side message).

2.2.3 Sensemaking

Sensemaking is defined as a concept for explaining the way the user of an analytic system understand the data. According my definition of HDI as a communication

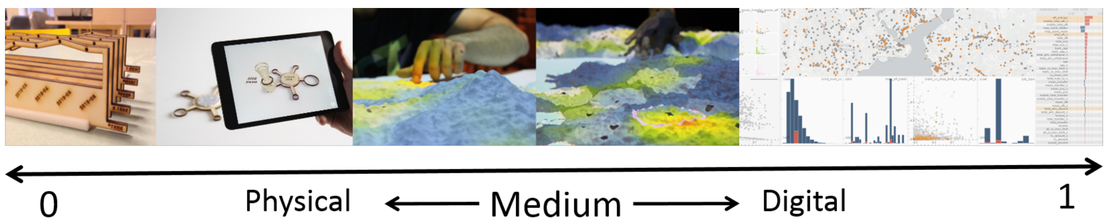

Figure 2.6: Physicality of the human-data interaction medium.

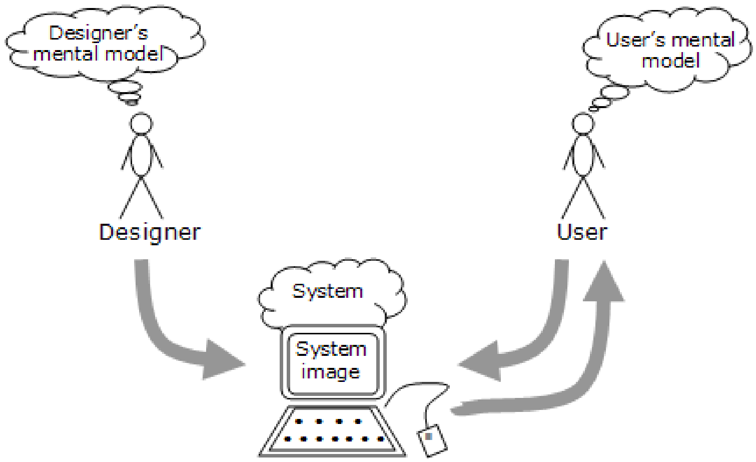

Figure 2.7: User's mental model. The illustration is the courtesy of Interaction Design Foundation [17].

model, making sense of the data is simply creation and progressively update of the user's mental model [16] based on the messages generated on the data side.

Mental models are the conceptualization and subjective understanding of a phenomenon pertaining to physical objects or virtual concepts. In other words, user's mental model is the subjective understanding of a person about how a system operates or a phenomenon occurs (Figure 2.7).

Mental models are continuously updated based on the information acquired through human sensing systems. In that regard, such updates naturally occur during the interaction with the data via the media on which the data side messages are conveyed. For example, a physical data representation sends representation messages both visually and tangibly.

The variety in the ways we communicate with the data might lead to different kinds of sensemaking. For example, let us assume two VA systems (VA1 and VA2) each having two media. Let VA1 dictates human-side messages be conveyed over a computer program (GUI), and data-side messages over physical medium (a web-based application connected to a 3D printer building physical representations of a given dataset). Contrarily, let VA2 accept human side messages over a tangible medium and data-side messages over a tabletop screen (a tabletop VA system accepting inputs with physical objects). Even if these two VA systems were to communicate the same dataset, mental models of a given user would develop differently due to different representations of the data. Due to this reason, it is of importance to accurately decide on the representation medium of a VA system in order to form the appropriate mental models.

2.3 Realms

Exploration space for HDI should be considered as a conceptual guide without any limits on any of the two axes. Furthermore, it should be noted that analytical studies should be positioned in the space with respect to objective methodologies explained in Section 2.6. However, positioning should be treated as an indicator showing which subregion of the space the analytic study belongs to. In that regard, precision of the position of an analytic system in the space is not important.

From technical point of view, analytic systems with similar properties are expected to be positioned in similar locations in the exploration space. Such proximity could simplify evaluation of a group of analytic systems according to their region, which I call *realms*.

Number of realms in an exploration space depends on the resolution of the overview of positioned studies (number of studies/analytic systems considered in the determination of regions). The number of analytic systems considered in this thesis led to a division of four realms, namely both-low, middle, both-high, and high-digital low-interaction realms.

Analytic system observations falling into these realms give us general insights about the realm in which the observation is made rather than making overall judgments about their characteristics. This is mainly due to the lack of enough observation per realm in this thesis. In this regard, observations reported in this thesis should be considered as probes [18] inserted into the realm in order to acquire general idea about them.

Determination of realms should be realized by applying conventional clustering algorithms on the interactivity and medium composition values which are calculated based on the methodology explained in Section 2.5.2. Nevertheless, in my thesis, the realms are identified based on the visual appearance of observations as shown in Figure 1.2.

2.4 Exploration Space as a Model

Articulation of exploration space model introduced in this thesis would merely be a wasteful attempt unless it could be used for good purposes. Permanent effort for the custodianship of an observation database as the data source for exploration space can bear unprecedented benefit to the information visualization and data analytics communities. In fact, such a reference system would serve as a mechanism to refer to analytic models by their design purpose and expected task set. For example, a data analytics tool that would be used for batch analytics tasks such as clustering would be mentioned as an analytic model in the high-digital low-interaction realm.

Such references would promote the use of a small set of concise terminology to talk about large set of analytic tools with common characteristics.

At the beginning of a new visual analytic system design, it is commonplace to investigate existing tools with similar tasks and responsiveness expectations. Given that expected tasks and operation characteristics of a proposed tool are well-defined at the beginning of design process, the purpose of the tool and and its interaction characteristics could be verified by referring to the exploration space.

2.5 Mapping Analytic Systems to Exploration Space

Mapping a given analytic system to exploration space is the process of plotting the system in the multidimensional space of HDI based on the analytic system properties. Plotting operation could be done based on some or all of the properties whose values are calculated based on the observation data of analytic studies reported in Chapters 3, 4, and 5. In this section, formal definitions of mapping and realm detection operations are presented followed by the introduction of analytic system properties and how they are calculated.

2.5.1 Problem Formulation

Let an analytic system S_i be a tuple $(r, c, p_0, \ldots, p_n)$ and set $S = \bigcup_i S_i$ be the set of all analytic systems that exists, where $i \in N$, r is the responsiveness level of S_i and $0 \leq r \leq 1$, c is the communication medium indicator and $0 \leq c \leq 1$, and $p_0, \ldots p_m$ are properties of S_i, and $m \in N$. A property p_j could be a numerical or categorical value, and can take a special value *not applicable*, abbreviated as NA, meaning that property p_j is not applicable to system S_i. m is the total number of all possible properties that are applicable to all analytic systems in set $\bigcup_i S_i$. m determines the depth of the investigation of each study.

Exploration space is a set of all possible tuples (a, b) such that $(a, b) \in [0, 1][0, 1]$. As explained in Section 2.2, exploration space could be established with a particular subset of other dimensions $\{r, c, p_0, \ldots, p_n\}$.

A mapping of an analytic system to exploration space is simply plotting system S_i to the first quadrant of a cartesian system with respect to r and c properties. Deciding the realms of the exploration space is a clustering operation on the dataset $\{(r_0, c_0), (r_1, c_1), \ldots, (r_N, c_N)\}$, where N is the total number of all analytical systems. Let $R_1, \ldots, R_K$ be the clusters of analytic systems that is found as a result of the realm determination process. $\forall R_i \subset S$. K is the total number of realms, and union of all clusters is equal to S (i.e., $\bigcup_k R_k = S$).

Let $(p_{11}, p_{12}, \ldots, p_{1m}), (p_{21}, p_{22}, \ldots, p_{2m}), \ldots, (p_{j1}, p_{j2}, \ldots, p_{jm})$ be the properties of analytic systems in cluster R_k, and $|R_k| = j$. Then the distributions $d_1, d_2, \ldots, d_m$ represent the characteristics of realm R_k where each d_i is the frequency distribution of the set $\{p_{1i}, p_{2i}, \ldots, p_{ji}\}$.

2.5.2 Properties of Analytic Systems

I define the properties of analytic systems as the set of all possible characteristics of all existing analytic systems. Based on this definition, responsiveness (r) and communication medium indicator (c) values, which I used as the dimensions of the exploration space, are also included in the property set. One might oppose to the idea of using a uniform set of properties for all existing analytic systems regardless of their differences. For example, progressive rendering of an information visualization system fed with stream data might be accepted as irrelevant for an analytic system with physical representations of data. That would be an appropriate opposition as the uncanny differences between analytic systems make it impossible to be represented with the same set of properties. However, it *could* be possible to represent all possible analytic systems under the same umbrella if we union their properties into a single set, and assign a special value such as not applicable (NA) to those properties that are not applicable to a given analytic system. For example, let systems S_1 and S_2 be analytic systems, and are characterized with property sets $\{p_1, p_2, p_3, p_4\}$ and $\{p_3, p_4, p_5, p_6\}$, respectively. It is possible to represent these two systems with a uniform set of properties ($\{p_1, p_2, p_3, p_4, p_5, p_6\}$), and S_1 could be characterized with a value vector of those properties in which p_5 and p_6 would equal to NA.

From theoretical point of view, there could be infinitely many properties due to the subjective nature of property identification process. New conceptual and technological advancements could bear new aspects of analytic systems, and hence, new properties. I suggest maintaining a large yet finite set of relevant properties in order to keep the problem tractable. In this thesis, I identify the following properties which, I suggest, will be an adequate set in order to prove the concept I propose in this thesis. Many more properties could be suggested for the analytic systems that I report in this thesis, however, I intentionally specify a small subset of properties in order to better explain the phenomenon with clear and simple examples.

2.5.2.1 Responsiveness

Responsiveness is defined as the average response time of a given analytic system messages generated on the human side. Average response time should be calculated by taking into consideration the responses on all available media. Then the average response time is normalized such that the resulting responsiveness constant (r) remains in the range $[0,1]$. Given the average response time as r_0, responsiveness property is calculated as follows:

$$r = \frac{1}{2} - \frac{\arctan(log(r_0))}{\pi}$$

One might notice that the formula produces values in range $[0,1]$. From theoretical point of view, 0 corresponds to immediate response by the system (i.e., response in no time) whilst 1 corresponds to no response. The median value $\frac{1}{2}$ is calculated for average response time 1 second, which is the suggested delta time for the second human time constant introduced by Card et al. [15]. Please see Section 4.2.1.1 for information on human time constants.

2.5.2.2 Communication Media Level

Let S be an analytic system with m media supporting the communication between human and data. Let H and D be the sets of the media for human- and data-generated messages, respectively. Let $|H| = h$ and $|D| = d$, then $h + d = m$.

Let $f_1, \ldots, f_h$ be the usage frequencies of media used by human side messages. Similarly, $g_1, \ldots, g_d$ be the normalized usage frequencies of media used by data-side messages. Then the communication media level c is calculated as follows,

$$c = \frac{\sum_i f_i a_i + \sum_j g_j b_j}{\sum_i f_i + \sum_j g_j}$$

where a_i and b_j are the i^{th} and j^{th} medium's physicality constant. In other words, the c constant is basically the weighted average physicality of all the media employed by system S. In order to map the constant c to the $[0,1]$ domain, I apply 0 and 1 values to a and b variables, respectively, implying that fully physical systems are represented with 0 whereas fully digital systems are represented with 1.

2.5.2.3 Unit Task Diversity

Amar et al. introduces 10 distinct tasks that could be performed on a given visual analytic system [19]. Amar and his colleagues conducted a study with 200 students in which they ask each student to select two datasets, and write as many questions and hypothesis as they can in order to analyze the data. They applied affinity diagram methodology on the student answer corpus, and developed a task taxonomy comprising ten task types, namely, *Retrieve Value*, *Filter*, *Compute Derived Value*, *Find Extremum*, *Sort*, *Determine Range*, *Characterize Distribution*, *Find Anomalies*, *Cluster*, and *Correlate*.

I employ Amar et al.'s task typology in order to characterize the user actions (human-side message) with analytic systems having taken place during my studies. In doing so, I transcribed the user inputs in terms of Amar's task types and prepare unit task execution frequencies. In that regard, I define unit task diversity as the normalized entropy of the unit task type executed by the user. In other words, the frequency of each human-side message type prorated to the rest of the messages are employed to calculate task usage entropy.

Let c_i be task type i execution count per unit time, then the unit task diversity is calculated as follows:

$$d = \sum_i \frac{-p_i log(p_i)}{log(T)}$$

where $p_i = \frac{c_i}{\sum_i c_i}$ and T is the total number of unique task types.

2.5.2.4 Human-side Message Closeness Factor

Let $t_0^k, \ldots, t_m^k$ be the human-side message execution timestamps for message type k, and δt_i be the time delta between t_i and t_{i-1}, and $\delta t_i = t_i - t_{i-1}$ in seconds. Then the closeness factor for message type k is calculated as follows:

$$T^k = \frac{log \sum_i 2^{\delta t_i^k}}{log c^k}$$

where c^k is the total number of k type messages.

2.5.2.5 Progressiveness Level

Exploratory data analysis necessitates high engagement of users with the data, and high latencies are not acceptable for the success of such analysis [20], [21]. High latencies typically occur due to high computation costs or large data sizes. Recently, a novel programming paradigm called *progressive analytics* has been introduced with the aim of splitting long lasting analytic tasks into chunks, and provide partial yet useful results that approximates to the final result [22], [23].

Progressiveness level p of an analytic system is an aggregation of the data consumption pattern of the data-side message types on all of the available media. For example, let $\{\lambda_1, \lambda_2, \lambda_3\}$ and $\{m_1, m_2\}$ be the data-side message types set and available media set, respectively. And let us assume that the λ_1 and λ_2 messages could be sent over medium m_1, whereas λ_3 could be sent over medium m_2. Then the p value of this analytic system is calculated based on an aggregation of progressive values evaluated for the combinations (λ_1, m_1), (λ_2, m_1), and (λ_3, m_2).

Let Λ be the set of all data-side message types and M be the set of all available media of an analytic system S. For all $\lambda_i \in \Lambda$ and $m_j \in M$, progressiveness level

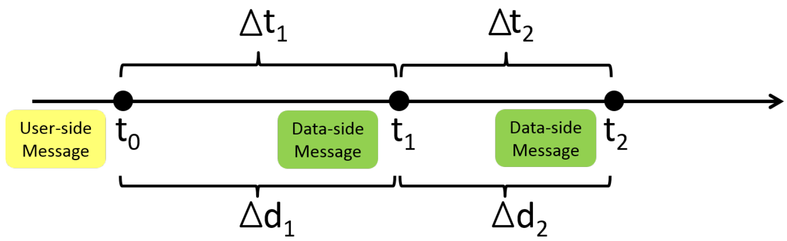

Figure 2.8: Illustration of an example data-side message sending pattern.

p_{ij} for each message type and medium combination (i.e., (λ_i, m_j)) is calculated as follows:

$$p_{i,j} = \beta\big(1 - \sum_k (\delta_k \tau_k)\big)$$

where δ_k and τ_k are the proportion of newly consumed data amount and fraction of time spent at each progression step, respectively, and β is the data consumption regularity constant. In Figure 2.8, please see an example data-side message sending pattern of a given message type (λ_i) on a particular medium (m_j). t_0 is the arrival moment of user-side message to the data communication model, t_1 and t_2 are the first and the second response moments of the data communication model (i.e., data-side messages). Δt_1 is the delta time between timestamps t_0 and t_1. Similarly, Δt_2 is the elapsed time between timestamps t_1 and t_2. Δd_1 and Δd_2 are the amount of newly consumed data during time periods $(t_0, t_1]$ and $(t_1, t_2]$, respectively. And let us assume that D is the total amount of data that can be consumed by message series m_j, T is the total amount of time in seconds needed in order to respond to user-side message in a single step. Hence, for a given timestamp k, $\delta_k = \Delta d_k/D$, and $\tau_k = \Delta t_k/T$.

$\delta_k \tau_k$ multiplication generates low values as the amount of consumed time and data decreases. Hence, high number of progression steps along with the low fraction of consumed data leads to higher p values. Nevertheless, this formulation lacks accounting for fluctuations [22] that might occur in progressive analytic systems. Fluctuations are the anomalies in representations of the data that happen due to the disorderly data consumption or malfunctioning of adaptive sampling mechanisms (Section 4.2.1.2). Big difference between the amounts of consumed data at each progression step may lead to misleading flow of progression animations. Regular patterns of data consumption can alleviate or solve the problem.

Ideally, regular data consumption pattern would require equal amount of data processing at each processing step. If we plot such a consumption pattern as a function of time, we could see that the slope of a linear fit on that function would be equal to 1 (Figure 2.9). Based on this idea, I developed a formulation penalizing

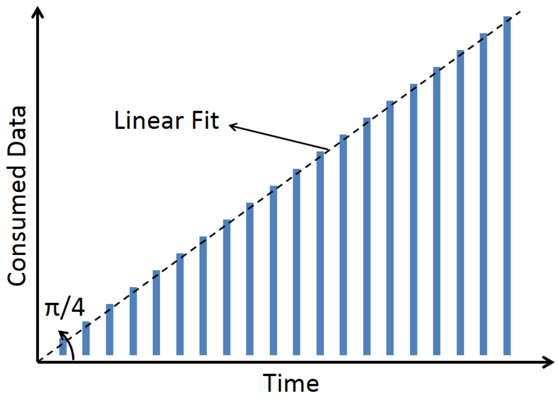

Figure 2.9: Linear fit on the ideal data consumption pattern.

the deviations of slopes from 1, and explain as follows. One might notice that, for this particular scenario, the slope (α) could have values in the range $[0, \infty)$, and $\arctan(\alpha)$ would map this range to $[0, \pi/2]$. Further calculations can lead to mapping to a range $[-1, +1]$ in which 0 correspond to slope 1:

$$-1 \leq \frac{4}{\pi}\left(\arctan(\alpha) - \frac{\pi}{4}\right) \leq 1$$

Complement of the unsigned value of the deviation value that would be found based on the above statement could serve as an indicator of the distance of slope from value 1. Hence the data consumption regularity constant is calculated as follows:

$$\beta = 1 - |\frac{4}{\pi}\arctan(\alpha) - 1|$$

Above formula generates values close to 1 as the slope of the linear fit on the cumulative data consumption function approximates to 1. Conversely, as the slope diverges from 1, β values closer to 0 are generated.

Progressiveness is evaluated regardless of the responsiveness of the system. For example, a very responsive system could both be evaluated as highly or lowly progressive. For example, an analytic systems solving a segmentation problem with large amounts of data could be attributed as highly progressive if generates partial results at every second based on partial data, and lowly progressive if shows a partial result right after receiving the user-side message and the final result after finishing the whole segmentation based on the rest of the data.

2.6 Methodology

Exploration of HDI space necessitates availability of structured or unstructured observation data from various analytic systems research. Unfortunately, availability of such public data is very rare, and available data does not seem to have common structure. Due to this reason, this thesis work attempts to *probe into* the exploration space, rather than *analyzing*, with the available data from five data analytic studies which I and my colleagues conducted during my doctorate education. The studies that I report here had been designed for various research interests with different structural and methodological approaches. However, as is common with all other research, we also collected data in various forms with the available capabilities. By making use of all the artifacts generated during the research conduct such as developed tools, collected human or sensory data, insights gathered during the design, realization, and evaluation of analytic systems, I was able to prepare a new dataset, which I will call *observation dataset*, pertaining to the analytic systems and their users' interactions with the data.

Observation dataset comprises

- Analytic system properties introduced in Section 2.5.2 (objective data),
- Self-reported reflections on the analytic system, engagement, sensemaking, and usability (subjective data), and
- Observations of users' interaction with the analytic system (behavioral data).

Preparation of observation dataset is a labor-intensive activity involving transcription of video and audio records, cleaning of sensory and system log data, and digitization of subjective and behavioral measures, which requires various methodologies such as controlled experiments and grounded theory as reported in Section 2.6.2.

Grounded theory is also applied to the behavioral and subjective data in the observation dataset in order to imply general principles about the analytic systems in the same realm.

2.6.1 Observation Dataset

The observation dataset is comprised of objective, subjective, and behavioral data.

2.6.1.1 Objective Data

As is accepted in social sciences, objective data involves bare observations from a phenomenon without any kind of human articulations. For example, number of correct answers given by participants of an experiment, or the average response time of a given analytic system are bare observations which are not affected by any human intervention, or subjective measures converted into countable or calculatable forms as a result of valid human transcription process. As reliable as they may be accepted, they could also bear misleading results especially due to carelessly designed experimental setups. Moreover, not all phenomena, such as creativity or emotions could be objectively counted or measured. Due to these reasons, they should also be supported by other forms of data such as subjective and behavioral measures.

Objective data that has been employed in my thesis is two fold: Calculated analytic system properties as explained in Section 2.5.2, and objectives measures that were collected during case studies or laboratory experiments of the reported studies such as number of insights made during the interaction with a particular system. Objective measures other than the properties of analytic systems are explained in detailed in the studies that I report in Sections 3, 4, and 5.

In order to calculate the analytic system properties, I had to establish a correspondence between the HDI communication model concepts and real user-system interaction primitives. Due to this reason, I determined the following characteristics of the interaction between the users and the systems that were employed in my studies:

- *Human-side Message Timestamps.* Any kind of user action in order to explore the data under any different condition (e.g., zoom, filter, rotate, position change of user, view details, pause/resume progression, aggregate) has been pinned as a form of human-side message. Such actions of users are identified, transcribed, and associated with execution timestamp.

- *Human-side Message Types.* Each human-side message has been mapped to a corresponding task type which are elaborated in Section 2.5.2.3.

- *Media.* Types and numbers of distinct media of the analytic systems has been determined based on the definition in Section 2.1, and formal expert review as explained in Section 2.6.2.1.

- *Medium Physicality Constants.* Based on the medium types, physicality constants have been determined based on a predefined procedure.

- *Data-side Message Timestamps.* Analytic tool responses to human-side messages have been transcribed as data-side messages. For example, representation generation, update, data consumption rate change, representation customization, and several other actions that are observable on the human-side are considered as data-side message, and each such action (i.e., message) has a timestamp.

- *Data-side Message Types.* Unfortunately, no task or action typology for what I call data-side messages seems to exist in the literature. Due to this reason, I developed a typology for data-side messages, and each transcribed data-side message has been associated with a task type.

- *Data Consumption Rates.* As an additional property of data-side messages, each action on the data side have also been associated with a data consumption rate. For example, if a data-side message executes a representation change due to a “zoom in” *by considering all of the data*, its data consumption rate would be evaluated to be 1. Similarly, if the action were to generate a partial analytical result with a quarter of the data, its data consumption rate would be 0.25. Data consumption rate is a measure independent from time.

2.6.1.2 Subjective Data

Self-reported measures of the users of our systems that I report in this thesis comprises Likert-type scales, structured and semi-structured interviews, expressions noted during the case and user studies. See Appendix-B for a detailed example of subjective measure collection methodology, and self-reported data. The subjective data employed in this thesis is reported in Chapters 3, 4, and Appendix-B.

I utilized unstructured subjective data in order to validate the task-system matching quality (Section 2.6.2.3). In order to explore implications on the success of the analytic system for the given tasks, I employed grounded theory methodology in the analysis of the subjective data.

2.6.1.3 Behavioral Data

The information acquired from the actions of users is called behavioral data. Number of times a student criticize her students [24], the pattern of where and when a customer make credit card purchases [25], the distance the human approach to a robot after a particular action of the robot [26] are all examples of behavioral data collection.

Behavioral data is typically characterized and employed in the analysis phase after a transcription activity of the collected raw behavior data such as video/audio recording. Transcription could also be done in real-time in the case of an ethnographic observation. The output of the transcription process is a set of coded data as a summary of the phenomenon under investigation. In order to prevent bias on the coded data, typically more than one coder are employed for the transcription task, and then, intercoder reliability check is performed.

2.6.2 Analysis and Realm Decision Methodology

The analysis and evaluation methodologies employed in my thesis are expert review, controlled laboratory experiment, insight-based evaluation techniques, and grounded theory approach.

2.6.2.1 Expert Review for Objective Data Collection

Expert reviews, focus groups, field and case studies are methodological techniques which are employed for qualitative evaluation of tools, systems or processes where quantitative approaches are not applicable due to lack of participants or cases. In some cases, there exist only a few users of a system in question making statistical assessment infeasible. Qualitative evaluation methodologies facilitates investigation of high-level cognitive tasks such as exploring and deciding on the next steps during the course of the interaction. Among those methodologies, I employed expert reviews in order to assess the extraction of objective data from the studies. Expert reviews [27] are usually conducted for the usability testing of given tools or systems. Few domain expert participants evaluate a given system by following specific guidelines or heuristics without any time constraint. The administrator of the study manages the session and takes notes and asks questions to experts in order to comprehend where the system succeeds or fails. Please see Tory et al.'s study [28] for examples.

The aim of the domain expert review employed in my thesis work was to objectively identify the media characteristics of the analytic systems based on the definitions given in Section 2.1. Experts initially interacted with the analytic tools, identified the tasks types, and then evaluated on which medium they sent their commands to the system. One might argue that such operation could easily be performed with novices such as the researcher conducting the experiment, however, in order to maintain the objectivity of the data expert review approach had to be performed.

2.6.2.2 Controlled Laboratory Experiments for Subjective and Behavioral Data Collection

Applied in almost all fields of science, controlled laboratory experiments are used to assess statistically significant inferences about a phenomenon based on the analyses of data collected in controlled environments. Typically, design of experiments involve identification of independent and dependent variables, and development of hypothesis statement. Independent variables are the cases that the experimenter is interested whereas dependent variables are the results that are possibly effected by the relevant independent variable. For example, we were interested in whether the gender has an effect on the height of human beings, independent and dependent variables would be *gender* and *height*, respectively, and "male" and "female" would be *levels* of the independent variable. Several other assessments need to be done for a proper controlled experiment such as identification of confounding variables, between-participant stratification, experiment execution strategy (e.g., within- or between-subjects), identification of measurements.

Controlled laboratory experiments have been frequently employed in the information visualization community [29]. Even though their reliability has been considered to be questionable in determining the usefulness of analytic systems [30], they are still of use in cases where specific conditions or models need to be quantitatively evaluated [31].

In the studies reported in Chapters 3 and 5, I and my colleagues employed controlled laboratory experiments in order to collect both behavioral and subjective measures on the usability and usefulness of the systems that we introduced. I reused that data in order to explain the realm in which the relevant study was located. In other words, the behaviors and subjective statements of users assisted me to facilitate general ideas and principles about the realms.

2.6.2.3 Insight-based Evaluation for Task-System Match Quality

The problem with the controlled laboratory experiments is that benchmark tests and performance measures such as task completion times need to be predefined by test administrators. Metrics such as accuracy and performance heavily rely on simple answers relating to potentially complex analytic systems. Unfortunately, simple answers lead to "Boolean usability" results which are typically two fold: Users liked the system or not, as stated by North [30]. He argues that main goal of utilizing an analytic system is to gain insight from data about a particular phenomenon, and further state that evaluation methodologies should measure the quality of analytic systems in terms of assisting their users in insight development.

North introduces insight-based evaluation methodology for information visualization systems, and Saraiya et al. applies this methodology in order to evaluate five microarray visualization tools with the domain experts [32]. During the interaction of domain experts with the tools, they identify the *research questions*, *insights*, and *hypotheses* gained or developed by the user. Each such finding is saved with its domain value, identification moment in the study, correctness, and several other such characteristics, and then the analytic systems were compared based on those characteristics.

For the evaluation of DimXplorer, a progressive visual analytics tool for exploratory analysis of high-dimensional data developed by Turkay et al. [22], we employed insight-based evaluation methodology. Please refer to Section B.3.2 for the saved insight and questions developed by the analysts during the case study that I report in Section 4.2.

Plotting the analytic systems in the exploration space with respect to their properties and visual judgment of their distribution cannot be useful without determining their success. Association of properties with a certain set of success metrics is necessary for the benefits I reported in Section 2.4. In this regard, evaluation of analytic systems in terms of their capability to support high-quality insight extraction has been performed for all the analytic systems reported in Chapters 3, 4, and 5, where I explain the realms that I introduce.

2.6.2.4 Grounded Theory for Finding Realm Characteristics

Proposed by Glaser and Strauss [33] in 1967, grounded theory is defined as a qualitative research method to develop a well-grounded theory based on unstructured yet systematically gathered data. In contrast to experimental research in which the data collection procedure relies on a preformed theory such as hypotheses, grounded theory starts with a set of empirical observation data collected with systematic and open-minded attitude. As suggested by Myers [34], creativity is the key to success of the grounded theory analyses.

The mainstream research adopting grounded theory follow a three-stage procedure. In the first stage, unstructured text or multimedia data is systematically collected with a set of preferred methodologies such as ethnography, and collected content is tagged or coded with a predefined or commonly used approach. Systematic and objective application of coding to all of the data and controlling for the subjective bias with intercoder reliability checks is vital for the reliability of the analysis. In the second stage, codes or tags are grouped together in order to form higher level concepts and categories, and then, the relationships between concepts are identified and the web of concepts is incrementally constructed while at the same time searching for higher level theories explaining these relationships. And finally in the last stage, inferential and predictive theories about the phenomenon under investigation are formed, and their verification might be performed with further coding of the collected data with a new set of coding scheme prepared based on the newly generated theories.

The aim of the qualitative analysis is to explain the important aspects of a problem, in detail, and based on collective unstructured data. In this regard, I applied grounded theory analysis on the behavioral and subjective data collected during the conduct of my studies with the aim of identifying important aspects of those analytic systems falling into the same realm. That analysis comprised systematic coding and identification of significant concepts pertaining to the properties of the studies. Based on these concepts, I explain general characteristics of the realms, and how these characteristics could inform the future analytic system designs that could be attributed to the same realm.

2.7 Application of Methodology to the Reported Studies

The methodologies described in Section 2.6.2 have been applied to all of the reported studies in my thesis to extract data, and realm identification and characterization. Not all the studies revealed the same set of observation data, and hence, analysis methodologies differ from case to case. Each such distinction will be clarified in the relevant sections of the following chapters. Below is the general overview of all the methodologies applied to the studies, and could be considered as the general overview of the rest of the chapters explaining realms.

2.7.1 Observation Data Preparation

Due to the different contexts and goals of each reported study, I had varying sets of readily available data prior to realm investigation. Due to this variety of the available datasets, I enumerate the available data at the beginning of each chapter, and explain the methodology to extract the unavailable part of the data.

For almost all of the cases, I had to apply expert review for the identification of human- and data-side messages and the extraction of objective data which were used to calculate the general characteristics of each analytic system as explained in Section 2.5.2. In doing so, the task types have been identified, and then the timestamps of human-side messages are tagged. As a result of expert review procedures, I determined the communication media level and physicality constants.

Self-reported judgments of users about the analytic tool and their interactions have been considered as subjective data. For those studies lacking subjective measures, a brief usability testing has been done in order to collect such data.

Behavioral data has been extracted from multimedia content collected during the conduct of the studies by applying tedious transcription process. Behavioral data comprised interaction patterns, and users' successful or failing actions.

2.7.2 Applied Analysis Steps

Following the data preparation information is the description of the steps taken in insight-based evaluation for determining the task-system match quality, and grounded theory for developing the general concepts about the identified realms.

Insight-based evaluation requires observation of intense human-data interaction activity. Based on the observation, characteristics of hypotheses or insights are extracted from the observation data by applying a certain set of methodology. In other words, insight-based evaluation requires conduct of some form of case study with expert human participants, which is not the case for all the studies reported in this thesis. For such cases, I carried out structured interviews as well as focus groups in order to approximate the conclusion that would otherwise be drawn with an actual insight-based methodology.

As for the insight-based evaluation, grounded theory is also conducted with observation data, however, from various resources such as interviews, scales, probes, ethnography methodologies, and controlled or quasi experiments. Such variety of the data is managed by rigorously objective application of widely accepted analysis processes typically done with human effort. Scientifically soundness and reliability of the results are strongly connected to application of commonplace methods by two or mode analysts, whose coded results are compared with intercoder reliability checks. Unlike the case for the insight-based methodology, I was able to directly apply grounded theory to all studies as the observation data was available for each of the analytic system.

2.7.3 Realm Identification and Characteristics

As explained in Section 2.5.1, realm identification is a clustering problem, and, for robust segment identification, reasonably intense data should be employed. In this case, an individual datum for clustering an analytic system and its corresponding research observation data, and only five analytic systems' data is available. Due to this reason, identification of realms are merely decisions based on visual distributions of the study. Therefore, each analytic system's observation data should be treated as *probing* into the exploration space, and realm identification and characterizations are approximations to what would be achieved with high amount of observation data. The principles and affirmations formed during the grounded theory processes are accepted as the realm characterization.

CHAPTER 3

HYBRID SYSTEMS REALM

Hybrid systems realm comprises analytic systems having moderate level of responsiveness, and supporting human-data communication on a set of media comprising both digital and physical parts. Such set of analytic systems seem to establish high level of engagement with the analytic system due to relatively high responsiveness. The experimental results imply that the partial physicality of the analytic system fosters the engagement due to the increased modality of the communication.

Systems in this realm necessitate moderate level of responsiveness in order to maintain the collaborative aspect, however, high responsiveness seems not to be a requirement. This flexibility permits adoption of progressiveness mechanisms which could be run in the background in between the loosely distributed interactions of the users.

Analytic systems and the associated studies that were employed for the identification of hybrid systems realm are reported in Sections 3.1 and 3.2. Having been designed for diverse set of purposes, these studies introduce the analytic systems in detail, and should be considered as a formal description of the observation data collection methodology. Characterization of hybrid systems realm necessitated intensive transcription and analysis effort of the observation data, which is explained in Section 3.3. The application of grounded theory and insight-based evaluation methodologies for realm characterization are explained in Section 3.4.

3.1 Study-1: Low-fidelity Prototyping with Simple Collaborative Tabletop Computer-aided Design Systems

Design processes encompass iterative elaboration and elimination of new and many ideas gathered from a wide range of resources. The higher the diversity of the resources, the higher the chances that the design process will bear expected outcomes. Following that idea, immense amount of effort has been devoted to the development of collaborative computer-aided design (CAD) systems, and process frameworks that drive those systems. We infer from the existing literature that collaborative CAD solution attempts involve holistic approaches in which all aspects

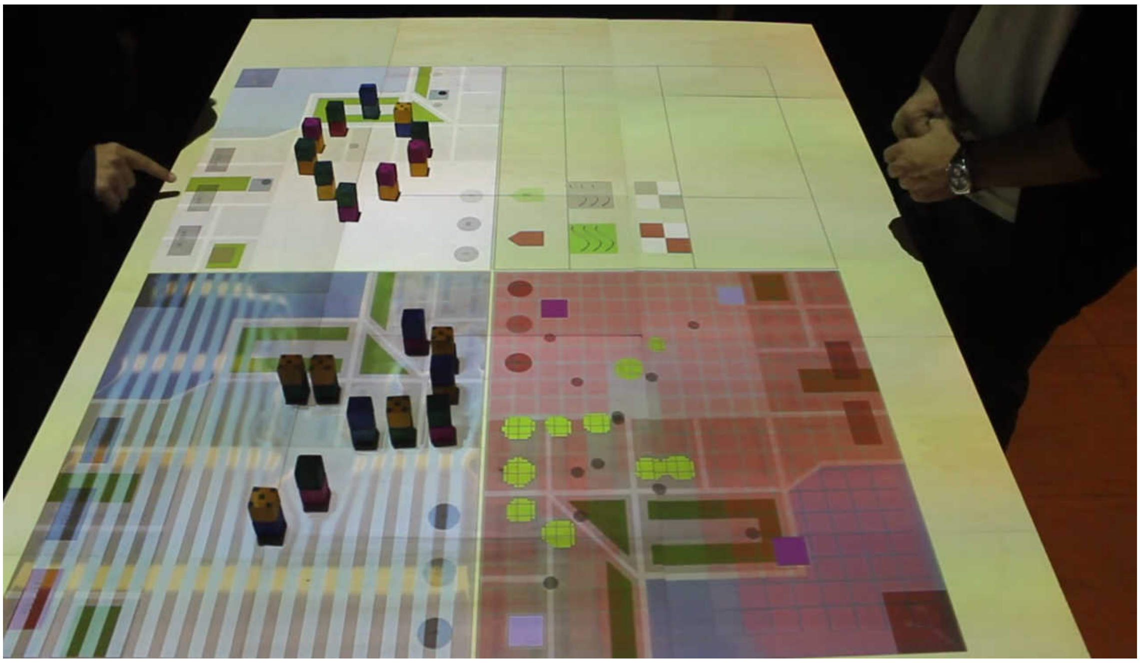

Figure 3.1: InitialInsights: A collaborative tabletop computer-aided prototyping setup. The setup supports multiuser interaction and rapid prototyping iterations.

of the problem (social and technical) are being addressed. As an attempt to address social and physical aspects of the problem, tabletop systems with complex structures have been proposed in previous work. Unfortunately, such complexity comes with the lack of reproducibility of the research work, and high evaluation overhead per prototype imposing a low limit on the number of design ideas to be investigated. Sophisticated systems might be required to solve the real-world problems, however, we argue that, with simple setups, rapid collaborative iterative prototyping could be achieved. Such simple setups could lead to high number of *good* ideas ready to be fed into off-the-shelf CAD systems lacking adequate support for collaborative design. We realized and evaluated this idea by implementing a tangible tabletop collaborative design system that facilitates fast and iterative prototype production for residential area design. Based on the case studies conducted with this setup, we show that synchronous collaboration for rapid prototyping could be achieved with lean setups, provide a list of design recommendations for such systems that we derive from our case study observations and existing literature, and finally contribute to the community with an open source tangible tabletop installation tool kit.

3.1.1 Previous Work

Collaborative design attracted much attention from many researchers as a challenging and complex problem [35]. Collaboration in design context seem to have many aspects such as the theoretical frameworks and processes [36], [37], technical issues or solutions [38]–[40], and social problems [41], [42]. In general, technical approaches involve integration of design processes [43] and the establishment of shared understanding in order to raise the perception [44].

Collaborative idea development has been an attractive topic since the emergence of computer-aided tools to address engineering design problems. C-Sketch [45] is an early attempt to explore collaborative idea generation processes with detailed lab studies providing inspiration for further studies [46].

Quick idea sharing and fast prototyping are key factors for the success as appreciated by the industry. For example, CAD tools such as SketchUp^{TM}[47] and AutoCAD^{TM}[48] have been excelled in fast model transformation and idea sharing allowing iterative low-fidelity prototyping at the early stages of the design process. We argue that our approach differs from these kinds of tools in several ways. InitialInsights has been designed to support trial (and elimination) of many different scenario ideas. With almost real time tangible interaction, it supports quick evaluation of *what-if* scenarios. On the other hand, SketchUp excels on design and production of models rather than working with scenarios. Moreover, most of the off-the-shelf products seem to support online or on screen sharing of ideas whereas InitialInsights supports synchronous local collaboration with physical objects and tangible almost real time interaction. And finally, InitialInsights aims at rapid and *concurrent prototyping* cycles on various scenarios in a collaborative fashion (e.g., wind and pedestrian movement) whilst SketchUp seems to focus solely on sequential prototyping and addresses problems of the later stages of the design process.

Involving multiple users (co-located or in remote location), the collaborative computer-aided design processes are inherently social interactions which happen on computer-mediated communication tools on remote collaboration [49] or face-to-face when the designers are co-located [50]. Several groups reported promising results in remote collaboration [51], [52], however, they still seem to be far from being common commercial solutions due to the major social problems such as communication [53].

As suggested by [54]–[56], tabletop systems have the potential to facilitate a social and collaborative working environment and enhance the idea sharing during the analysis and design processes as a remedy to the communication problems emphasized by the literature [57], [58]. Up to date, numerous approaches have been integrated with the CAD systems to benefit from the capabilities of tabletop systems with different focuses and challenges such as implementation of augmented reality on tabletop systems [59], [60], gestural interaction [38], tactile interaction with the digital 3D model and also projection of the simulated information onto three-dimensional model [61], and real time tactile interaction for controlling 3D terrain models [62], [63].

Tabletop approaches aim to alleviate the social problems by facilitating co-location of the designers. Furthermore, current body of research has several other alternative enhancements for human-related issues of the problem. For example, such approaches involve the use of design templates to eliminate possible mistakes [64], application of integrative group decision making [65], experience-based design team formation [57], and well-structured communications [66], [67].

3.1.2 Low-fidelity Prototyping with Tangible Collaborative Tabletop CAD Systems

In this study, we stress the simplicity and reproducibility of our approach for introducing rapid and iterative low-fidelity prototyping for design process. Simple yet effective systems will not only facilitate feasible level of performance during the fast *design-feedback* cycle, but also better help CAD research community improve current state of the research.

Counterintuitively, however, designing *the simple* while providing the expected functionality might turn out to be complicated. This is what we experienced when we developed our low-fidelity prototyping tool, namely InitialInsights. Here we list the design recommendations (**DR**s) that could assist the designers in producing *good-enough* solutions based on the existing literature and a case study that we performed. We also refer to inferences that we made from the case study explained in Section 4 whenever appropriate.

3.1.2.1 Communication

Collaborative design systems should promote high level of social capabilities such as communication and shared understanding. In order to achieve this ambitious goal, the system should support high interactivity and face-to-face communication which is attributed as the golden standard [68].

DR-1 *Support between-designer communication at the highest possible quality.*

3.1.2.2 Responsiveness

In their seminal work, Card et al. introduce the concept of human time constants [15]. These constants are attributed as the important limits that should be applied in interactive systems in order to facilitate reliable communication between human and computer. The *perceptual processing* level corresponds to the perception of still images as animation by humans. The concept requires iterative systems to have at least 10 Hz update frequency. As the second rule of the concept, *immediate response* level is introduced. The communication between human and computer is assumed to be interrupted if computer fails to respond in less than one second. Card et al. state that the same effect is also valid for the human-to-human communication. The third level, *unit task* constant (10-30 seconds) corresponds to the time in which a unit task is to be completed.

In order to achieve continuous interaction both in-between the designers and between designers and the tool, human time constants should be applied in the implementation. In particular, the *immediate response* constant seems as a crucial factor that can facilitate fast and productive prototyping iterations.

DR-2 *Responses to interactions should obey the human time constants in order to sustain the rhythm of the prototyping cycle.*

3.1.2.3 Synchronization

Low-fidelity prototyping typically requires synchronized creative team work. Such team work should be supported with concurrent interaction and rendering functionalities.

DR-3 *Facilitate collaborative design environment by considering all aspects of synchronization for interaction, data processing, and rendering.*

In the multi-user interactive tabletop systems, the common case is that the race condition might occur during the user interaction. We also observed this situation in the earlier versions of InitialInsights even when the conventional concurrency methodologies were applied. In order to prevent race conditions while obeying the human time constants, priority hierarchies of the available resources such as shared memory spaces should be designed.

DR-4 *Design the hierarchy of priority of the consumers of the resources in order to prevent race conditions which might occur during multi-user interaction.*

3.1.2.4 User Friendliness

High number of design ideas in limited time is the expected outcome of low-fidelity prototyping. The time that will be spent while learning the tool might increase the overall cost per design idea. Even worse, the complexity of the prototyping tools might disincline the team members to use the tool. In that regard, the prototyping tool itself should be user friendly and easy to learn. It should be able to present affordances inspired by domain-specific features.

DR-5 *Provide the designers with easy-to-learn tools and graphical user interfaces.*

3.1.2.5 Support for Individual Contexts

Cooperation between the users and engagement with the system may dramatically decrease if the inputs of different users are not handled according to their supposed functions. Similarly, creation of interaction context customized for each user expertise should also be considered as designers in the group can be assigned to different tasks. For example, in the case study that we report in Section 3.1.4, designers could momentarily switch between the *wind flow* and *pedestrian flow* modes. Quick mode switching capability facilitated not only a flexible work share among the designers, but also maintained the rhythm of the prototyping cycle.

DR-6 *Provide customized and easy-to-switch contexts for the designers.*

3.1.2.6 Provenance and Versioning

Oftentimes the problem associated with the collaborative independent work is the version control of the work in progress and merging the checkpoints of different parallel works. Occasionally, depending on the problem, re-creation of a particular combination of artifacts comprising a version of a product can be very compelling, if not impossible, which usually causes the loss of valuable findings or difficulty in the documentation of the process.

DR-7 *Enable mechanisms to re-create, tag, and manage the versions of the design process.*

3.1.3 InitialInsights: A Collaborative Tabletop Prototyping Setup

The hardware setup of InitialInsights is depicted in Figure 3.2. The setup is comprised of a desktop computer, Kinect for Xbox One [69] as RGB-depth camera, an HD projector, and a wooden table without any sensor or electronic parts. The software part of the InitialInsights is basically a Processing [70] application which takes all its user input via RGB-depth camera, sends its output to a projector.

The ability to capture the user input from the bare wooden surface of a table with no additional electronic appliances is a valuable capability for the purposes of quick iterative design for the initial phases of a product design. Along with the input detection library that we developed, we argue that our lean hardware setup will be able to serve as a base for low-level prototyping in several design domains (Section 3.1.3.1). In this study, as an example, we chose to develop a residential area prototyping platform for which we additionally needed to build 3D surface detection, wind flow design model, and pedestrian flow design model (Section 3.1.3.2).

3.1.3.1 Input Detection Capability

RGB-depth camera had the capability to generate depth and infrared maps of a given view area with high frequency. A thread in InitialInsights continuously scans and processes these infrared and depth maps in order to locate and track the changes happening in the viewport. For given two time intervals t_0 and t_1, let us assume that the depth (D) and infrared (I) maps captured by Kinect are D_0, I_0, D_1, and I_1. The

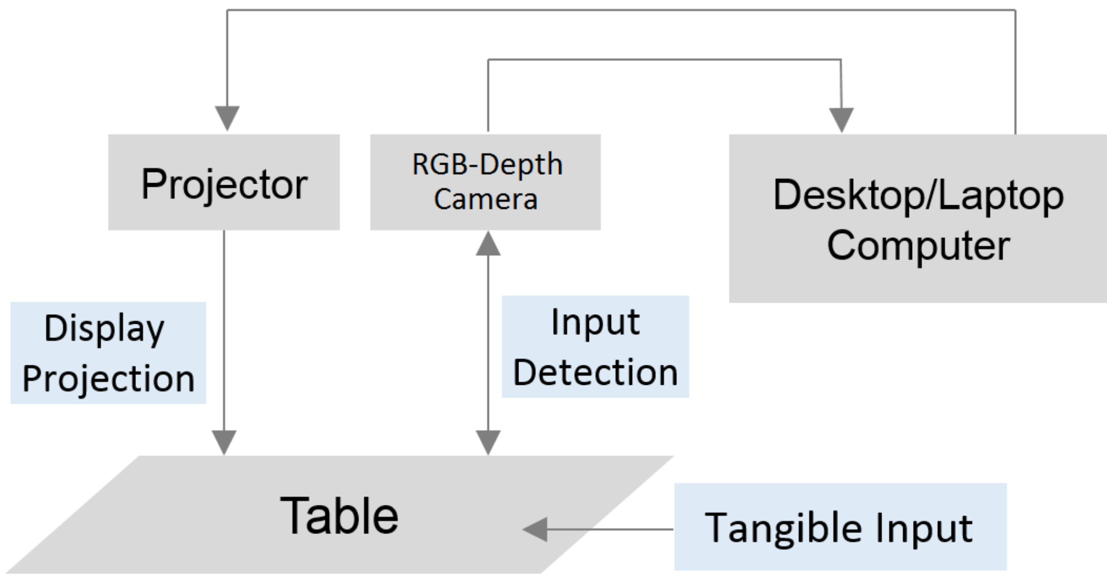

Figure 3.2: Hardware setup of InitialInsights. Simple collaborative interactive low-fidelity prototyping setups can be quickly built with a computer, projector, and an RGB-depth camera.

delta images of depth (ΔD) and infrared (ΔI) maps are calculated as $|D_0 - D_1|$ and $|I_0 - I_1|$, respectively. The delta images are binarized with respect to a threshold (τ) and merged into a composite delta image ΔC with the $(\Delta D > \tau) \vee (\Delta I > \tau)$ operation. Composite delta image (ΔC) is fed into the finger detection module developed by utilizing Processing FingerTracker library [71]. After elimination of small defected contours, a vector of the finger tip point locations and major contours are returned by the module. The index finger of users are considered as the one used for the interaction (e.g., pressing button).

Touch Detection—Whenever the index finger's location (i.e., a pixel) is updated by the finger tracker module, the depth values of the neighbor pixels around the index finger location are aggregated into a single average depth value (d_f). Similarly, on the base depth image (D_b), average depth value of the pixels that are in the neighborhood of index finger location are averaged into a single value (d_b). If the difference $|d_f - d_b|$ is below a certain threshold, the index finger is considered to be touching the table. The aggregation process has been a necessity due to the relatively high level of noise of the data acquired from RGB-depth camera.

Object Detection—The physical objects placed onto the table are basically detected with the difference maps of the depth images acquired before and after the placement of the objects. In contrast to finger detection, object detection is a process initiated by the user and processed only once per refresh request. Such different mode of processing has proved to facilitate better user interactivity by clearly separating the hand detection from object detection.

Let us assume that the depth maps before and after object placement (or displacement) are D_0 and D_1, respectively. The difference map (ΔD) is calculated as $|D_0 - D_1|$. In an ideal world, we would expect the depth difference to be zero for the regions of the table where no physical difference was made, however, we observed that the noise in the depth data leads to further processing in order to achieve the actual physical changes. In order to locate physical objects placed on the table, firstly, we applied a filtering based on empirically identified threshold τ such that the difference values below τ are set to zero. Following this operation, we applied a Gaussian-like filter on the difference map in order to smooth out the noise existing on the upper surfaces of the placed objects.

Buttons—The buttons are presented to the users as geometrical objects in the InitialInsights. Each button has its own listener and worker thread which are activated whenever a user is considered to touch the location inside the region of a button. Having separate threads for each of the buttons ensures required performance for multi-user interactivity. The race conditions on the update of shared data sources are resolved by making use of a work queue in the main thread to which all other threads report their data manipulation actions. Whenever the user interface is updated (e.g., mode change), all the relevant buttons are erased or redrawn on the viewport by the main thread, and new worker threads are assigned to the newly generated buttons.

3.1.3.2 Modules Developed for Residential Area Design

Residential area design is the example case that we preferred to demonstrate on InitialInsights. Having the input detection mechanism and hardware setup, the development of residential area design case has involved the development of the following modules.

Pedestrian Movements—Pedestrian movements are modeled with the concept of moving two-dimensional objects (e.g., circles) from user-defined source to destination locations which are marked as rectangular areas. Users can add as many source and destination pairs as they desired. The pedestrian movement model also enables the sharing of sources or destinations between pairs. For example, pedestrian flows originating from a certain source can *find* their way towards two or more destinations.

The pedestrian *agents* moving between source and destination points are also aware of the physical objects placed on the table such that the agents can find their *shortest* paths to destinations by circumventing the obstacles (i.e., physical objects). In doing so, the physical objects located by the input detection module of InitialInsights are fed into the pedestrian flow model. The shortest path is searched with the A* algorithm working on a *maze* created with square grids whose edges are

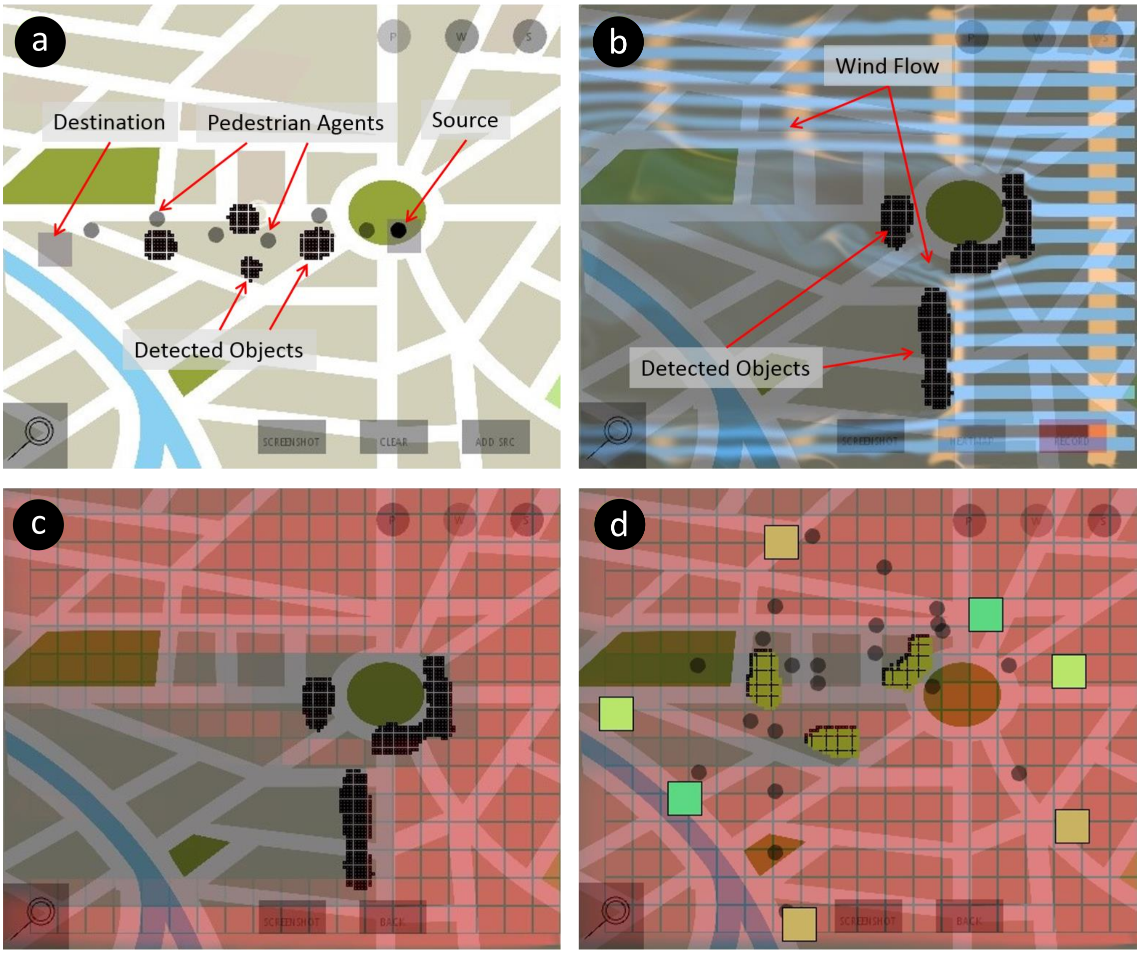

Figure 3.3: Modules developed for residential area design. The designers could build (a) pedestrian flows by adding source and destination locations. Pedestrian agents follow the shortest path between the source and destination without crossing over the physical objects. (b) Wind flow module simulates the wind flow which is effected by the physical objects placed on the table. (c) Wind flow exposure could be calculated by accumulating the periodically sampled wind force vectors, and visualized as heatmap. The moderator could merge the models designed by the designers in (d) the composite view.

1 pixel. At each step of the agent movement, the neighboring grids are enqueued in a priority queue, and next best step is decided based on the items in the queue which are not part of the obstacles detected by the input detection model.

Wind Flow and Exposure Prediction—Wind flow model has been implemented based on the PixelFlow library [72] which is developed for high performance GPU-computing. The obstacles detected by InitialInsights are fed into this library as blocks standing against the wind flow generated by the library. The library also manages thousands of wind force vectors uniformly distributed on the viewport. For our residential area prototyping setup, we utilized those vectors in order to calculate accumulated wind exposure levels for different parts of the viewport. The wind exposure levels are then visualized via a heatmap ranging from transparent to red color in order to represent the range from low and high level of exposure.

Control Panel—Our residential area prototyping tool enabled the designers to generate and send model (e.g., wind, pedestrian) outputs to the moderator so that he/she can evaluate designs sent by the designers by investigating the outputs individually or merging them into a composite view. The merge operation collates (i.e., alpha blending or overdrawing) the outputs from different designers onto each other in a specific order in which the pedestrian model output is always displayed on top. The control panel allows the moderator to quickly switch between different views.

3.1.4 Evaluating Design Recommendations

In order to evaluate the validity of the design recommendations, we conducted a case study in which we observed three students (two undergraduate and one graduate) and an assistant professor all from architecture domain performing residential area design.

At the beginning of the study, we conducted semi-structured interviews with the participants in order to learn about their past collaborative design experiences. We also requested them to make design speculations about how a collaborative prototyping tool should be designed. Upon the interviews, we introduced them our tabletop collaborative prototyping tool and carried out a training session to help them get acquainted with the tangible interactions. During the case study, we did *fly-on-the-wall* observations [73] without making any interventions in order to avoid any bias in their design processes. During the case study, participants worked on the same design problem (i.e., residential area design) together interactively. We recorded all the sessions and transcribed the conversations that had effect on the course of the design activity.

3.1.4.1 Residential Area Design Scenario

Residential area design is a dynamic process that involves many aspects such as placemaking, drainage systems, natural disaster risks, parking areas, landscape, sunlight optimization, and crowd management. Practically, the design process is challenging due to the necessity of collaborative work, time constraints, different skills of the designers, and capabilities of the tools at hand. The *whole* design process should consider all such factors, however, putting a limitation on the number of the initial ideas that could potentially be best candidates for the final design choice. To alleviate this problem, physical low-fidelity prototypes could be used, however, this could put another limitation on the number of the design ideas since physical model generation requires considerable amount of time.

For the design of InitialInsights, we employed *pedestrian movement* and *wind effect* factors during the initial residential area prototyping process. We established a design setting in which the users could generate their own residential area prototypes, manage their own design context, and share their findings with a moderator who could merge and compare effects of different factors on their own screen (Figure 3.4).

During the case study, teams with three participants were formed where two of which were assigned as designers and the remaining one as the moderator. The *designers* were expected to build various residential area design prototypes with different distributions of building blocks. They were able to improve their models with pedestrian movement and wind exposure models. During the time they modify their prototypes, they were able to share the momentary state of their model with the moderator. The *moderator* was able to view one or more of the wind exposure, pedestrian movement, and building block distribution models in his/her own viewport. He/she was also able to switch between the views of both *designers*. The participants frequently changed their roles during the design session so that the stratification between the roles and the participants were realized.

3.1.4.2 Design Tasks

The design cycles of the designers and the moderator required a number of tasks that could be grouped as in the following paragraphs.

Task-1: *Placing Building Blocks.* Designers introduce the building blocks' locations and sizes to InitialInsights by simply placing them on the table inside their own viewports. Each *designer* could place any number of building blocks in any possible physical distribution.

Task-2: *Recognize Residential Distribution.* Whenever a new distribution of building blocks was done by the *designers*, it had to be recognized and prepared by the InitialInsights. After recognition of the location and sizes of the building blocks, pedestrian and wind models were initialized with the new distribution.

Task-3: *Generate Pedestrian Flows.* Pedestrian flows were created with the definition of a number of source and destination points. *Designers* could generate as many source-destination pairs as they desired. Furthermore, they could make additional destination bindings to existing sources.

Task-4: *Save Pedestrian Flow.* At any moment during the design process, *designers* could capture the momentary state of the pedestrian movement, and send that information to the *moderator*.

Task-5: *Generate and Save Wind Flow.* Wind flow could be initiated at any moment by the *designers*. Based on the distribution of the building blocks, the wind model reflected the flow through and around the blocks. As in Task-4, designers could capture any moment of the wind flow and send it to the *moderator*.

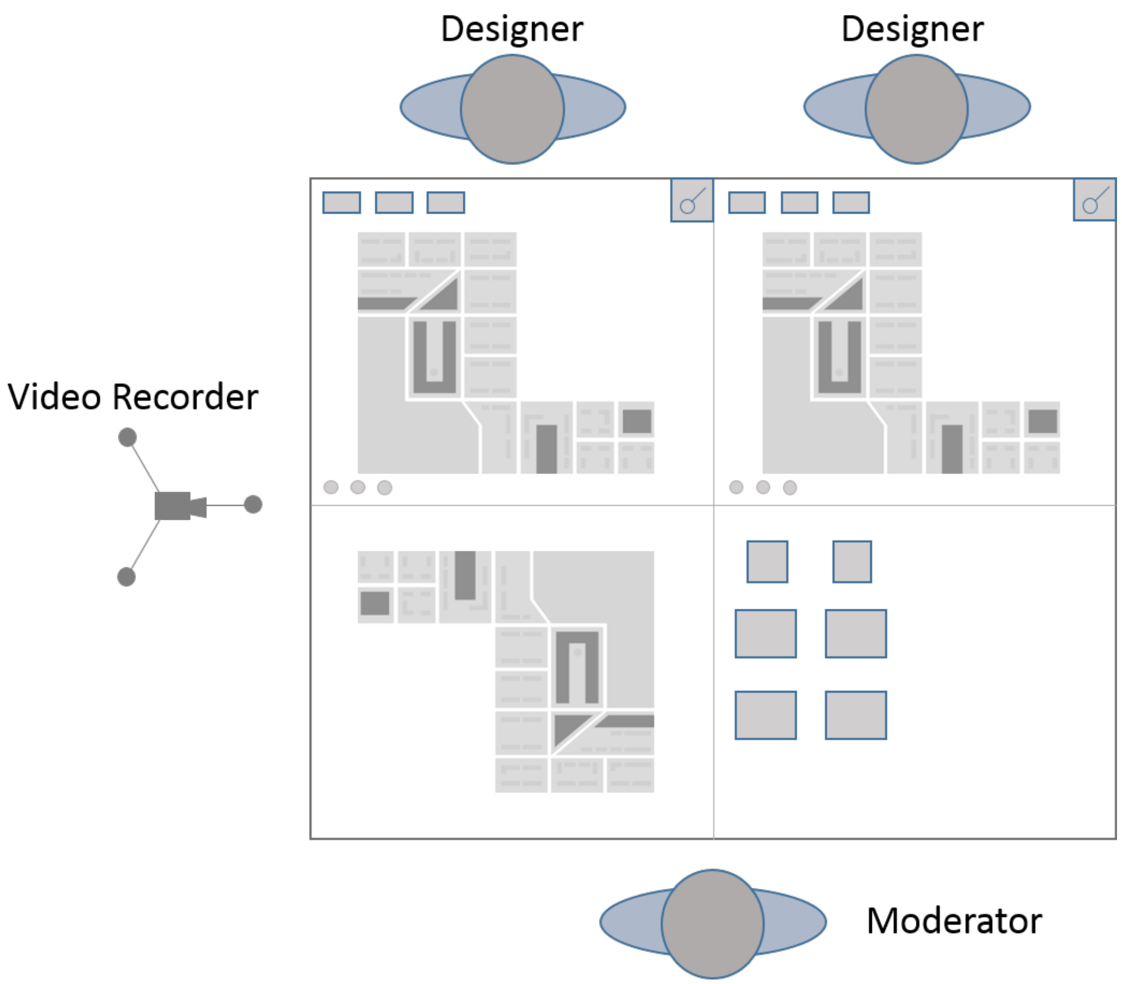

Figure 3.4: Experimental setup. The case study was conducted with a design team which included two designers and a moderator. Designers iteratively developed many models and shared with the moderator instantly. The moderator could investigate the models of the designers individually or in composite views. The whole session was video recorded.

Task-6: *Predict Wind Exposure. Designers* could calculate the wind exposure by saving the wind flow for a certain period of time. During the wind flow save operation, InitialInsights periodically collected the effect vectors' cardinality, and aggregated these cardinality values into a composite wind exposure score. Wind exposure score distribution with respect to location was visualized with a heatmap representation ranging from transparent (no exposure) to red (maximum exposure).
Task-7: *Save Wind Exposure. Designers* could save and send the wind exposure information to the *moderator*.
Task-8: *Merge Models. Moderator* could merge any possible combination of the pedestrian flow, wind flow, wind exposure, and building block distribution. The merge operation rendered the combined models in a way that representation of each model could be correctly observed.

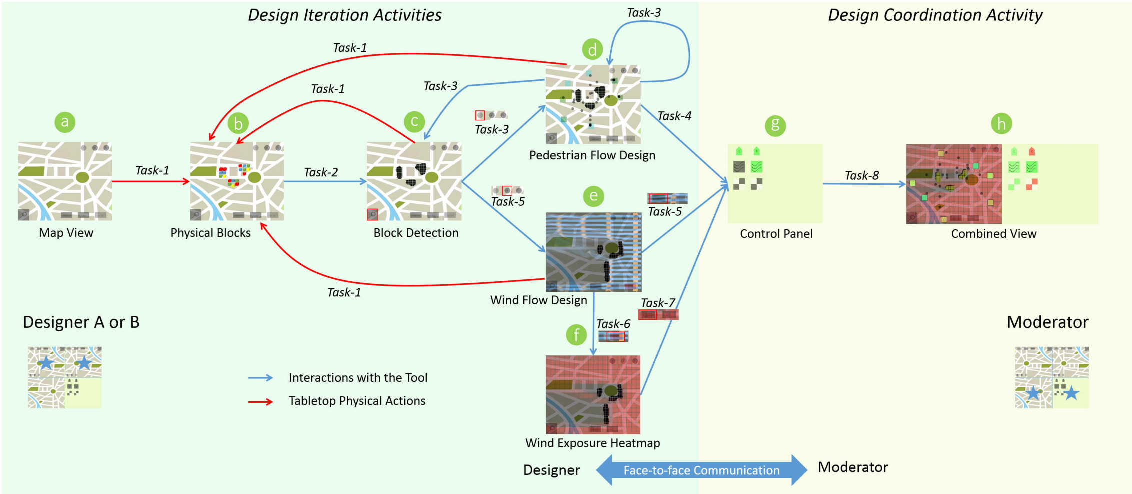

Figure 3.5: Design and evaluation workflow. Depicted is the typical workflow of the residential area design process on the InitialInsights. The tool always showed (a) a city map in the background. The users could (b) place physical objects (building blocks) on the table at any moment of the design cycle. In order to make the objects effective, they had to (c) initiate object detection. Upon the detection of objects, designers could either start (d) pedestrian flow design or (e) wind flow design. While in the wind flow design mode, designers could also make the tool calculate (f) a wind exposure heatmap. At any moment of the design cycles, designers could send their current design to the moderator. Whenever the moderator received prototypes from the designers, (g) the elements on the control panel were highlighted in order to reflect the prototypes available to the moderator. With the help of control panel, (h) the moderator could either investigate the prototypes separately or merge them into a combined view.

3.1.4.3 Prototyping Workflow

InitialInsights allowed *designers* and *moderator* to follow infinitely many design workflows. Figure 3.5 depicts a typical workflow that was performed by our participants. *Designers* could generate any combination of the aforementioned model outcomes (e.g. wind exposure, pedestrian flow) in any order and send it to the moderator. Likewise, the *moderator* could view the aggregated model whenever a model outcome was sent by the *designers*. For example, *designer-A* (*DA*) distributed the building blocks in a particular way, generated a pedestrian flow (*A-Pf*), and finally initiated a wind flow (*A-Wf*) and shared all the outcomes with the *moderator* (*M*). At the same time, *designer-B* (*DB*) distributed the building blocks in a different way and generated a wind exposure model (*B-We*). *M* could view *A-Pf* and *A-Wf* in any order, or merge these two models and view the aggregated model to see any possible combined effect. Nevertheless, *M* could not view *B-We* since it was not sent by *DB*.

3.1.5 Observations and Discussion

The case study was conducted for investigating the design recommendations rather than the capability of the InitialInsights as a prototyping tool for residential area design. Therefore, emphasis was more on the behavioral observations which we associate with the design recommendations (DR) discussed in Section 3.

As is expected, tabletop environment of the prototyping promoted high level of social interactions. Except for rare technical issue moments, the participants covered whole session period by arguing about the improvement of the design decisions that were being currently investigated.

Physicalization of the design environment also arose a feeling of entertainment among the team members which in turn promoted creativity and solution building. Interestingly, participants formed a game-like set of rules for their prototyping process by themselves without being urged to do so. The rules involved some of the domain specific attributes such as *'All building blocks should have a lake view.'* or *'In this round of prototyping, designs should not involve more than 7 blocks.'*. They stated that rules helped them focus better and increased their productivity. During the case study, our participants tried 127 low-fidelity *design ideas*, and overall they were able to try and evaluate each design idea in around two minutes ($Mean$: 136.7 *Sec.*, SD : 52.3 *Sec.*).

One of the designers also mentioned that, in their own work environment, they usually brainstorm a new idea by just bare discussion and often decide on its feasibility without trying on a map sheet. The preparation of a map sheet having a particular residential area design drawn is considered as time consuming task. They stated that having a tool similar to InitialInsights during the whole design process of a project may accelerate the iteration and save considerable amount of time.

With the help of the physical representation of the residential area design, all designers were able to view a particular instance of building block distribution from different views. They discussed that the perception is promoted with the physical objects used on the residential area map resulting in the consideration of minor details in the design that were otherwise not noticeable. For example, they were able to better perceive the directions of the windows or doors of the buildings which is an important aspect in the large scale residential complex planning.

3.1.6 Discussion

We conducted this case study with the aim of identifying the methods and guidelines that may form as an attempt for a collaborative tabletop prototyping tool framework. In order to achieve this goal, we developed InitialInsights, and expected a group of participants to design and develop landscape solutions, find relationships between open and closed spaces with respect to possible pedestrian movement, and wind flow scenarios. Through the development process of InitialInsights and the case study conduct, we gained invaluable insights on and experience in the design of collaborative tabletop prototyping tools.

In principle, low-fidelity prototyping is a process which elaborates the whole production process by facilitating evaluation of many initial ideas. In order to achieve high throughput for idea investigation, the tools and materials that are used in the prototyping process should be easy to use and manipulate. For example, for the development of a web interface, large number of different paper mock-ups could be generated and tested in smaller amount of time compared to what would be spent for the full web page implementation. In a similar fashion, we designed InitialInsights tabletop collaborative prototyping framework with only a few hardware components, namely a high-resolution projector, an ordinary table, and an RGB-depth camera. As explained in Section 3.1.3, the hardware setup involves manual basic projection adjustments in order to gain best focus and zoom levels. From our point of view, the only challenge could be the interaction mechanism implementation which we share as a fully-documented open source project. In that regard, our study contributes to the community with an easily deployable collaborative tabletop system setup tool kit and the relevant documentation. Current state of our framework allows developers to build their own prototyping environments by just implementing their domain-specific design context (e.g., residential area design, urban planning, container ports). Furthermore, we aimed at making the collaborative design set up as reproducible as possible which led us to design a system with less constraints and limited requirements. We claim that, for the reproduction of InitialInsights setup, no expertise will be necessary on either tabletop systems or tangible interfaces.

Much attention has been devoted to remote collaborative design solutions, and similarly, many case studies report that communication is still an important challenge in the remote collaborative prototyping. We argue that not always remote collaboration could be the best choice, particularly when the designers would like to develop and test initial ideas at the early stages of a product life cycle. In such cases, the cost of gathering team members for a short iterative prototyping process can outweigh the costs of unsuccessful product design due to the associated problems of remote collaboration.

We emphasize that the lack of collaborative prototyping support of state of the art CAD tools could be due to the fact that they are designed to bring the early and late prototyping phases of the design cycle in the same context. Early design stages require attendance of all team members work in short cycles, which necessitates fast and simple tools that can help production of good-enough early models of the design. In contrast, late stages of the design process involve usage of sophisticated *production* tools to bring a prototype close to a product. This stage of the design cycle typically involves individual work targeting the completion of a list of requirements until a project checkpoint meeting. In this study, our goal is to show that these two stages of the design cycle could be tied together by making use of the best parts of both worlds. We should still finalize our design with commercial CAD tools, however, we can also utilize the power of low-fidelity prototyping tools. For example, after a prototyping session, InitialInsights could generate design metafiles which could be fed into main design tools in which the high-fidelity prototyping or final production could be done.

The participants of our case study stated that, as part of their coursework, the design process of an urban or residential area involved (a) 2D or 3D physical model generation (e.g., home/building models, geographic formation templates, map sheets), (b) digital modeling on CAD programs (viewable on computer/tabletop displays) and validation, and (c) updating their design decisions. They reported that digital modeling involves immense amount of effort so that they usually feel reluctant to make major changes. InitialInsights seems to combine all these design phases into a single cycle as explained in the workflow in Section 3.1.4.3.

One might notice that the design of InitialInsights seems not to follow the third human time constant mention in Section 3.2, which declares the constant as being a value between 10-30 seconds whereas our observed average design iteration duration during the case study is around 2 minutes. With the third human time constant, namely unit task constant, Card et al. [15] emphasized the necessity of a limit on the unit task completion time varying depending on the goal and context of the human-computer interaction rather than providing a specific recipe applicable under any circumstance. Due to this reason, they avoid being specific for the third constant (i.e., 10-30 sec.). For instance, the third constant could be a few seconds for a tactical display of a real-time command control system of a warship while it could rise up to tens of minutes during final product design on a CAD system. We suggest that it could be a sound approach to suggest the readers to follow a limit on the duration of the unit task rather than imposing a specific constant. The lower the fidelity of the prototyping, we suggest, the lower the duration of the unit task.

We propose InitialInsights as an implementation of the concept of rapid collaborative prototyping framework supporting tangible objects and interaction. It not only supports idea sharing among designers, but also enables *collaborative*, and *practically rapid* and concurrent prototyping cycles. In general, we state that its main purpose could be the trial and elimination of tremendously many ideas until having a set of good candidates. Furthermore, synchronous and independent design evaluation features, which can be considered as *conflicting tasks*, are also supported. For example, designer A can work on the wind effect scenarios while designer B can work on pedestrian movement patterns. We claim that handling of conflicting tasks and concurrent activities while maintaining high quality of communication channel should be accepted as the vision of the next generation of CAD systems. Nevertheless, it should also be noted that task independency and conflicting task handling bear a complication in the consolidation of the *subsolutions* which should definitely be coordinated with human mind, hence the vitality of communication.

Sophisticated programs and tools are of necessity for final products or high-fidelity prototypes, however, we argue that with simple setups, collaborative environments for iterative low-fidelity prototyping could also be facilitated. As an outcome of such simple setups, many ideas could be implemented and tested in short iteration cycles and finally be fed into off-the-shelf CAD tools for final product design. In order to test this idea, we implemented a tangible tabletop collaborative prototyping tool for residential area design, and conducted a case study with a group of architecture students and a professor. Our observations show that synchronous collaboration for rapid prototyping could be achieved with simple yet functional setups. In order to increase the reproducibility of our study, we share the setup details and code of our collaborative prototyping framework as an open source project on which various other prototyping environments (e.g., urban design) could be implemented. And finally, we provide a list of design recommendations for such systems that we derive from our case study observations and existing literature.

We argue that despite having many successful examples, current state of collaborative design/prototyping research seem to lack a critical property, namely reproducibility, without which the progress of the advancement becomes limited to individual efforts. We state that this lack of reproducibility mainly comes from the full-fledged development of tabletop CAD systems. Such systems also come with the complexity such that rapid prototyping process is typically replaced with individual model development efforts. There seems to be a gap in the literature of the collaborative design systems for rapid prototyping where the aim is the early and quick test of many ideas whilst supporting concurrent design activity and high bandwidth of communication channel. We emphasize that our approach tries to fill in this gap rather than to outperform existing approaches. Furthermore, we leave a thorough comparison of InitialInsights with relevant existing tools and approaches as the future work.

During the development of InitialInsights and our case study, we noted a number of key factors that have an impact on the design of tabletop collaborative design systems such as communication, synchronization, and provenance. We claim that such capabilities could still be achieved while maintaining simplicity of the development of the prototyping tool. With our open source project, we aim to incrementally add them to our tool kit, and thus, support the reproducibility of collaborative tabletop design research. The scope of the InitialInsights project can be seen as a platform on which tabletop tangible interactive rapid prototyping applications (e.g., residential area design in our case) can be built without requiring much expertise on the confounding fields. The success of the orchestration of these characterizations in the real-world applications will completely rely on the improvements in the collaborative design research.

3.2 Study-2: Incorporating Tabletop Visual Analytics into the Decision-making Process: A Case Study of Retail Banking

The large-scale use of office tools and statistical analysis applications indicates that they have sufficed well for some of the everyday tasks in our work cycles such as analysis, presentation, reporting, and decision-making. Nevertheless, they were designed in an era when business data was not "big" and complex enough. The ever growing avalanche of the data that we collect for our businesses compels us to find new means of understanding, sharing, and reporting the underlying ideas, and of making decisions for the future. We implemented a tabletop hybrid visual analytics system comprised of projection and physical visualization with the aim of supporting these tasks better by making the data physically available. We conducted a case study with an analysis team of a nationwide bank and completed a series of observations and interviews during their data analysis and decision-making sessions. Our study revealed that the hybrid visual analytics system approach promotes idea sharing and the contribution of all members of the group during the presentation sessions. This approach also seems to transform the common one-way structure of the communication in the presentation sessions into an alternative structure that encourages everyone to take the floor. Research questions that will be investigated in future work are also discussed.

3.2.1 Hybrid Visual Analytics System

We propose a hybrid visual analytics system with the purpose of supporting analysis and decision-making processes. Our current setup involves physical visualization, projection, and a set of preliminary interactions (Figure 3.7).

3.2.1.1 Physical Visualization

We have manufactured a 3-dimensional map of Istanbul on a CNC machine with rigid plastics. Specifically, we have physically visualized the southern central part of Istanbul, where the population density and thus the data density are high. The size of the physical map was 108 cm long, and 143 cm wide. The map uses a scale

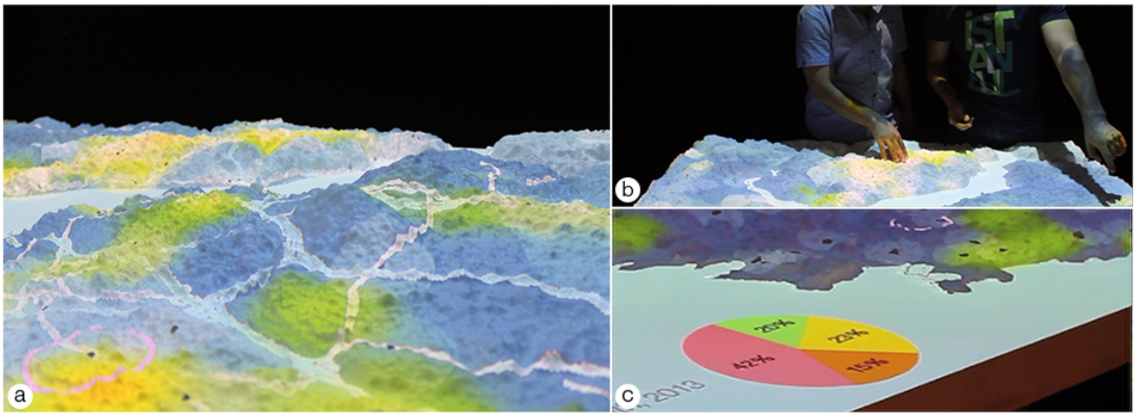

Figure 3.6: A hybrid visual analytics system (a) has been implemented by employing physical visualization and projection of data visualization. This approach seems to (b) facilitate a collaborative data analysis space. The projected visualization can be augmented with (c) various visual elements that are bound to data.

of 1:11.244. Altitude and geographical data has been acquired from ASTER GDEM V2 [74] and OpenStreetMap [75], respectively. The altitude data in the physical visualization was interpolated so that the outcome of the CNC manufacturing was rather a representation of altitude formations than a high resolution plan of a city. Large landmarks such as Bosphorus bridges and stadiums can be identified whereas the smaller structures such as buildings cannot be distinguished.

3.2.1.2 Projection

A high resolution projector is positioned over the physical visualization so that it faces the physical visualization orthogonally as shown in the Figure 3.7. The calibration and alignment of the projector with the physical visualization has been achieved so that virtual elements (e.g. shore lines, roads) are projected exactly on their physical counterparts. The projection of spatial data visualization on physical medium forms the basic version of our proposed spatial display. The projection does not only realize the visualization of data onto the physical visualization, but also it serves as a base on which augmented reality features and various interaction methods can be applied.

3.2.1.3 Interaction

In spite of the limitations posed by the stasis of the physical visualization, spatial display of our visual analytics system provides opportunities for various kinds of interactions. In our current implementation we adopted Kinect [76] in order to implement basic actions such as moving the cursor, resizing visual elements, feature filtering, and selection. Swiping hands horizontally to the right or left side causes the cursor to hop from a visual element to another. Vertical upward or downward

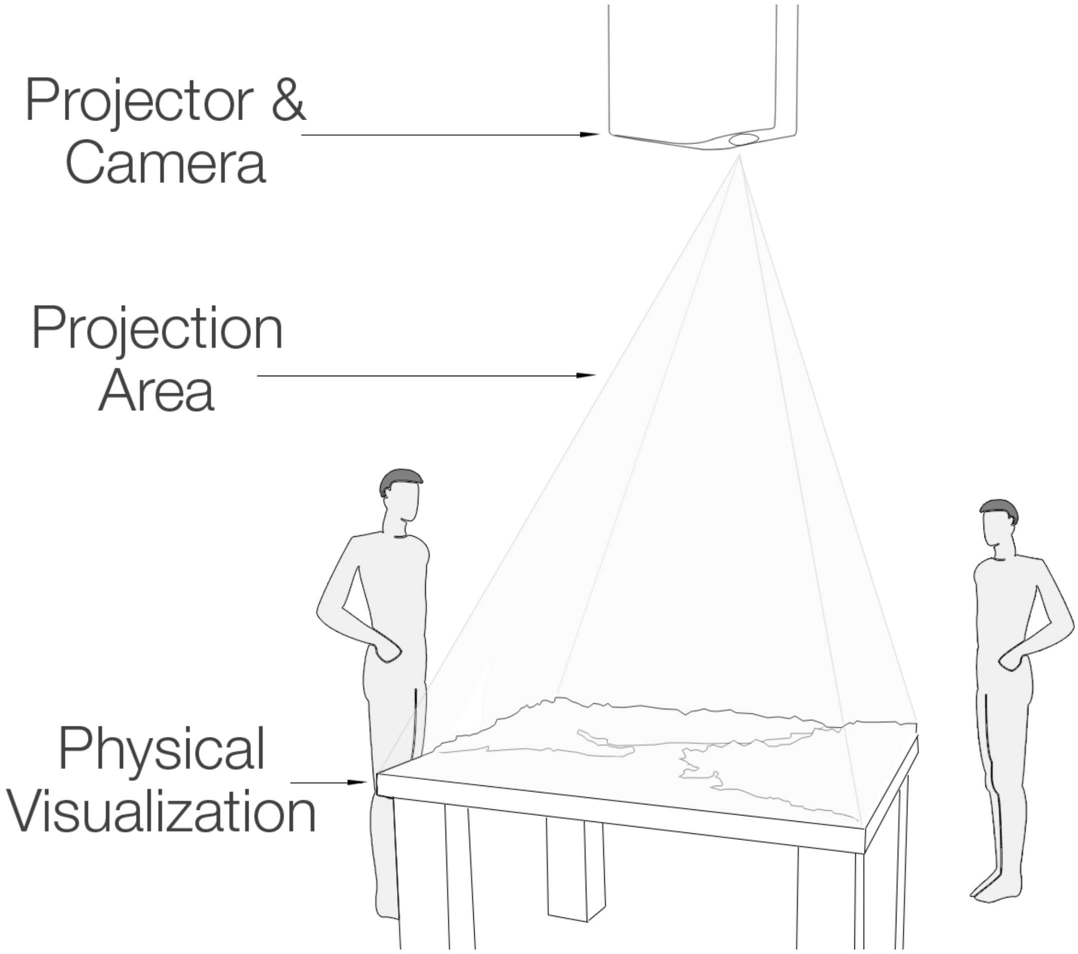

Figure 3.7: We designed a tabletop hybrid visual analytics system comprising physical visualizations and projection with the aim of supporting the collaborative analysis of spatial multivariate data and decision-making processes.

hand movement increases or decreases the value of a visualization parameter (e.g. size of circles), respectively.

3.2.2 Case Study: Retail Banking

Results of our prior informal observations lead us to conduct a case study in order to identify problematic areas and formalize research questions.

3.2.2.1 Purpose of the Case Study

As articulated in [77], the description of a new phenomenon can be accomplished through ethnographic methodologies such as observations, surveys, interviews, and focus groups. Investigation of new concepts and ideas usually starts with incomplete information; however, case studies can provide early feedback and insights about the solutions, and they can help the formalization of research questions [78]. In particular, informal case study approaches can be adopted to decide whether "a new idea is worth pursuing" or to acquire an initial understanding of a new problem domain. Informal case studies can relax the necessity of a theoretical background and some of the guidelines and procedures. Similarly, in order to better understand the use of tabletop visual analytics combining physical visualization and projection with the purpose of multivariate data analysis and decision-making, we conducted an informal case study. Our goals for the study were (a) to gain deeper insight on the effects of our novel concept on the decision-making processes, (b) to infer future design implications, and (c) to formulate research questions as the starting points for further research on the hybrid visual analytics approach. We did not intend to make rigorous comparisons of the hybrid approach and the previous systems and to make general claims.

3.2.2.2 Case Study Conduct

We conducted a case study with the participation of a group of analysts of Akbank, a nationwide bank of Turkey, to explore whether and how our novel visual analytics approach enhanced analysis and decision-making processes. For the purposes of the study, we prepared our visual analytics system according to the needs of the bank's analysis team. Preparations involved the implementation of visualizations such as catchment areas and density maps, and interaction. Upon training, the bank's analysis team conducted a number of catchment area analyses with our hybrid visual analytics system, during which we assisted with the interaction and usage of the system. As they visualized and investigated various summaries of their data, we observed their discussions about the inferences they extracted from the data. We also videotaped their brainstorming sessions and how they communicated their findings and decisions to their managers.

3.2.2.3 Catchment Area Analysis

The domain problem investigated in this case study is the catchment area analysis of bank branches. Typically, banks have a number of branches located in a region such that they collectively cover the geography and the customer demand within. Whether or not the customer catchment areas of these branches have the most *correct* coverage, the locations of new/moving branches, and optimal distribution of workforce across the catchment areas are common problems for retail banking. The catchment area of a bank branch refers to the area around each bank branch that represents an effective customer catchment radius of the branch, and the goal is to find out how well these areas align with where the customers of the bank usually transact. Hence, we resort to using our visual analytics system for an assessment;

the more the two catchment areas overlap, the closer the bank is to its customers and thus has greater opportunity to do business with them.

3.2.2.4 Analyzed Data

The data visualized for our case study comprises an anonymized transactional database of credit card purchases made in Istanbul, and the locations of the bank branches. The dataset does not include the transactions of online purchases. In both datasets, locations are given in latitude and longitude form. Each row in the transaction dataset includes anonymized *identification* of the customer making the purchase, the *timestamp* of the purchase with date and time information, the *location* of the purchase, and the *value of the transaction*. The bank branch dataset includes the *street address* and the *location* of the branch. There exists no relationship between the transaction dataset and the bank branch locations meaning that none of the transactions by no means can be linked to a bank branch. Transaction data spans over three months with more than 400.000 anonymous account owners, and it contains more than 4 million transactions in total. As our extent was limited within the central part of Istanbul, the branch data contained only over 140 branches.

3.2.2.5 Visualization and Interaction

In order to assist the catchment area analysis, we have visualized transaction pattern of bank customers by employing a kernel density map representation along with the circular bank catchment areas. The first layer of the visualization, namely the density map, was generated by using the transaction data. Two variants of the density map have been generated based on the monetary value and density of the transactions. The density has been encoded to color spectrum starting from green, transitioning into yellow, and then ending at red. Tones of green indicate lower density of transaction or lower average transaction value. And in contrast, tones of red indicate the opposite situation. If there exists no transaction for a given time period and region, only the background color is shown.

The second layer holds the catchment areas that were drawn as white circular areas centered at the bank branch locations which are marked as black dots (Please see Figure 3.6c and 3.8). The opacity of the circular area represents the density of the credit card purchases falling into the catchment area of the corresponding branch. As the radii of the catchment areas change, the opacity of the circular areas are updated depending on the transactions caught.

The south-west corner of the visualization covers a portion of the Marmara Sea where physical visualization has a flat surface since the altitude is zero. By utilizing this opportunity, we placed short data summaries of the selected branches on this area. As shown in the right side of Figure 3.8, the catchment areas are selected with a circular magenta cursor. Moreover, the visualization supports animation along the time dimension so that the analysts could track transaction pattern changes over time. The borders of the districts, highways, and arterial roads are also shown to aid the analysis endeavour.

The analyst team members could move the cursors from one branch to another by waving their hands horizontally from left to right or in the opposite direction. Branch data summary is updated whenever the cursor is moved. If the size of the catchment areas is increased or decreased, the size of the cursor is also updated accordingly. The radii of catchment area could be altered interactively by raising hands up and down. Users could also toggle some other visual features such as roads, district borders, and branch marks.

3.2.3 Related Work and Discussion

Accepted as a subfield of computer-supported cooperative work (CSCW) [79], collaborative visualization, like all other CSCW scenarios, can be investigated from space (co-located vs. distributed) and time (synchronous vs. asynchronous) aspects [7], [80]. Moreover, Isenberg et al. [79] categorizes collaborative visualizations based on different levels of engagement with the visualizations, namely viewing (e.g. PowerPoint), interacting and exploring (e.g. CPOF [81]), and sharing and creating (e.g. ManyEyes [82] and Tableau Public [83]). As our co-located visual analytics system physically supports different views of visualization and enables social interaction space through which findings can be exchanged, it could be considered as a sharing and creating system.

Many visualization and analysis systems have been designed and developed for collaborative information processing activities. Lissermann et al. introduce the concept of Permulin [84], which comprises a large touch surface and shutter glasses providing private and shared space for the users. The use of a head-mounted device limits natural face-to-face communication, which seems not to be the case in our system. Another example of two-dimensional tabletop approaches, OrMiS [85] supports collaborative analysis, planning, and interaction for simulation-based training purposes. These examples of tabletop visual analytics approaches are similar to our system in that they facilitate a base for collaborative analysis and idea sharing. In order to extend these valuable contributions, we incorporate the physical aspect to the visualization as evidence demonstrates that the physical representations of the data can augment perception. In their experimental work, Jansen and Dragicevic [86] report that physical visualizations are superior to their virtual counterparts in information retrieval tasks. The ability to touch the data and perfect visual realism of physicalism facilitates the successful extraction of inferences from the data.

The benefits of data physicalism were also utilized in CityScape [87], and Relief [62]. An obvious advantage of these tabletop approaches is the active physical visualization that can be adapted to various parameters of the visualization. However, the dynamism of the physical visualization leads to a low resolution in the spatial aspect of the data due to technical constraints. In contrast, the continuous (thus high resolution) form of static physical visualizations seems to outperform from this perspective.

Static physical visualizations have numerous limitations due to their passive nature. One of the obvious inherent constraints is the lack of support of very basic actions such as zoom in and out. Compared to the improvements in the two-dimensional tabletop visualizations such as the zoom lense enhancement on OrMiS [38], the stasis of physical visualizations may seem as an irresolvable problem. However, we argue that, with the support of peripherals such as transparent displays or mobile devices equipped with cameras, the techniques available in the augmented and virtual reality fields can aid the alleviation of the problem. For example, we can zoom into the regions of interest in separate views that can be generated in the scopes of transparent displays. Alternatively, the details of data can be viewed through the cameras of wearable computing devices or of mobile phones [88].

In the case study, we focused on three retail banking processes; the innovative analysis of transaction data, idea flow and brainstorming between the members of the analysis groups, and the communication of findings to decision-makers. During our preliminary investigations, the bank personnel reported that these processes were conducted via commercial analysis tools and reported by conventional tools such as presentations and spreadsheets. Not surprisingly, some of the personnel reported that the summaries or charts of vast amount of transactional data lead to missing details as well as misunderstandings. Moreover, the dashboards and branchwise representation of information degrades the recollection of previously discussed issues. As the presenter highlights the important aspects of a particular topic, previously presented information remains out of sight, and leads to a discontinuity in the big picture.

The results of our preliminary interviews show that analysts and decision-makers seemed to have issues with their conventional analysis and information representation methodologies. The reason was likely due to the bandwidth of their communication channel where they shared their ideas and information. Based on our preliminary observations, we formed our research questions as follows:

RQ1 Are the analysis, idea sharing and decision-making processes involving conventional methods (such as spreadsheets, presentation, statistical analysis tools) and interactions (such as 2D displays, mouse, and keyboard) sufficient to effectively discuss extensive and complex scenarios?

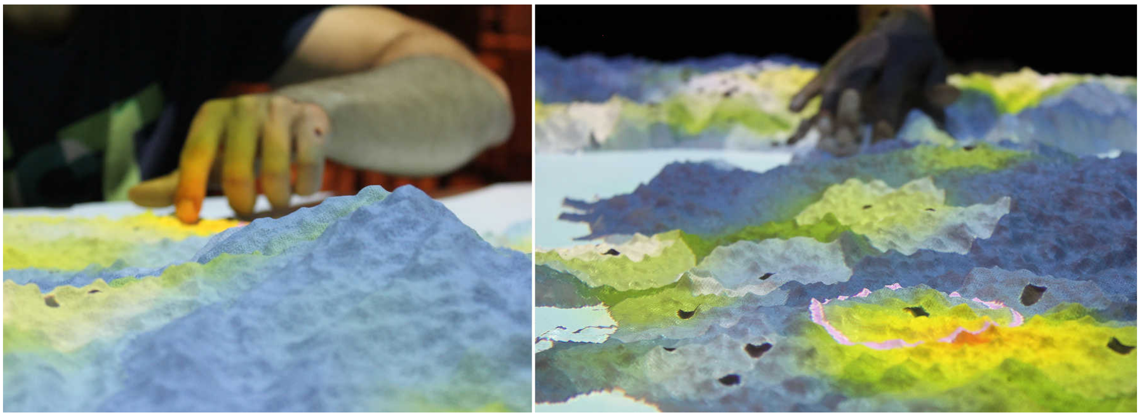

Figure 3.8: Physically availability of the data augments the realism of the subject matter attracting attention and promoting conversation.

RQ2 From which perspectives do new intra-group interactions, conversational hierarchy, and idea flow emanating from hybrid visual analytics systems differ from those conventional ones?

RQ3 What are the feasible means for telepresence during analysis and idea sharing processes other than rooms with large displays supporting multimodal communication?

During both informal observations and the case study, we observed that the users of the tabletop visual analytics system tend to communicate their ideas to or start a discussion with other users around the table. More surprisingly, this usually occurs without any intervention. One analyst participating in our study reported that he found the visualization and the capability of touching the data aesthetically appealing and, thus, it promoted conversation and idea sharing among the analysts, which led to new questions and discussions. Another analyst supported this reasoning by stating that the physical existence of visualization and the facilitation of all team members in a natural social setting encouraged social interactions. To this respect, we state that visual analytics system involving physical visualizations may significantly enhance the decision-making process since the increase of idea flow and conversation positively affect the creativity and productivity of the group as articulated in [89].

Statistical analysis applications, office applications, reporting software, and presentation tools have been serving efficiently for the support of business processes for more than three decades. Summarization of sales, charts of productivity, and various reports specialized according to the needs of the companies have been effectively used to spread information and explain ideas to other employees. However, the data that has to be dealt with has accumulated so much that it requires wider channels and new means of conveying valuable inferences. Big data is not only big, it is also complicated in the sense that new analysis methods need to be developed to make better sense of it. Statistical methods seem to fall short in dealing with the high dimensionality of the data. The exploratory analysis of the data might benefit from humans sophisticated visual skills to make the judgements. That said, we argue that recent improvements in human-computer interaction and the visualization worlds can be employed to build visual analytics systems that enhance the capabilities of our current tools. Can we use tabletop systems, large displays or physical visualizations accompanied with state-of-the-art interaction techniques to create hybrid visual analytics systems that enhance collaborative analysis? Would that improve our retention rate at the end of analysis sessions if we were able to interact with the data with our senses such as sight and touch? After our observations, we had enough evidence demonstrating that research for answering these questions might reveal valuable insights.

The rate of the idea flow in a group of people directly indicates decision-making and strategy refining processes [89]. Flow of ideas seems to be critical in any group of people in a sense that the creativity and productivity depends on this social phenomenon. One of the data analysts participated to our study reported that they use presentation or data visualization provided by their analysis tools during decision-making processes. However, he also added that the physical visualization made them communicate more compared to when they use other conventional techniques. Moreover, the physical visualization not only improved the occurrences of conversations but also helped to form a conversation network in which all participants took part. We postulate that the physical availability of data enabled analysts to view the data from different perspectives and led to diverse ideas. Besides several factors such as the susceptibility of individuals to new ideas, the contribution of group members to the communication is essential. From this perspective, our hybrid visual analytics system seems to have an important role in the conduct of idea flow.

Whether hybrid visual analytics systems can also affect the structure of the conversation during the decision-making processes, particularly during the presentations of findings to the relevant audience, remains as an interesting question. For example, in a common setting where a presentation is made over a common presentation application, an audience tends not to interrupt the presentation by keeping their feedbacks or criticism to a predefined time slot. This situation seems to form a one-way communication structure where the presenter has an implied control of the setting. On the other hand, we have observed a slightly different situation where communication during all phases of this study was two-way. This observation implies that physically available data and its visualization might change the habitual behaviors that we adopt during presentations and briefings, which we will explore in one of our future studies.

Telepresence plays an important role in remote collaboration and decision-making as the companies are getting increasingly global. Being one of the blessings of the digital age, remote collaboration via video conference and large displays are current telepresence systems widely employed by many organizations; however, face-to-face communication, in which individuals are able to make more use of voice prosody, facial expressions, gaze, hand and bodily gestures to convey the intended messages, remains as a "golden standard" [90] of communicative effectiveness. From this point of view, making all aspects (users, data, inferences) of remote analysis and collaboration activities as physically available as possible is an important step towards the golden standard of communication. The use of techniques available in the spatial augmented reality with physical visualizations can lead to design of tools that substantially enhance remote collaborative analysis. For example, we can illustrate visual elements such as bar or pie charts over a landmark on physical visualization in the virtual space in which we can browse through the lenses of our computing devices. Then we can obviously send our inferences to our colleagues so that they can view them via their own devices supporting browsing in the virtual space. This kind of communication primitives can help us build more sophisticated and engaging telepresence systems that are more realistic compared to their conventional counterparts.

3.3 Observation Data

The evaluation of both studies has been conducted with case studies whose participants were domain experts. Except for the training sessions, interactions of experts with the tool has been video recorded, and notes have been taken by the experimenters. The analyses conducted by the experts were partially guided by the experimenters in order to cover as many aspects of interactions as possible. At the end of the experiments, participants were asked to report their subjective evaluation of the analytics tool and their interaction experience. Moreover, both of the analytic systems were configured in a way that the data-side messages and part of the human-side messages were logged. Based on these data sources, objective, behavioral, and subjective data for realm characterization has been extracted as explained in the following subsections.

3.3.1 Objective Data

Prior to calculation of analytic system properties explained in Section 2.5.2, objective data has been extracted from the case study sessions whose methodologies described in Sections 3.1.4 and 3.2.2. Those case study sessions mainly comprise users' interaction with the analytic system for considerable amount of time. Except for the initial training times, the duration of the sessions has been enough to observe the correct user message patterns. In order to prevent any personal bias on the data, no time limitation has been applied in neither of the sessions.

Analytic systems developed for the studies had the capability to record the user interactions as well as their time stamps. Furthermore, the system also recorded its own reactions to user inputs. By careful categorization of user inputs and system reactions, I characterized the human and data side messages that would be used for the calculation of system properties.

Nevertheless, such log data was often not clean and frequently incomplete. That was mainly due to the quick decision change of the users where users sent different messages without waiting for the response of a prior one. In order to complete the data, video recordings of the analysis sessions have been employed. Contiguous user-side message occlusions have been dissolved by making cross checks with the user input actions.

The result of the tedious log data compilation was the objective data set whose attributes are the characteristics of interaction listed in Table 3.1. For each of the analytic system the data source for each of the characteristics is shown in the table. Human-side message timestamps was determined based on the cleaning and preprocessing of the analytic system logs accompanied by the video record of the user interactions. The message types was attributed by the experts by making use of the system logs. All possible media over which the human- and data-side messages are carried and their physicality constants was identified by the experts. Data-side message timestamps were simply filtered from the analytic system log. After a careful investigation data-side messages, their types were determined by applying heuristics. Data consumption rates have been automatically recorded by the system.

Table 3.1: Study-wise data sources for interaction characteristic data calculation. Objective data sources for each analytic system study is listed. *Log* corresponds to the interaction logs on the data side, *video* is the records of the sessions, *expert review* is the data extraction process carried out by analytic system's expert, and *heuristics* is a predefined procedure for the calculation of the corresponding data set.

Characteristic	System 1	System 2
Human-side Message Timestamps	Log + Video	Log + Video
Human-side Message Types	Log + Expert Review	Log + Expert Review
Media	Expert Review	Expert Review
Medium Physicality Constants	Expert Review	Expert Review
Data-side Message Timestamps	Log	Log
Data-side Message Types	Log + Heuristics	Log + Heuristics
Data Consumption Rates	Log	Log

The calculated analytic system properties are shown in Table 3.2.

Table 3.2: Analytic system property values for systems 1 and 2. Property values for the analytic systems are calculated based on the interaction characteristics data. These values form the objective part of the observation data.

Property	System 1	System 2
Responsiveness	0.47	0.54
Communication Media Level	0.39	0.44
Unit Task Diversity	0.25	0.69
Human-side Message Closeness Factor	2.78	13.21
Progressiveness Level	0	0

3.3.2 Behavioral Data

Behavioral data comprises user actions that can not be considered as part of the communication between human and data. This definition excludes the user inputs such as zoom in or out, filtering, comparison, and data retrieval. In that sense, behavioral data becomes more important in the multiuser systems where users establish additional communication channels between each other.

Behavioral data transcribed during the analysis sessions are as follows:

- *Directed technical communication.* Questions or statements made by the users of the systems that are not directly related to the analytic task at hand. These behaviors involve analytic system interaction requests of peers such as context or mode change, object detection.

- *Directed task-related communication.* Questions or statements made by the users of the systems that are directly related to the analytic task. This communication pattern involves the statements that could be attributed as task completion efforts. For example, questions asked to other peers such as "Do you think young customers spend more during weekend in region X?" can be considered in this behavior category.

- *Hypothesis, insight, and question statements.* Any form of statement that could be qualitatively judged as a hypothesis, insight statement or research question are considered in this category.

- *Idling communication.* Any form of communication that is not made for technical or task-related reasons are considered in this category.

- *Bodily gesture towards peers.* Bodily gestures that are directed toward the analytic system's interface or peers are considered in this category.

3.3.3 Subjective Data

Subjective data comprises self-reports of the participants of the data analytic studies. This form of data is used for the characterization verification of the realms, and is listed as follows.

- *Pre-study self-reports.* Users of the analytic systems were expected to describe their analysis strategies for the given tasks prior to the case study sessions. This subjective data was employed in order to measure the extent that the analytic system affords itself as for its capabilities.

- *Post-study self-reports.* Post-study scales and semi-structured interviews was conducted in order to measure the task-tool match quality from the users' perspective. Users were asked to express their own considerations on how successful they felt in completing their tasks, and collaborating with their peers.

- *Assisted self-reports.* Moments of interests have been cropped from the recorded video, and were shown to the users in order to get deeper explanation on the root cause of their behavior or interactions with the tool.

3.4 Realm Characterization

As can be seen in Figure 3.9, two studies reported in this chapter are projected onto the middle region of the exploration space. The fact that these analytic systems have moderate values in each of the responsiveness and physicality axes is inline with the specifications of the analytic tasks: Both tasks require high collaboration with highly interactive system whilst supporting a tangible interface. The interactivity, however, is only required at the lowest level supporting continuous communication between human and the system as inline with the second time constant of Card et al. [15]. Card et al. suggest that, in human-human or human-tool interaction, the immediate response time should be less than one second in order to have a healthy communication among the agents. Inline with this constant, the median value for the responsiveness axis is fixed as one second, which is translated as 0.5 value.

Grounded theory methodology has been applied on the observation data acquired during the conduct of the analytic system studies in order to develop plenary statements about the hybrid systems realm. Results comprise statements from wide range of aspects of the analytic systems. Wherever possible, the statements are verified with respect to objective data, scale or interview data, or insight-based evaluation data.

One of the most prominent properties of the analytic systems residing in this realm seems to be that they have better support for the communication as evidenced

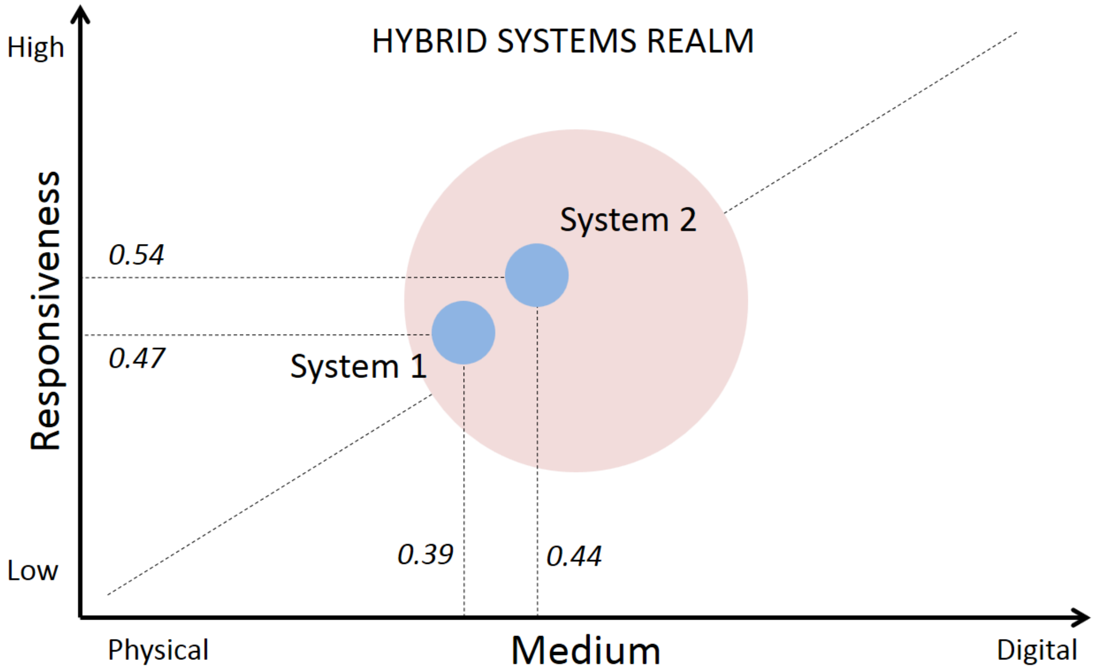

Figure 3.9: Hybrid systems realm. Plot of system 1 and 2 on the exploration space of HDI is represented.

by our observation during the case studies. As an explanation for this case, the tabletop-like shape of the analytic systems could be asserted, however, we found out that the partial physicality of the system also promotes the communication. The physicality adopted in the systems such as cubes representing building blocks or three-dimensional geographical space acts as the catalyst for the communication. This effect can be noticed more saliently in system 1 where movable physical parts are used which increases the physicality indicator. In system 1, participants seem to communicate more compared to those did with the system 2.

As North et al. suggest [30], analytic systems' main role should be to convey insights from a given data set. He argues that the comparisons of the data analytic systems should be made based on the number and significance of the insights and hypothesis extracted. Being highly cognitive tasks, extracting such information from a given data set necessitates high level of communication distributed with a balance among the group members [89]. In this regard, hybrid systems realm seems to support the analytic tasks requiring brainstorming and idea generation. During the interviews, the domain experts stated that they conducted the analysis sessions in a more collaborative atmosphere, and they stated they were more productive with these analytic systems in terms of exploratory data analysis.

Idling conversations occurred a few times during both of the case studies. When closely investigated, I have seen that these short breaks were taken by the participants mainly in case of technical issue moments or when the systems' response time

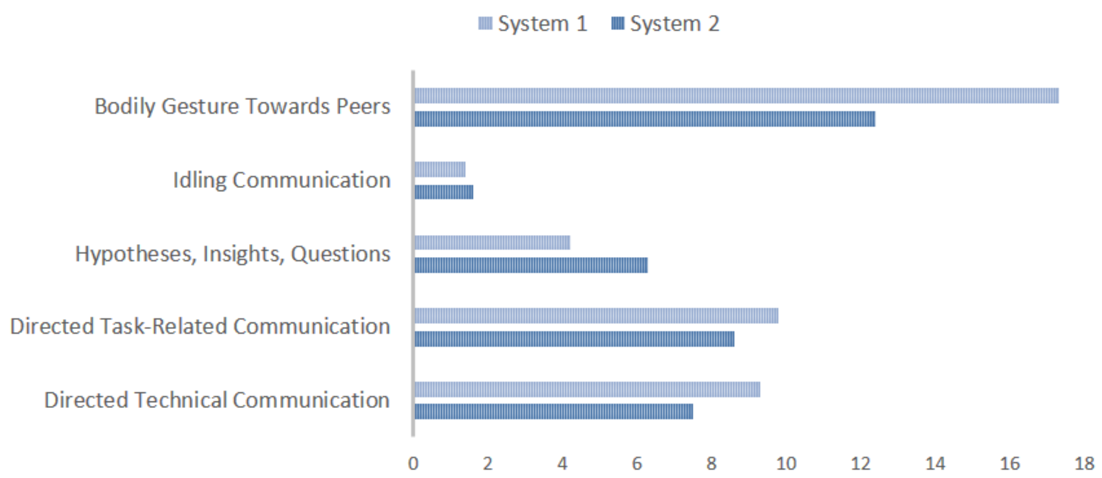

Figure 3.10: Behavioral data summary. Behavioral data collected during the case studies of analytic system 1 and 2. The values represent the mean counts per hour.

was higher than their normal. In such cases, participants usually lose focus and sometimes momentarily change the flow of the analysis. In some cases, they did not recall their last insight right before the idling conversation, and started a new analysis session from scratch. Engagement to analytical task seems to be strongly tied to the responsiveness level of the analytic system, and response times longer then 4-5 seconds seem to be distracting the users of the tool. Except for the idling conversations, based on the behavioral and subjective data, I state that the case study analysis sessions facilitated highly continuous communication environment mainly due to the average response times of the systems slightly less than one second.

It should be noted that high responsiveness does not necessarily mean high interactivity. The former is related with the response time of data-side messages whilst the latter corresponds to high number of human-side messages per unit time and responded in considerably short amount of time. For example, a task might necessitate issuance of human-side messages once in 10 minutes, and short response time. And another task might require very intense human-side message issuance, say once in 30 seconds, and short response time. Both tasks are said to require highly responsive systems, however, not highly interactive systems. Similar case was that of the Study-1 and 2 both of which had similar responsiveness value while they had quite different human-side message closeness factor as can be seen in Table 3.2. Human-side message closeness factor is a measure for the system's average contiguous human-side message frequency. Closer the human-side messages are, lower the value is. In this regard, hybrid systems realm could be the set of analytic systems designed for responsive tasks, however, not necessarily for highly interactive systems.

CHAPTER 4

RESPONSIVE DIGITAL SYSTEMS REALM

Responsive digital systems realm is the region in the exploration space of HDI where the typical desktop or web-based visual analytic systems running on relatively small data sets with little or no physical parts reside. However, it should be noted that these analytic systems are not expected to process batch analytic tasks on big data that could dramatically decrease responsiveness. Contrarily, they are highly responsive systems facilitating communication between human and data mainly on digital media. Such systems typically lend themselves as highly interactive exploratory data analysis tools. They may or may not support visual analytic operations, and they could also perform progressive analytics on large and/or stream datasets. As long as the system is highly responsive and communicates on the digital media, such systems are considered as belonging to the responsive digital systems realm, i.e., located on the upper right corner of the exploration space.

Due to the highly digital nature of such analytic systems, they tend to be analytic systems mainly for individual data exploration. The collaboration aspect seems not to be as strong as in the hybrid systems realm. Even though the collaboration could be established with multiuser and highly digital analytic systems, as evidenced by the behavioral data explained in Section 4.3, they lead rather brief discussion for the on-the-fly results than long brainstorming sessions as in hybrid systems realm. I have partial evidence implying that the main cause for this is that the lack of physical part of the analytic systems in this realm.

The studies from which the observation data was extracted for the identification of this realm are reported in Sections 4.1 and 4.2. Having different validation methodologies, the observation data extracted from each of these studies varies, and explained in Section 4.3. In Section 4.4, the characterization of the responsive digital systems realm is done by applying grounded theory to the available observation data.

4.1 Study-3: An Empirical Discussion on 2D and 3D Representations of the Spatio-temporal Data

Evaluation of information visualization techniques and visual analytic systems with controlled laboratory experiments is commonplace when the usability and technical capabilities are of concern. In order to gain experience in the experimental evaluation of information visualization systems, I and my colleagues conducted a research with the aim of comparing 2D and 3D representations of spatio-temporal data entitled as "Do 3D Visualizations Fail? An Empirical Discussion on 2D and 3D Representations of the Spatio-temporal Data."

Having particular aspects diverging from other kinds of data, spatio-temporal data may require conventional visualization techniques and tools to be modified and tailored for analysis purposes. However, there could be a number of pitfalls for the design of such analysis tools that completely relies on the well-known techniques with well-known limitations possibly due to the intrinsic characteristics of the data at hand. In this study, we present an experimental study to empirically testify whether advantages and limitations of 2D and 3D representations are valid for the spatio-temporal data visualization. To form the basis for our experimental design, we interviewed domain experts from a location-based services company and identified different cases of spatio-temporal data analysis. Furthermore, we implemented two simple representations, namely density map and density cube, and conducted a laboratory experiment to compare aforementioned techniques from time and correctness perspectives.

Results of our experiment revealed that care must be taken while applying highly accepted well-understood properties of 2D and 3D visualizations to the spatio-temporal data. On one hand, none of the techniques was superior to the other in terms of task completion time except for some particular cases. On the other hand, our 2D density map implementation led to more accurate inferences for tasks related to trend detection and making comparisons, as vastly suggested by the previous work. As suggested by these examples, the validity of the generally accepted aspects of 2D and 3D visualization needs to be reconsidered for the analysis of spatio-temporal data, hence the purpose of our study.

4.1.1 Previous Work

4.1.1.1 Properties of 2D and 3D Representations

The choice of 2D or 3D representations in information visualization seems to be a highly debatable and complex task. Furthermore, previous body of work suggests that well-understood commonly accepted advantages and drawbacks of one representation over the other may depend on the context and data analysis objectives. For example, as Munzner states [91], selecting 3D becomes meaningful when the 3D representation is implicit in the dataset and when that representation fits to the mental model of the user about the phenomenon.

It would be a subject of a different work to enumerate all the studies comparing the 2D and 3D representations. Common known properties of these representations will briefly be explained with a few example studies.

The effectiveness of 2D representations in terms of both effectiveness (e.g. time to complete task) and accuracy (e.g. error rate) have also been reported in a number of studies. For example, users were able to locate regions of interest more effectively and accurately when they view the blood flow visualization with 2D representations [92]. In their experimental study, Hicks et al. [93] compared various performance levels of 2D and 3D visualizations of e-mail usage log, which they consider as temporal data. They reported that 3D representations poorly performed in terms of task completion time whereas the very same representation aided their participants in answering the comparison questions more accurately where the participants had to view the whole data. Their findings differ from ours possibly due to the natural differences between temporal and spatio-temporal data, which we discuss in Section 5.

Another property of the 2D visualizations is related to “spatial memory” as suggested by previous work. In their study [94], Cockburn et al. suggest that their subjects’ ability to quickly locate web page images deteriorated as their freedom to use the third dimension increased. Case studies and formal user studies demonstrated that 2D data encodings and representations are generally more effective than 3D for tasks involving spatial memory, spatial identification, and precision.

The advantage of the 3D representations becomes clear with the tasks requiring the holistic view of the data [31], [93], [95]. As Hicks et al. state, cognitive error degrades when the users are to view the data as a whole. The experimental findings

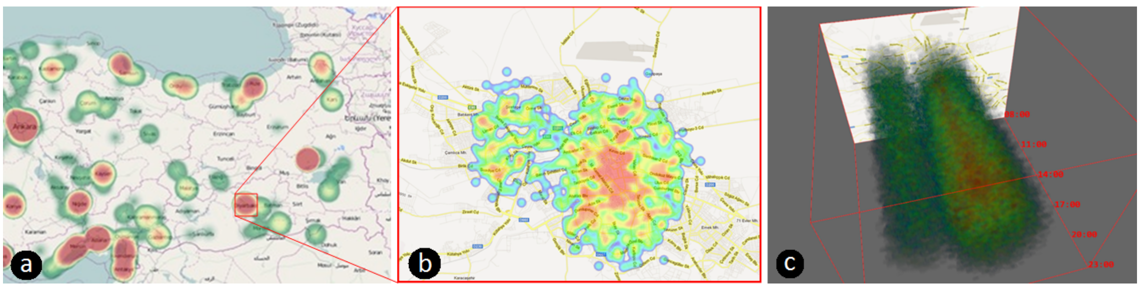

Figure 4.1: Views from our experimental tool showing the 2D and 3D visualization of the spatio-temporal data from a commercial friend finder application. (a) 2D Density Map, (b) Zoomed Density Map, and (c) 3D Density Cube. Note that the third dimension for the Density Cube is time.

of Kjellin et al. also supports this idea as their participants were able to better track multiple vehicles when they were equipped with 3D visualization techniques [96]. Nevertheless, the holistic view might bring its own issues such as occlusion, one of the common problems associated with 3D representations [93]. Perspective foreshortening is another problem with the 3D representations. Users might suffer from justifying the measures of the objects in the 3D visualizations due to the perspective projection while performing trend comparison tasks.

4.1.1.2 Spatio-temporal Data

In their extensive work [97]–[99], MacEachren and Kraak explore different techniques and the research challenges in geovisualization field. They address the important points of geovisualization such as representation of geospatial information, integration of visualisation with computational methods, interface design for geovisualization environments and cognitive/usability aspects of geovisualization. Different cartographic techniques have been used to represent geospatial information. Among many other visualizations that use geographic maps, thematic mapping [100] techniques are designed to show a particular theme connected with a specific geographic area. Density map [101] technique is also adopted for geographic data visualization and data analysis in several important examples. Fisher proposes an interactive heat map system [102] that visualizes the geographic areas with high user attention in order to understand the use of online maps. Mehler et. al also use a geographic visualization technique similar to heat map [103] in which they geographically analyze the news sources. Another interactive framework taking advantage of heat maps is introduced by Scheepens et al.[104] in which they aim to visualize the trajectory data of vessels. The density map implementation employed in our experimental study is heavily influenced by heat map visualization technique.

Perhaps one of the most important and ubiquitous data types is the one with references to both time and space, usually referred to as spatio-temporal data. The concept of spatio-temporal data is defined in both geographic information systems (GIS) [105], data mining [106], and visualization [107]. Visualization of spatio-temporal data involves the direct depiction of each record in a data set so as to allow the analyst to extract noteworthy patterns by viewing the representations on the displays and interacting with those representations. Increasing number of studies on management [108] and analysis [14], [109] of spatio-temporal data in the last decade indicate the importance of the analysis of this data type. The analysis of such data with references both in space and in time is a challenging research topic. Major research challenges include [110] *scale*, as it is often necessary to consider spatio-temporal data at different spatio-temporal scales; the *uncertainty of the data* as data are often incomplete, interpolated, collected at different times, or based upon different assumptions; the *complexity of geographical space and time*, since in addition to metric properties of space and time and topological/temporal relations between objects, it is necessary to take into account the heterogeneity of the space and structure of time; and finally the *complexity of spatial decision making processes*, since a decision process may involve actors with different roles, interests, levels of knowledge of the problem domain and the territory.

When the spatio-temporal data sets are very large and complex, existing techniques may not be effective to allow the analyst to extract important patterns. Users may also have difficulty in perceiving, tracking and comprehending numerous visual elements that change simultaneously. One way to deal with this problem would be the aggregation or summarization of data prior to graphical representation and visualization [111]–[113].

Infinitely many visualization studies have been conducted regarding spatio-temporal data visualizations. For visualizing the spatial change over time in data, Scheepens et al. propose an interactive visualization framework, which analyzes the trajectory data of vessels to understand their behavior and risks [113]. After the space-time cube method has been revisited for the analysis of geographic data in many works [99], [102], it has been used frequently in visualizing spatio-temporal data [114], [115]. The space-time cube approach bore the idea of using the third axis for representing time. 3D visualization techniques have been used on visualizing hierarchies that change over time in a geo-spatial context [116]. Several time-oriented visualization methods are also presented [117], [118] to analyze and support effective allocation of resources in a spatio-temporal context. In their analytic review, Andrienko et al. [107] discuss various visualization techniques for spatio-temporal data, with a focus on exploration. They categorize the techniques by what kind of data they can be used for and the kinds of exploration questions can be asked.

Being a very brief summary of existing literature, these works include evaluations of the visualizations to some particular extend; however, when and if their findings for the spatio-temporal data apply also to other types of data seems to be rather unheeded. Effects of the representation type may be different than expected when the data to explore is spatio-temporal, where the dimensionality inherently increases. Kjellin et al. [96] reports different cases how 2D and 3D representations outperform each other. Their 2D representations of movement plots assisted the participants to perform better while they were trying to estimate the intersection of two vehicles. Contrarily, 3D representation helped the users better when they had to estimate a similar measure for more vehicles, supporting our idea that the nature of the spatio-temporal data might effect the generally accepted drawbacks of the 3D representations.

4.1.1.3 Data Description

Real Data—The geographic data employed in our study comprises spatial event records with geographic coordinates with GPS-accuracy as well as a seconds-precision timestamp. Approximately 2.5 millions of such records have been collected by a LBS company. Each data record corresponds to when and where a mobile phone application was invoked. Dataset involves records saved between February 2nd, 2011 and April 1st, 2012 from almost all provinces of Turkey. Even though the density of the data adequately enabled us to infer possible interesting trends, we opted to use fictious data so that we could generate a number of patterns that were not existent in our real data.

Artificially Generated Data—For our experimental study, we decided employing artificial data merely to generate a wider spectrum of different scenarios that were not available in the real data.

A data generator has been designed to reflect the nature of spatio-temporal data. For a given region centered at L_C and a specific time interval T, the generator creates data records each of which corresponds to a point L in the data space with latitude $(L_{C,Lat})$, longitude $(L_{C,Long})$, and time $(L_{C,T})$ dimensions.

For a given time period T, i sub-intervals with even durations are created such that the duration of each sub interval $t = T/i$. Based on the requirements of a predefined scenario, a distribution function $f[t]$ is defined as the *planned* total number of data points per sub interval t. The *actual* total number of data points per sub interval is calculated as

$$N_t = R(f[t](1-\alpha), f[t](1+\alpha))$$

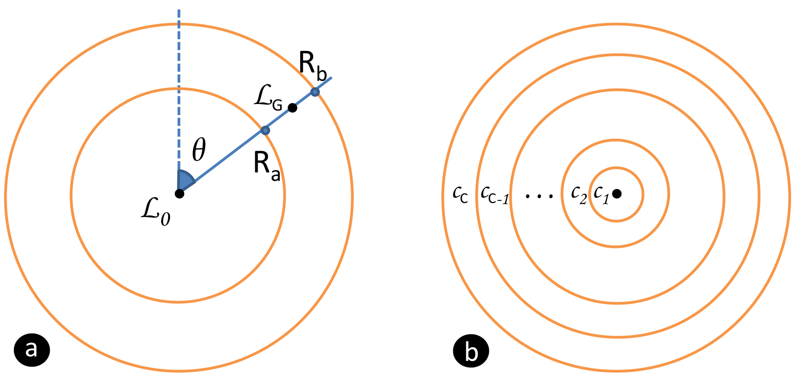

Figure 4.2: Given an origin point L_0, (a) a data point is generated with randomized θ angle and a radius selected from the interval $[R_a, R_b]$, (b) which specifies the inner and outer radii of the selected annular area c_j.

where α $(0 \leqslant \alpha \leqslant 0.5)$ is the noise parameter controlling the distortion of the number of the data points to be generated, and the function R(a,b) selects a random value between a and b.

The following equations are employed to generate data points $L_G(L_{G,Lat}, L_{G,Long}, L_{G,T})$ each of which are comprised of tuples of latitude, longitude, and timestamp.

$$L_{G,Lat} = rcos\theta + L_{0,Lat}$$

$$L_{G,Long} = rsin\theta + L_{0,Long}$$

$$L_{G,T} = R(t_m, t_n)$$

where $r = R(R_a, R_b)$, $\theta = R(0, 2\pi)$, L_0 is the origin point based on which the data generation is performed (Please see Figure 4.2(a)). In principle, the data generator is provided with the region center L_C before the data generation process. The region center L_C is slightly shifted to a randomly created point (L_0) to add noise to the data generation. The calculation of L_0 will be explained in the next subsection. t_m and t_n are the beginning and the ending times, respectively, of a given sub interval t. Being the inner and outer radii of the selected annular subarea, R_a and R_b, will be explained further in the next section.

The geographical scatter of the generated data points was based on the "Lottery Scheduling Algorithm" developed by Waldsburger et al. [119]. In the lottery scheduling algorithm, each process to be run on a CPU is assigned a number of tickets. At the beginning of each run period, scheduler picks a ticket randomly and selects the process holding this ticket. Apparently, the process holding more tickets than the other rivaling processes will have more chance to be scheduled on the CPU. Our data generator tool benefited from this algorithm for the scatter of data. Such that, the area of interest is divided into C concentric annular nested areas as in Figure 4.2(c), and each area is assigned a number of tickets ω_j, in other words, scatter parameters, associated with area c_j. The number of tickets, ω_j, defines the probability of a data point to be generated in area c_j. For a group of sub-intervals $[t_a, ..., t_b]$, the scatter of the data can be arranged by simply modifying ω_j scatter parameters. The number of scatter parameters (and also the number of nested areas), j, is specified manually depending on the requirements of the task scenario. The selected area c_j is defined by two radii, R_a and R_b, which is calculated as $R_a = \frac{D(j-1)}{C}$ and $R_b = \frac{Dj}{C}$, where D is the diameter of the circular area for which the data will be generated.

To generate data looking more real, we considered another noise factor β, namely scatter noise. Scatter noise parameter creates distortions on the scatter of the data by randomizing the latitude and longitude values of the origin point L_0 within a predefined limited area. $L_0(L_{0,Lat}, L_{0,Long}, L_{0,T})$ is calculated before the generation of each data point with following formulae,

$$L_{0,Lat} = r_r cos\theta + L_{C,Lat}$$

$$L_{0,Long} = r_r sin\theta + L_{C,Long}$$

$$L_{0,T} = 0$$

where $r_r = R(0, D/\beta)$, $\theta = R(0, 2\pi)$, L_C is the central point of region with a radius D selected by the experimenters according to the requirements of the scenarios.

4.1.1.4 Spatio-temporal Data Analysis Needs

We conducted interviews with system operators at a LBS company in order to form a practical basis for our experimental study. We found out that two spatial analysis cases were of their firm's critical importance: (a) Measuring the effectiveness of a service promotion (e.g. analyzing the effects of a promotion for a particular period of time after the advertisement) and (b) identification of local service disruptions (e.g. being able to find of 30-minute-long disruption in a week-long worth of data). It was important for them to quantify the effective breadth and duration of a promotion to understand the market dynamics. Fast and correct spatial analysis of service disruption cases was found out to be crucial. For example, a local service drop could trigger a false alarm which may lead to a legal dispute in particular cases. Even though tracing back from the usage logs of the service might be considered as a reasonable approach, the process had usually been evaluated as exhaustive and unfruitful. It was also important to locate and identify the general usage patterns, customer behavior patterns, and user profiles for the service providers, since the data specifically tailored for some special user groups might bring potential benefits.

4.1.1.5 Visualization Techniques

In order to represent 2D and 3D techniques in our experimental study, density map and density cube representations were implemented.

Density Map Animation—Density map animation has been implemented as the sequential animation of raster images. Firstly, the density map of the artificial data has been created by making use of kernel density estimation. Calculated intensity values of pixels were scaled into the range of 0-255 to produce a gray-scale intensity map image. Upon the calculation, we modeled each geographic location as a radial gradient, a filled circle having a gradually descending intensity as the distance from the center increases. To calculate the intensity of each pixel with this method, additive blending technique has been utilized where the intensity values of geographical

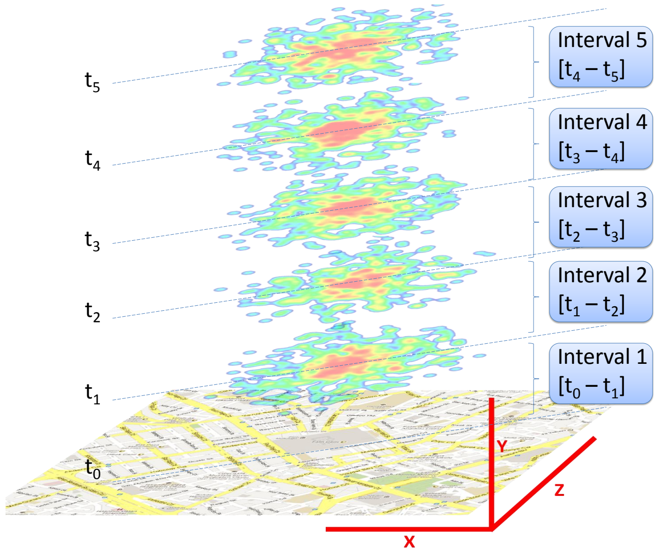

Figure 4.3: Generation of 3D texture by stacking density maps: Density maps are generated for each time interval with the data lying in the corresponding interval. Then, these density maps are stacked according to the order of the time intervals to generate the 3D texture.

points occupying the same pixel have been summed up. Consequently, at the beginning of the colorization process, gray-scale maps have been created by scaling the intensity values of each pixel between 0-255. Finally, appropriate color schemes have been chosen to be applied on the gray-scale intensity maps [104]. The still images generated with this method are used as the frames for the animation.

Density Cube—As a time-space cube variant for spatio-temporal data visualization, density cube represents the data of interest as a whole on the display so that the cognitive load on analysts is reduced by obviating the need for remembering the previous frames for tracking the changes. Density cube technique overcomes this limitation by utilizing a 3D texture, which actually is a stacked version of consecutive density maps (Figure 4.3). The 3D texture is then rendered by employing a GPU ray-caster to visualize all time slices in a single frame. Similar visualization techniques have been used in medical imaging field in recent years [120]. Imitation of the 3D texture conveys the continuity of the data. Since the frames has readily been created for density map animation technique, density cube required less computation load. As for interaction, the participant was able to change the orientation of the cube to better analyze the desired portions of the data. The participant could also change the time scale which would either compress or extend the visualization along the time axis (axis-y in Figure 4.3).

4.1.2 Experiment

The methodology to analyze the effectiveness of visualization techniques has benefited much from Munzner's Nested Model for visualization design and validation[10]. While many studies on information visualization demonstrates "how" they measure potential usability of their techniques, we considered it useful to specify "which" evaluation method should be employed based on the contribution category of our findings. As Munzner states, the contribution of an information visualization study should be well defined since each contribution venue requires a different evaluation technique. According to her evaluation model, possible contributions of information visualization studies can be classified as (a) domain problem characterization, (b) operation and data type abstraction, (c) visual encoding and interaction design, and (d) algorithm design. In order to assist readers to comprehend how our study fits in the existing literature, we defined the level of the possible contributions of our study. The density map and density cube techniques have been implemented to aid the users to notice the possible trends and outliers in spatio-temporal data. On the one hand, the change in the data is abstracted to particular visual formations, which could be visually analyzed by the users. On the other hand, users are provided a number of interactions with the assistance of which they can rotate, manipulate, and scale the visualization to make more sense of the data.

4.1.2.1 Methodology

Much research has been done to investigate how effective different venues of visualization techniques are to aid the analysis. While many researchers suggested that 2D representations should not be the only technique of choice [31], [121], there are considerable number of studies suspecting potential benefits of 3D representations. We argue that the reconsideration of well-known characteristics of 2D (e.g. reasonable level of detail) and 3D (e.g. occlusion) representations when the spatio-temporal data visualization is considered might yield unexpected results. That is in other words, we aim to empirically evidence whether the commonly accepted characteristics of this two classes of representations are also valid in the spatio-temporal data visualization. The typical metrics for the performance of a visualization technique could be the "time" spent by a user on a given task and could be the "correctness rate" of the inferences made based on the observed data. From this vantage point, a study aiming to explore possible advantages of one technique over another might benefit from these metrics.

Experimental Design—To explore the advantages of the visualization techniques, we designed a within-subject experiment with a single independent factor (Type of the visualization technique: Density map vs. Density cube). As implied by

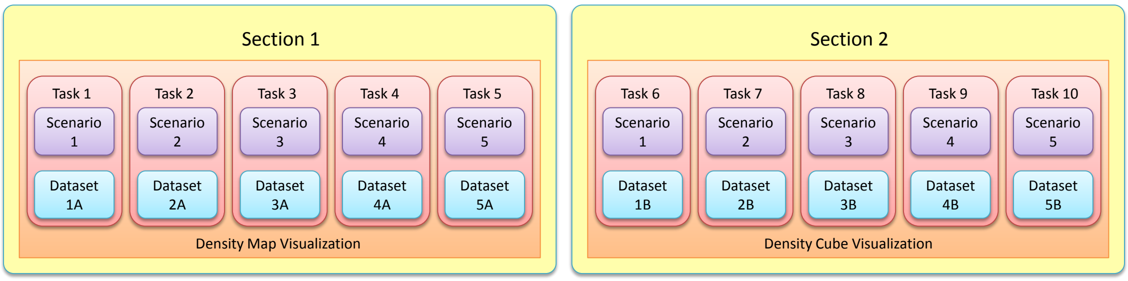

Figure 4.4: Scenarios, tasks, and datasets: Task sequence for experiment Group 1. Sequences for other groups can be derived from Figure 4.6.

the design of the experiment, each participant viewed the same scenario twice with different visualization technique. The performance of the participants was measured in terms of time and correctness rate.

We have conducted interviews with three data analyst in a location-based services company in order to formulate our experimental scenarios and tasks. To better simulate the cases that might exist in real data, five scenarios were created. As can be seen in Figure 4.4, the experiment comprised two sections, each of which included five tasks and was dedicated to one of the two visualization techniques, namely density map or density cube. According to our experimental design, participants were expected to finish both sections which include tasks generated based on the same scenarios, which could have caused learning effect[2]. To prevent possible learning effects, each scenario was realized with two datasets sharing the same data formation but generated for different location and sampling frequency. For example, if scenario 1 emphasizes an increasing trend in the data, two datasets (dataset 1A and 1B) generated for this scenario must have the same trend, but the datasets should be generated for different locations, and should span across different periods of time.

The order of the visualization techniques viewed by the participants was of importance to prevent bias on the collected data. Furthermore, the dataset-visualization technique combination had to be counter-balanced such that each combination had to be viewed by the participants with the same frequency. To address these problems, we generated four participant groups. As shown in Figure 4.6, dataset-visualization technique combination and the ordering has been fully stratified. For example, participants in Group 1 viewed the scenarios realized with Datasets A and visualized with density map technique in the first half of the experiment, while they viewed the same scenarios realized with Datasets B and visualized with density cube technique. Similarly, participants in Group 2 viewed the same sequence except for the order of the datasets.

[2]Learning effect is a case where the participants gain ability to accomplish tasks with better performance metrics as a result of repeatedly working on similar tasks.

Scenarios. We postulated that evaluation of visualization techniques should benefit from the formations (e.g. trends) that could exist in a real-world data. Such formations could inform the generation of scenarios that would be used during the experiments. After examination of the real data and the results derived from interview, we established five scenarios whose characteristics were, we suggest, capable and diverse enough to aid us in investigating how and in which ways a visualization technique would outperform the other.

Five scenarios employed in the experiments are demonstrated in Figure 4.5. In the first scenario, the usage of the application in a city has been demonstrated. By employing this scenario we aimed to facilitate an evaluation setting where users were expected to locate different levels of increase in the usage of the LBS service. The first scenario dictated data generation process with a moderate increase followed by an exponential one.

A typical workday in a city's downtown and suburban areas has been simulated in the second scenario. Moreover, a service failure in the downtown area between 12-3pm has been added to the scenario design. As can be seen in the Figure 4.5, a steep increase in the usage rate between 3-9pm implies a common case where the application is used more frequently at the end of a work day. In the suburban area, a sudden increase in the usage rate at around 3pm is shown.

The third scenario, similar to the second scenario, demonstrates a usage rate pattern in a city's downtown and suburban areas. However, in this scenario a weekend day has been simulated where the usage rate in downtown increased consistently during the day. A slight increase followed by a decrease in the suburban area has also been added to the design.

Commute of farmers from their villages to farm fields in rural areas was simulated in the fourth scenario. In this scenario, the usage rate decrease in all villages earlier in the day followed by an increase in the afternoon was demonstrated. The usage rate peak around 1pm in the farm field has been generated to give the insight that farmers could use the application more often while they work in the field.

Finally in the last scenario, participants were expected to infer a periodical trend of the usage rate in a city of moderate size. A typical work day usage rate trend has been repeated three times with random noise added to the slope of the trend.

Datasets. To demonstrate the scenarios in the experimental tool, we generated datasets with our data generator explained in Section 3. Our design dictates that participants will view each scenario twice; however, with different visualization techniques and datasets combinations (Figure 4.4). To prevent possible learning effects, two separate datasets, hereinafter *twin datasets*, for each scenario has been generated. The twin datasets share the same data formation specified according to the corresponding scenarios. However, they differ in the noise they include, the location and time information where and when the scenario takes place, and finally the length of the time period along which the scenario spans. For example, twin datasets 1A and 1B have the same trends as specified by Scenario 1. However, the scenario takes place in the cities Gonen and Tomarza in October 2011 and November 2012 in datasets 1A and 1B, respectively.

Tasks and Groups. As illustrated in Figure 4.4, each run of the experiment was comprised of 10 tasks, each of which is a combination of a scenario, an accompanying dataset, and the technique to visualize the dataset. The order of the scenarios was specified as shown in the Figure 4.4, and this order was never modified throughout the experiments. However, the dataset and the accompanying visualization technique that will be used for a given task was determined according to the experiment groups (Figure 4.6), which were created to prevent learning effect. For example, participants in Group 1 viewed tasks 1 through 5 which were created with datasets 1A through 5A visualized with density map technique while they viewed tasks 6 through 10 generated with datasets 1B through 5B visualized with density cube technique. All possible dataset and technique orderings were stratified among these four experiment groups.

Questions. In each task, users were expected to answer four or five multiple-choice questions. Please see Figure 4.7 for the number of questions asked in each task. We asked three types of questions to our participants during the experiment:

- *MinMax*, finding the minimum or maximum usage rate in the data set,

> *Q) In which of the following time periods was the service used most?*
> *a) May 10th, 2011 – May 14th, 2011*
> *b) May 22nd, 2011 – May 26th, 2011*
> *c) June 4th, 2011 – June 8th, 2011*
> *d) June 8th, 2011 – June 12th, 2011*
> *e) This tool is not informative enough to answer this question.*

- *Comparison*, comparing the usage rates at two different time stamps,

> *Q) Which of the following is true?*
> *a) The usage rate on July 13th, 2011 is higher than the one on July 18th, 2011.*
> *b) The usage rate on July 18th, 2011 is higher than the one on July 13th, 2011.*
> *c) The usage rates on July 18th, 2011 and July 13th, 2011 are almost at the same level.*
> *d) This tool is not informative enough to answer this question.*

- *Trend*, detecting trends in the data,

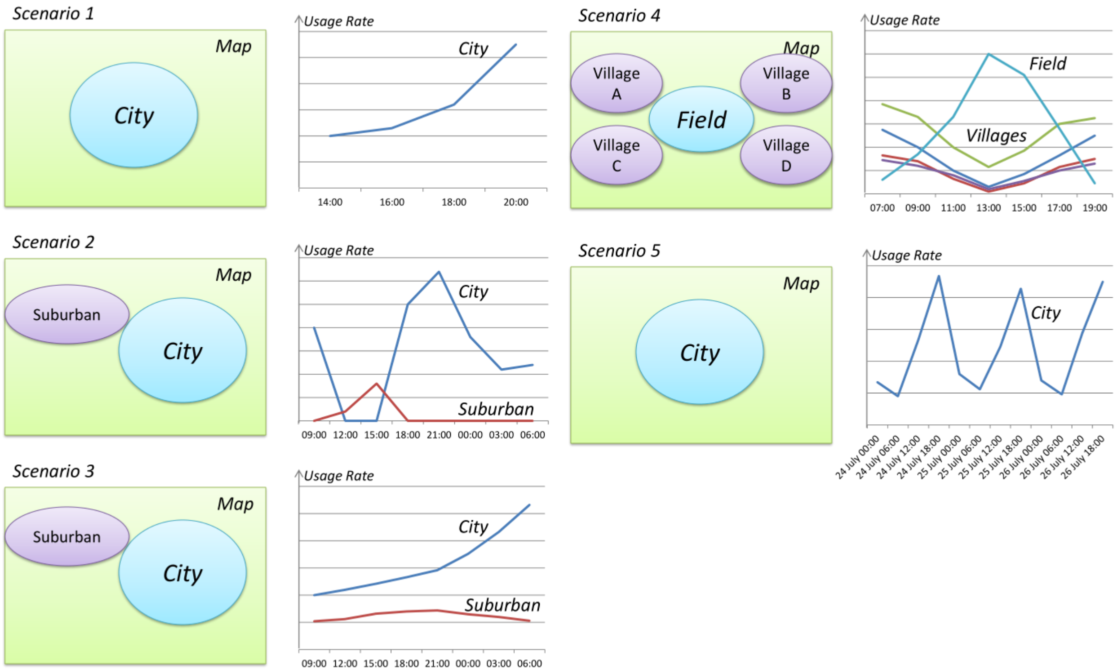

Figure 4.5: Scenarios and accompanying trends: Five scenarios created for the experiment. In the left column of the figure, general spatial distribution of the data is shown. In the right column, noise-free trends of application usage rate as a function of time has been depicted. The actual data represented in the experimental tool includes some level of noise in order to complicate the tasks and to imitate the real-world data.

Q) Which of the following is correct about the usage rate variances around city shown to you?

a) The usage rate moderately increases during the first half of the period. During the second half of the period, usage rate increases much more faster than it does in the first half of the period.

b) The usage rate moderately decreases during the first half of the period. During the second half of the period, usage rate increases saliently.

c) The usage rate moderately decreases during the first half of the period. During the second half of the period, usage rate remains at an almost constant level.

d) The usage rate moderately increases with a constant acceleration during the whole period.

e) This tool is not informative enough to answer this question.

In total, the participants answered 46 multiple-choice questions (23 for both density map and density cube techniques) during the experiment.

Procedure—Experimental procedure is shown in Figure 4.6. At the beginning of the experiment, participants were shortly briefed about the experiment followed by a visual capability test. Upon the completion of the test, participants were randomly assigned to one of the experiment groups. Depending on the experiment group, the first section of the experiment started with an instruction part for either density map

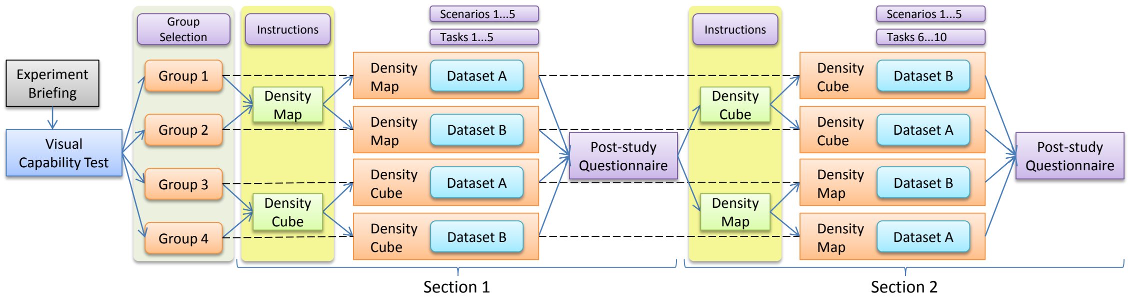

Figure 4.6: Experiment conduct flow.

Scenario	Tasks		Number of questions			
			MinMax	Comp.	Trend	TOTAL
1	Task 1	Task 6	2	1	2	5
2	Task 2	Task 7	2	1	1	4
3	Task 3	Task 8	2	1	2	5
4	Task 4	Task 9	2	1	2	5
5	Task 5	Task 10	2	1	1	4

Figure 4.7: The number of questions asked per task is demonstrated. Participants were expected to answer four or five questions in total for each task: Two *MinMax* questions, one *Comparison* question, and finally one or two questions related to *Trend* analysis.

or density cube technique. In instruction part, each participant received about 10 minute of training with our tool so that they had adequate amount of time to explore a small example dataset. Every question asked by the participants were answered by the experimenter to ensure that they understood the functioning of the tool and gained sense of how the data should be interpreted. Following the instruction part, participants were shown the first five tasks with the sequence explained in the previous subsection. For example (Figure 4.6), a participant in Group 2 took a training about the density map technique and completed five tasks generated with datasets B and visualized with density map technique. The second section of the experiment has been completed with the same sequence of processes.

Measures—Our measures included both objective and subjective data. We defined the descriptiveness capability of each technique in terms of *time* to solve the questions and the *correctness* value indicating how accurately the participants answered the questions. The *time* measure was collected for each task separately. The *correctness* value, collected both as task-based and as total, has been normalized so that it was ranged between 0 and 100.

Participants—14 participants (4 female), graduate and undergraduate students in a university in Turkey, volunteered for the experiment. Nevertheless, one of the cases was discarded due to the color vision deficiency (CVD) [122] of one of the participants.

4.1.2.2 Results

Analysis of the collected data was conducted in three steps; (1) effects of 3D visual capability baseline, (2) analysis of task completion time, and finally (3) analysis of correctness rate. To comply with the question types discussed in the previous subsection, analysis of task completion time and correctness rate was evaluated separately for minimum-maximum, comparison, and trend questions.

Effects of 3D Visual Capability Baseline—We performed general linear model (GLM) repeated measures in order to analyze the variance of the correctness and time along with the 3D visual capability as the covariate. As the result of our analysis, we found out that visual capability did not have an effect on time measure, $F(1,6) = 1.82$, $p < .23$, while it was marginally effective[3] on correctness measure, $F(1,6) = 4.652, p < .1$, implying an opportunity for further investigation with visual capability as a factor of interest.

Analysis of Task Completion Time—Task completion time performance of our participants was evaluated separately for minimum-maximum, comparison and trend

[3]While reporting the statistical results, marginal effect is referred to the p value between 0.05 and 0.1.

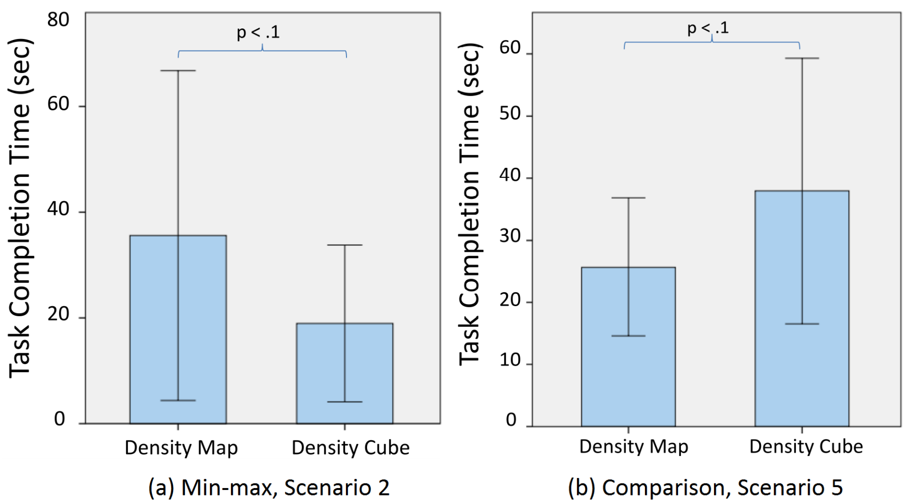

Figure 4.8: Exceptional cases in the analysis of task completion time: (a) Task completion time analysis of minimum-maximum questions of the second scenario: Participants answered the questions faster in minimum-maximum question category when using density cube visualization. (b) Task completion time analysis of comparison questions of the fifth scenario: Participants answered the questions faster in comparison question category when using density map visualization.

Question Type	Visualization Technique	Means and Standard Deviations	Significance Levels
Trend	Density Map	M=53.3, SD=18.5	t(12) = 1.283, p > .1
	Density Cube	M=38.5, SD=31.9	
Minimum-Maximum	Density Map	M=38.0, SD=17.6	t(12) = .054, p > .1
	Density Cube	M=37.6, SD=18.6	
Comparison	Density Map	M=21.1, SD=18.6	t(12) = -.276, p > .1
	Density Cube	M=23.0, SD=22.8	

Figure 4.9: Statistical results for analysis of overall time measure for each question type: No significant difference could be found between the task completion time measures of the two visualization techniques (all $ps > .1$) (Means of task completion times are in seconds.).

questions. For each question type, analysis of each scenario (3 x 5 paired t-tests) and total time spent across all scenarios (3 x 1 paired t-tests) was conducted.

The comparison of the total time periods spent on each technique (for each question type) showed that neither technique assisted participants in solving the tasks in less time than the other technique did as demonstrated in Table 4.9. However, a notable exception occurred for the minimum-maximum questions of the second scenario, $t(12) = 1.862, p < .1$, where the density cube technique ($M = 18.95, SD = 14.85$) outperformed the density map technique ($M = 35.58, SD = 31.2$) (Figure 4.8(a)). Another exceptional case existed for the comparison questions of the fifth scenario, $t(12) = -1.861, p < .1$, where the density map technique ($M = 25.72, SD = 11.07$) aided users to better compare the usage rates than the density cube technique did ($M = 37.97, SD = 21.33$) (Figure 4.8(b)).

4.1.2.3 Analysis of Correctness Measure

Analysis of correctness measure revealed significant findings about how visualization technique was predictive on our participants' success. For each question type (trend, minimum-maximum, and comparison), analysis of overall success rate of our participants was conducted. According to the results of paired t-tests applied on the correctness measures, we found out that participants were able to answer both trend ($t(12) = 2.215, p < .05$) (Figure 4.11) and comparison ($t(12) = 3.482, p < .01$) (Figure 4.12) questions more correctly when they viewed the data with density map technique. Statistical results are presented in Table 4.10. However, there was no significant difference between the correctness measures of the two visualization techniques ($t(12) = 1.443, p > .1$) for minimum-maximum question type.

Question Type	Visualization Technique	Means and Standard Deviations	Significance Levels
Trend	Density Map	M=74.62, SD=12.66	t(12) = 2.215, p* < .05
	Density Cube	M=59.23, SD=21.78	
Minimum-Maximum	Density Map	M=70.00, SD=19.15	t(12) = 1.443, p > .1
	Density Cube	M=58.46, SD=18.64	
Comparison	Density Map	M=80.00, SD=20.00	t(12) = 3.482, p** < .01
	Density Cube	M=58.46, SD=20.75	

Figure 4.10: Statistical results for analysis of overall correctness measure for each question type: A zero value means that all users answered the questions wrong, where a value of 100 demonstrates the opposite. Participants answered the questions more accurately when they viewed the data with density map animation technique for *trend* and *comparison* question types. Please note that p values for *trend* and *comparison* questions are significant.

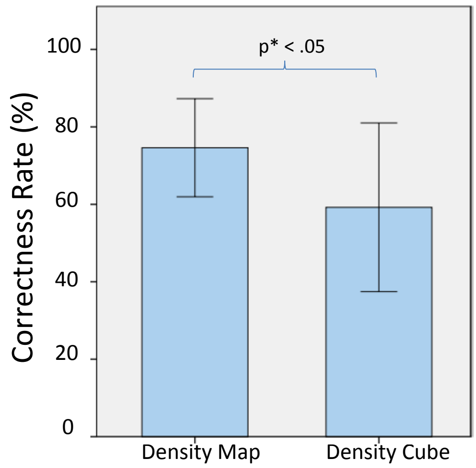

Figure 4.11: Comparison of density map and density cube techniques according to overall correctness rate in trend questions: Participants were able to answer trend questions more accurately when they viewed the data with density map technique.

4.1.3 Discussion

Listing the benefits and drawbacks of 2D and 3D representations for each kind of data and task is out of scope of this study. We argue that there exists enough evidence in order to argue that 3D representations in the visualization of spatio-temporal data should remain as a valid option as 2D representations are accepted to be by the existing literature. This is to say that, some common and well-known drawbacks of 3D visualizations seems not to be significantly present when the spatial and temporal data is to be analyzed. Even more, 3D representations of spatial and temporal activities can aid the users perform significantly better in particular tasks where 2D representations accepted as the de facto leading option [96].

The analytical use cases that can occur in spatio-temporal data are inexhaustible in terms of the possible formations (i.e. patterns), and tasks to be performed. The

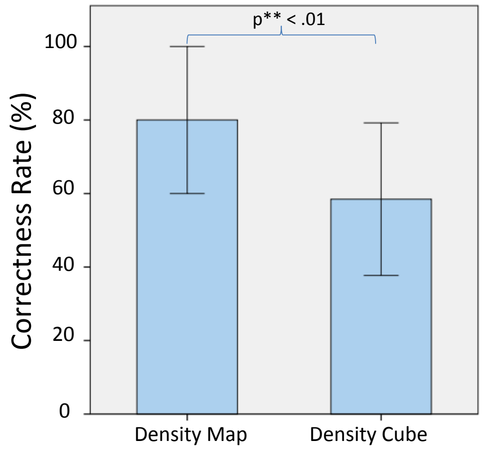

Figure 4.12: Comparison of density map and density cube techniques according to overall correctness rate in comparison questions: Participants were able to answer comparison questions more accurately when they viewed the data with density map technique.

setup for the visualization and the role of the users of the visualization are other factors contributing to this variety. In our experiment, we were able to cover only five of them that were found to be the best representatives of LBS data analysis by the field experts. As suggested by Andrienko and Andrienko [14], the tasks remain analogous across various visualizations, hence their task typology divided into two main groups as "elementary" and "synoptic" tasks. Some tasks may require a holistic view of data, while some others might lead users to make analysis based on a particular point in the visualization. According to this typology [14], our *minimum-maximum*, *comparison*, and *trend* questions correspond to *inverse comparison*, *direct comparison*, and *pattern identification*, respectively, task groups. This correspondence forms a basis for our discussion on whether and why 2D and 3D representations fail for particular tasks in spatio-temporal data visualization compared to the visualizations of other data types.

While analyzing the data represented in the second scenario, users observed the whole dataset as a set of 3D objects in our experimental tool enabling them to draw decisions more quickly than they did with density map animation. This is highly likely due to the fact that users tend to model the data in their mind as a whole rather than interpreting data in smaller chunks, causing them to browse through all data as fast as possible [95]. Similar findings have also been reported in [31], [93]. Hicks et al. claim that "computational offload[4]" that occurs during the holistic observation of the data assists to complete the undirected comparison tasks in less time. Earlier works of Tversky et al.[123] and Baudisch et al.[124] imply that the static representation of motion may be more effective than animation. Robertson et al. stated that "static

[4]Computational offload refers to the gain in the cognitive effort of the participants in their experiment.

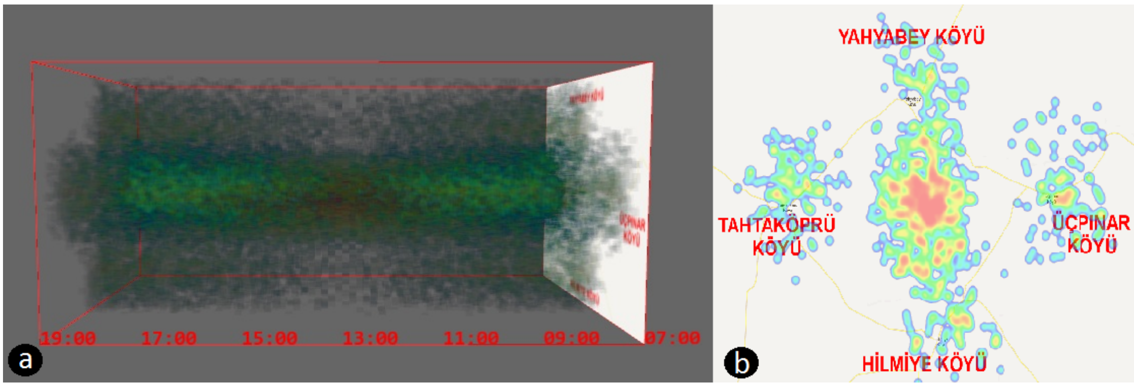

Figure 4.13: (a) A frame from density cube visualization: The problematic occlusion in Scenario 4 is depicted. (b) The density map animation of the same scenario: Occluded regions are viewed more clearly.

depictions of trends appear to be more effective," supporting our findings about the density cube visualization, which is actually a static representation of all the data for a given time period and region. According to the previous work along with our findings, the holistic overview capability of 3D representations suggested in these studies seems to be valid also for the spatio-temporal data visualization.

In their experimental study, Hicks et al. [93] report various performance levels of 2D graph, 3D plot, and 3D helix visualizations of e-mail usage log, which they consider as temporal data. They conducted a laboratory experiment to compare the performances of these visualizations against three groups of tasks, namely information retrieval, directed, and undirected comparisons which can be classified as *information look up*, *direct comparison*, and *inverse comparison*, respectively, according to [14]. They reported that 3D representations poorly performed in terms of overall task completion time compared to 2D graph. However, in our experiment, we were not able to observe any significant superiority of 2D representation over 3D in terms of overall task completion time. The scenario-wise task completion times of the two representations also were not significantly different. The fact that 2D representation were not able to outperform 3D could potentially be due to the higher dimensionality of our spatio-temporal data compared to 2D temporal data used in [93]. Our participants spent more effort while analyzing with density map technique probably since they had to navigate to the other timestamps of the animation. On the other hand, participants were able to partially view the 2D temporal data used in [93]. Density map animation technique leads analysts to focus on a representation of data for a unique moment of time causing the need for traversing back and forth in time dimension of the visualization tool during the analysis process, which was also reported in [31]. Kjellin et al. also reported that 2D visualization of spatio-temporal data led to a better analysis performance particularly when the tasks required detect structures in the visualization were invariant to Euclidean and similarity transformations [96]. Nevertheless, we could rarely observe the significant superiority of the 2D representation in terms of either time or accuracy during our experiment.

As Hicks et al. reported, 3D plot of the temporal data facilitates convenience and accuracy with the undirected comparison tasks [93]. Nevertheless, we did not observe similar results in our experiments where our participants were able to make more accurate inferences with density map technique (our 2D implementation). Particularly, they completed the tasks with more accurate answers while they were analyzing the fourth scenario where the data were more cluttered compared to other scenarios. As seen in Figure 4.13(a), locating minimum or maximum usage rates (inverse comparison task) with 3D representation has been a challenging task due to occlusion as reported by both our participants and the experiment results. On the other hand, density map technique as our 2D representation implementation allowed users to delve into the details of the data and draw more accurate conclusions. The occlusion problem inherent in 3D visualizations is more apparent in the visualization of spatio-temporal data possibly due to the increased dimensionality compared to temporal data.

Another difficulty with 3D visualization is complexity of making accurate size estimate due to the perspective foreshortening [91], [125]. Heights and widths that are at different distances from the user complicates comparing patterns (categorized as behavior comparison in [14]) as suggested by [126]. Our participants were able to locate trends (as behavior identification in [14]) significantly more accurately with the 2D representation of the location-based services data. This result is inline with [14], [125], [126] suggesting that the perspective projection effect in 3D representation is also a deterious effect in the visualization of spatio-temporal data.

Subtle regular patterns (e.g. periodically repeating formation as in our fifth scenario and information retrieval tasks in [93]) might aid revealing invaluable inferences with prominent importance to the analysts. As suggested by our results and [93], 3D techniques might better unveil subtly changing patterns over time. However, the choice of 3D technique is of importance since not all kinds of 3D representations could aid the detection of regular patterns without introducing occlusion.

As discussed above, the inferences previously made about the visualization techniques seem to be failing for the visualization of the spatio-temporal data. Due to its intrinsic properties, given the findings of our experiment, analysis of spatio-temporal data requires reconsideration of the widely accepted properties of 2D and 3D representations. Nevertheless, more research has to be done to investigate how and in which tasks 2D and 3D representations are effective.

Visualization of spatio-temporal data still remains as a daunting problem not only because of the variability of the cases that could be observed, but also due to the fact that particular cases usually require their intrinsic approaches. Even more, well-known aspects of 2D and 3D representations in visualizations of other kinds of data might not directly be applied when the data at hand is spatio-temporal. Given the complexity of the visualization context space comprising the task type, data type, and the user role, it is almost infeasible to cover all research space in order to conclude that 2D representations should be the first option to be considered [91]. We conducted an experiment where we treated our participants with five scenarios defined by field experts each of which is visualized with both density map and density cube techniques, relatively simple examples of 2D and 3D representations. Based on the findings of our experimental study, along with the findings of the previous work [93], [96], we claim that there is enough evidence that 2D representations seems not to be significantly better than 3D representations while analyzing spatio-temporal data.

4.2 Study-4: Designing Progressive and Interactive Processes for High-Dimensional Data Analysis

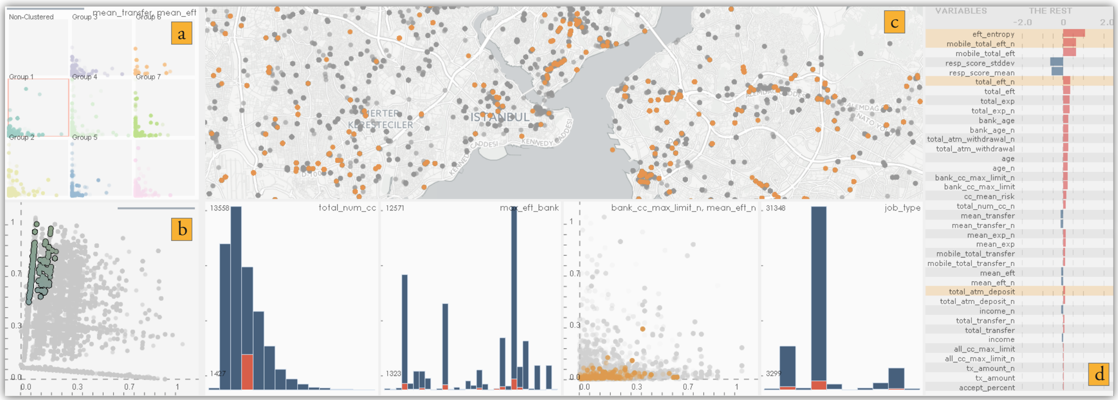

Figure 4.14: Iteratively refining a credit card transaction segment (details in Section 4.2.3) using progressive computations that are realized through a prototype built according to our design recommendations for temporally optimized analytical processes. Transaction segments are generated either through clustering (a) or through selections on a plot showing principal component analysis (PCA) results (b). Both the clustering and PCA computations are done "*online*" and the visualizations continuously update (according to the *three levels of operation*) either until the user changes the conditions to re-initiate the computations or until all the data is consumed. Subsegments are further refined through accompanying views (c, middle views), and the difference view (d) describing the segment.

In interactive data analysis processes, the dialogue between the human and the computer is the enabling mechanism that can lead to actionable observations about the phenomena being investigated. It is of paramount importance that this dialogue is not interrupted by slow computational mechanisms that do not consider any known temporal human-computer interaction characteristics that prioritize the perceptual and cognitive capabilities of the users. In cases where the analysis involves an integrated computational method, for instance to reduce the dimensionality of the data or to perform clustering, such non-optimal processes are often likely. To remedy this, progressive computations, where results are iteratively improved, are getting increasing interest in visual analytics. In this study, we present techniques and design considerations to incorporate progressive methods within interactive analysis processes that involve high-dimensional data. We define methodologies to facilitate processes that adhere to the perceptual characteristics of users and describe how online algorithms can be incorporated within these. A set of design recommendations and according methods to support analysts in accomplishing high-dimensional data analysis tasks are then presented. Our arguments and decisions here are informed by observations gathered over a series of analysis sessions with analysts from finance. We document observations and recommendations from this study and present evidence on how our approach contribute to the efficiency and productivity of interactive visual analysis sessions involving high-dimensional data.

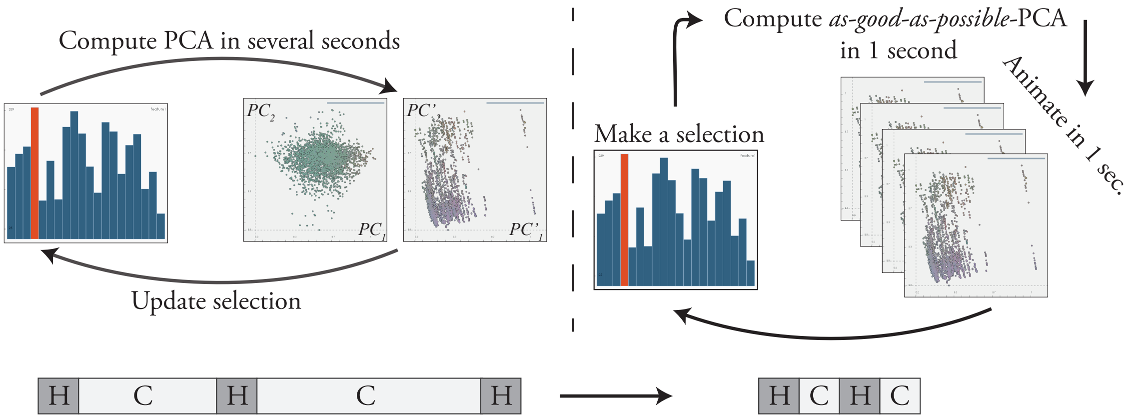

Figure 4.15: An example for optimizing an analytical process against the three human time constants [15]. In a conventional setup (left), the user first requests a (re-)computation of PCA results (with a selection of variables), then she waits a certain time. This waiting could potentially interrupt the dialogue between the user and the computer (let's check my Inbox!). In order to address this, our suggested optimization (right) computes PCA results *as good as possible* within one sec. in response to a selection by the user and the visualizations animate in one sec. to display the new results. The shown H–C–H–. . . - abstraction indicates the pattern of interaction (the lengths indicate the time spent and H: Human, C: Computer).

4.2.1 Designing Interactive Progressive Analysis Processes

Here we present the fundamental building blocks to facilitate interactive and progressive analytics processes within the context of high-dimensional data analysis. We discuss the design of such processes from four different perspectives to address four key questions: **Q1:** how to configure analytical processes such that ***the temporal capabilities*** of analysts are respected? **Q2:** how to integrate ***underlying computational mechanisms*** that are capable of performing progressively? **Q3:** how to devise suitable ***interaction methods*** to facilitate processes that embrace progressiveness? **Q4:** how to inform analysts on the various aspects of progressive computations through appropriate ***visualizations***? In the following, we consider these questions and present a set of techniques and methods to address these. Where appropriate, we externalize the justification for our methodological decisions in the form of design recommendations (**DR**s). These recommendations are informed both by the two-month-long case study we carried out and by other studies on in high-dimensional data analysis [127], [128].

Wherever appropriate, we relate to the recent literature on theoretical and technical frameworks [23], [129]–[134] and design studies on progressive analytics [131]. Schulz et al.'s [132] framework comes the closest to our work where the authors provide a theoretical framework and a data/computation model to facilitate incremental visualizations. Their approach, although, not giving specifics on visualization and/or interaction design, provides valuable pointers to the different aspects to consider such as selecting metrics and determining a quality/quantity trade-off.

4.2.1.1 Respecting the human-time constraints

Based on their investigation of psychology literature, Card et al. [15] present three *human time constants* that characterize the temporal characteristics of related human capabilities. These constants are reported to be highly important to achieve an optimal communication between the user and the computer. The first constant relates to the *perceptual processing* level at which humans are able to perceive changes in consecutive images as visually continuous animation. To achieve a visually smooth animation, the images need to be updated at least 10 times per second. The second constant addresses the *immediate response* level at which the parts in a communication are exchanging, forming a dialogue. The communication is interrupted if there is no response from the other party within about one second. The third time constant is the *unit task* constant which determines the limits for an elementary task to be completed during such a dialogue. This constant is reported to be more flexible and defined in an interval between 10 to 30 seconds. With the guidance of the human time constants, we aim to improve the dialogue between the human and the computer during visual analysis sessions. We achieve this by adjusting the system to operate at three levels (at three time scales of interaction). These levels correspond to the three human time constants and thus, are associated with certain temporal limits as shown in Table 4.1. In this study, we limit our discussion to a subset of visual analytics methods, including i) linking & brushing, ii) the integration of computational tools and interactive methods, iii) a visual representation of the computational analysis results. We argue, however, that our model is more general and that also other processes in visual analytics fit well into it. Briefly, we consider a *unit task in visual analytics as a sequence of actions and reactions where the reactions can be given by animated visualizations* and the three levels of operation moderate how such a task can be carried out at an optimized fashion by respecting the human time constants. Our first design recommendation emphasizes this role of human time constants:

DR1: *Employ human time constants as the underlying theoretical framework that governs the pace of interaction in analytical processes.*

Table 4.1: The three levels of operation, the corresponding human time constants [15], and the associated time limitations

Level	Operation Level	Human time constant	Response time (sec.)
Level 1	Visualization update	Perceptual processing	0.1
Level 2	Human-computer dialogue	Immediate response	1
Level 3	Analytical task completion	Unit task	10 - 30

Level 3: Unit task completion—This level determines the temporal range in which an analytical unit task is completed. Such an analytical task is performed to answer a specific question related to the data. Such a task involves a sequence of inputs from the user and corresponding responses from the computer. Our consideration of analytical tasks here is more high-level and pragmatic than those suggested in the literature [9], [19] and can be likened to data analysis related ones [135], such as finding groups, investigating relations, etc., however, we also allow for simpler tasks such as "changing a selection of items in one view (in order to explore inter-dimensional relations) and observing how the according focus+context visualization in a linked view changes". We facilitate this level mainly through the *keyframed brushing* (Section 4.2.1.3) where the tasks are bounded with fixed completion times, i.e., 10 sec., 20 sec., or 30 seconds.

Examples of such a unit task include (a) changing a selection of items in one view (in order to explore inter-dimensional relations) and observing how the according focus+context visualization in a linked view is changing, (b) observing the grouping structure of data items when different subsets of the variables are used, for instance, in clustering. The time constraints for this level of operation are more flexible according to Card et al. [15] and depend on the analytical unit task.

In order to ease the construction of different patterns of selections, we developed an interaction mechanism called *keyframed brushing* (see Section 4.2.1.3). With this mechanism, it is possible to frame unit tasks with fixed completion times, i.e., 10 sec., 20 sec., or 30 seconds – Card [15] leaves this window slightly flexible. Such a unit task is then a sequence of actions and reactions between the human and the computer as discussed in the following.

Level 2: Human-computer dialogue—This level is mainly responsible to maintain the dialogue nature of the visual analysis process. It ensures that the communication between the user and the computer is not interrupted. Specifically, this level focuses on maintaining a guaranteed response time (1 sec.) when integrated computational tools are utilized. This mechanism realizes an uninterrupted dialogue by making sure that the immediate response capability of the user is exploited.

Maintaining the one-second response time is not straightforward when the computations are complex and the data is large. Our solution to approach this problem is to compromise the quality of the results by computing the best possible result within the limited time frame. Similarly, in computer graphics, reducing the quality of the rendering process to maintain an interactive frame rate is common practice and related methods are usually referred to as *progressive refinement*. [136]. In order to achieve this, we make use of online algorithms together with an adaptive sampling strategy (more in Section 4.2.1.2).

Level 1: Visualization update—Animated transitions have been proven to be helpful when the motion of the data items are of importance [137] and can help to avoid *change blindness* [138]. Therefore, we make use of animated transitions between the different computational results that are generated as a result of the dialogue occurring at the second level of operation. The visualization update level moderates the update rate of animated visualizations and secures the perceptual processing of the animations in the visualization. In order to create animations that are smooth in the eye, the lower bound for the update rate should be 10 Hz [15].

Similar limitations on frame rate are highly important in computer graphics and virtual reality fields [139]. According to a study on the effects of update rate on the sense of presence in virtual environments [140], 15 Hz is an optimum rate for updating animations. In this study, we include the different update rates as reported in literature, and animate the visualizations at either 10, 15, or 20 Hz.

4.2.1.2 Incorporating Progressive Algorithms

Here we describe how computational methods can be integrated within interactive analysis systems and provide the details of how we seamlessly integrate different algorithms. We also present an adaptive batch sampling strategy.

Online Algorithms— In order to maintain the temporal limitations set forth by the human time constants, we develop a mechanism where the computational tool guarantees to respond within a fixed period of time (i.e. one second). To achieve this, we make use of *online algorithms*, which are capable of processing the data piece-by-piece, sequentially [141]. These algorithms do not need the whole data to operate and can update the results as new data becomes available. Such algorithms are popular in machine learning literature in cases where the datasets are very large to fit in the memory or it has a streaming nature [142].

In the machine learning literature, there are online versions of computational tools that are frequently used in visual analytics, such as principal component analysis [143] and clustering [144]. One common method to use online algorithms is to pass the data in small batches to have a lower memory footprint [141]. To be able to utilize this incremental computing nature of online algorithms, we use them in combination with a random sampling method that adaptively adjust the batch size to ensure efficiency in the iterative computations (see Algorithm 1).

This approach assures that the computations are finished and the associated visualizations are updated within the temporal limits. However, due to the fact that the results are computed on a limited sample, the results are usually not as accurate as one would achieve if the whole dataset was used. Therefore the algorithm continues to run (in a separate background thread) after the first response is given and consumes more and more of the data every second. This implies that the results are getting more and more accurate as the user observes the result without being disengaged from the communication. In the context of this work, we incorporated the online versions of two popular computational tools, PCA and clustering (a version with similar principles as the k-means algorithm). In Algorithm 1, these tools correspond to the module O. It is also important to mention that these algorithms should be able to run on user-selected subsets of both the items and the dimensions of the data. We make the following recommendation that relates to the underlying technical framework.

DR2: *Employ online learning algorithms that are capable of handling data in sub-batches to perform computational tasks.*

Online PCA—Online PCA algorithms make use of an incremental updating of the singular value decomposition (SVD) of the data matrix [143][5]. In this study, we refer to online PCA and incremental PCA interchangeably. Here, we use the incremental methodology described by Ross et al. [143]. At each iteration, the SVD is updated with the incoming data and the principal components loading are updated accordingly. However, *despite the PCs modeled on only the seen data, i.e., partially fitted, the final projections are applied to the whole dataset.* The results are then visualized through scatterplots where the axes are the first two principal components (see Figure 4.15).

[5]Incremental PCA is utilized from http://scikit-learn.org/.

Algorithm 1 Online computation with random batch sampling

procedure ComputeInFixedTime
  $O:$ *Online computation module*
  $D:$ *Data*, $size:\ n \times p$
  $Q:$ *Random sampling queue*, $size: n$
  $t_{lim}:$ *human time constant* ▷ Fixed to 1 sec.
  $t_0:\ currentTime()$
  $timeLeft:\ t_{lim}$
  $b: batchSize$ ▷ A conservative size, e.g., b = 100
  while $Q.notEmpty()$ **do** ▷ Until all samples are used
    while $timeLeft > 0$ **do**
      $i \leftarrow Q.pop(b)$ ▷ i is a vector of size b
      $x \leftarrow D[i]$ ▷ x is a matrix of size $b \times p$
      $\Delta \leftarrow O.update(x)$ ▷ Δ = computation time
      $b \leftarrow adaptBatchSize(b, \Delta, t_{lim})$
      $timeLeft:\ t_{lim} - (currentTime() - t_0)$
    end while
    $O.returnResults()$ ▷ Visualization is updated
  end while
end procedure

Online Clustering—Similarly for clustering, we use an online clustering algorithm that can operate on sub-batches of data incrementally defined by Sculley [144][6]. This algorithm takes a k parameter as an upper bound on the number of clusters. At each iteration it includes a new batch and appropriately merges/splits the clusters. Notice that in clustering, at each iteration of Algorithm 1, only a subset of the items is "added" to the clustering model to revise the cluster centres. This is followed by a step where all the "seen" points are associated to these revised cluster centres. The results are then presented on a small multiple visualization where each cluster is represented by a multiple and a distinct color (taken from ColorBrewer [145], see Figure 4.14-a). The non-clustered (those not yet processed) however are displayed within the first multiple. As computations iterate, the non-clustered items are distributed over to the other clusters. Although such an approach does not provide the labeling for all the items, it provides an overview of the clustering structure, i.e., the number and relative sizes of clusters. Therefore, we visualize the results of such a clustering using an abstract visualization, where each cluster is represented by a circle with a distinct color (taken from ColorBrewer [145]) where the size represents the number of items it includes (Figure 4.14-a). This view is also animated and shows how many items are migrating between different clustering results (Figure 4.14-b). Here, we prefer an abstract visualization due to the lack of an inherent spatial mapping of clusters to 2D.

[6] MiniBatchKMeans is utilized from http://scikit-learn.org/.

Online Statistical Computations—In addition to these algorithms that often operate on the dimensions, there are also computational methods to estimate statistics from the data. In our prototype, for instance we have a view that dynamically computes statistics (Cohen's D [146]) between the selected and not-selected items (see Figure 4.14-c). Although we do not incorporate a progressive version of these calculations and the visualization (partly due to the lower computational costs), we include them under this section for the sake of completeness.

Adaptive Sampling—Ensuring the temporal constraints within our approach is of key importance and sampling is the key mechanism to ensure a good quality / efficiency trade-off. In order to ensure efficiency in the computations while still maintaining the temporal constraints, we developed an adaptive sampling strategy to improve convergence times by adjusting batch sizes (i.e., the amount of data that is handled within an iteration) adaptively. The function $adaptBatchSize(b, \Delta, t_{lim})$ (Algorithm 1, line 14) is where we perform this iterative revision. In this function, we start by finding a multiplier $m = t_{tim}/\Delta$ and if $\Delta < t_{tim}$ we increase the batch size b with $b = b * \sqrt{m}$ and if $\Delta > t_{tim}$ we reduce the batch size by $b = b/m^2$. This ensures that the sampling method finds an optimal batch size that can be computed under the time limits while ensuring reduced overall completion time. Note that we have chosen conservative multiplicator factors $(\sqrt{m}, m^2)$ to account for the non-linear nature of computational complexity. The performance of such adaptive sampling strategies can even be improved further through techniques such as predictive caching [147] or binning [130].

DR3: *Employ an adaptive sampling mechanism that estimates suitable sample sizes for computations to ensure efficiency in convergence while still respecting the temporal constraints.*

4.2.1.3 Interaction Mechanisms

The above mentioned online computation methods are used in integration with the conventional linking & brushing and the keyframed brushing mechanism. In the following, we present different approaches to enable the interaction within such systems. We are running separate computational threads to carry out these two operations. Both of these threads manage the time limitations as outlined in our algorithm. When the results of the computations are ready, a signal is sent to the visualizations to update the results in the corresponding visualizations. We also evaluate the performance of online computations in terms of the stability of the results and the amount of data that could be processed within the time limitations and present these in Appendix B.

Immediate response & progression granularity—Our online computation mechanism immediately responds to user input such as a new selection of a group of dimensions (similar to Figure 4.15). The interactive input triggers our algorithm which returns the first, approximate result within one second.

DR4: *Facilitate the immediate initiation of computations in response to user interactions that limit the domain of the algorithms.*

Fine-grained progression may not be always desirable. In particular circumstances, frequent updates on the analytical model can generate additional cognitive load on the users, thus leading to frustration. Refer to Section 4.2.3.3 for an example situation that arose in our case study.

DR5: *Provide users with interaction mechanisms enabling management (pause, step size, re-run) of the progression.*

Visual analytic solutions with multiple progressive views may lead to a problem which we call *fluctuation*, a case where views process the data at different rates, and reach to their final states at different moments. We observed that such cases frequently led the analysts to confusion (Section 4.2.3).

DR6: *During the interaction design of visual analytic solutions, consider the effects of possible fluctuations due to unaligned progression in multiple progressive views.*

Keyframed brushing— The *keyframed brushing* mechanism is intended to reshape (a certain subset of) analytical tasks as a dialogue while keeping the user engaged. This methodology has been shown to generate dynamic visual summaries [146] and structured selection sequences [148] and we employ this technique here. The user defines two or more brushes (according to his/her analytical goal), similar to defining key frames in computer-assisted animation [149]. Using these *key brushes*, a sequence of *in-between* brushes is generated automatically. After the brush sequence is computed, the system starts traversing through this sequence without the need for further input by the user. Depending on the user's preference, the complete sequence is traversed in 10 sec., 20 sec., or 30 sec., and moving from one brush to the next takes 1 second. Here, traversing the whole sequence can be considered as *a task operating at Level 3* and *moving from one brush to the next as operating at Level 2* as defined above. Keyframed brushing enables the user to focus on the linked views that display the results of the animation rather than paying attention to moving the brush in a particular fashion. Refer to Section 4.16 for a demonstration of cases where keyframed brushing proves to be helpful in cases that are hard to investigate with manually modified brushes. This mechanism has a utilization both as an automated linking & brushing operation and as a method to interact with the computational tools. In order to construct brushing-based animations, we enable the specification of key brushes through conventional visualizations, such as scatter

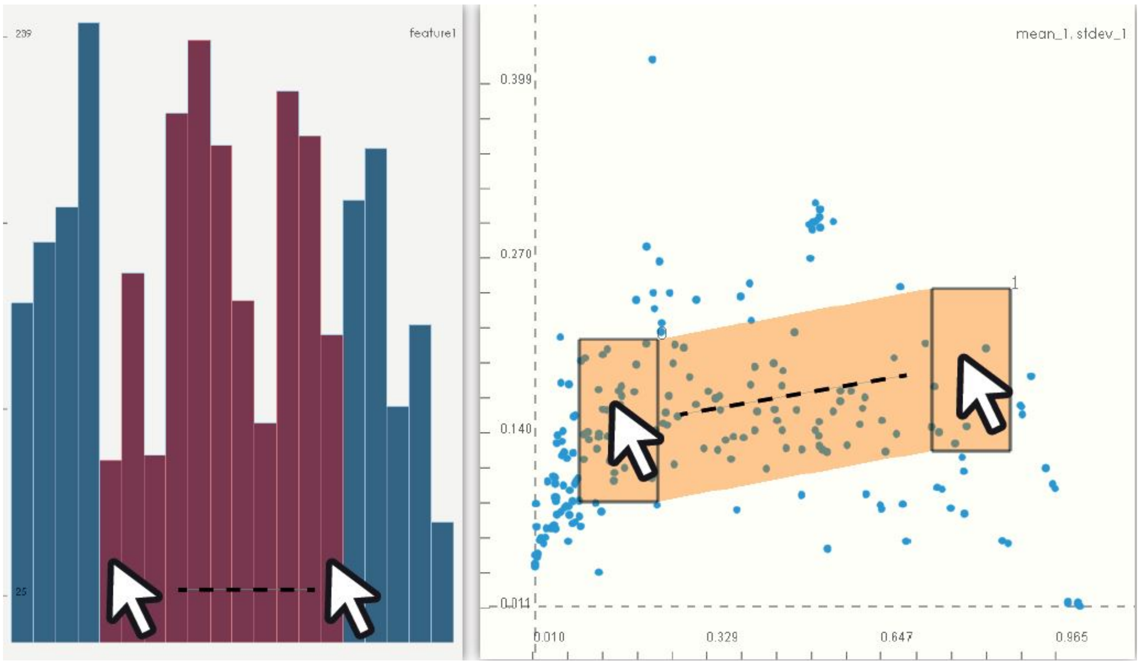

Figure 4.16: Keyframed brushing can be performed in a histogram or in a scatterplot. Interactively determined start and end selections ("key brushes") are accompanied with computed "in-between" brushes.

plots and histograms. In Figure 4.16, the interface to define a brush sequence can be seen. We draw overlays to abstract the range of the final brush sequence. The use of such interactive mechanisms led to the following recommendation:

DR7: *Provide interaction mechanisms to define structured investigation sequences for systematic generation and comparisons of computational results.*

4.2.1.4 Design considerations for visual representations

Here we discuss the decisions made whilst designing visual representations for progressive analytics. We focus on how to effectively incorporate animation and communicate progress and uncertainty.

Animated Transitions—In our approach, we use animated transitions to support the interpretations of changes while comparing different results of a computational tool. Although the usefulness of animations in visualization has also been disputed [150], there are several good examples where animations proved to be useful, both as transitions between different graphics [137] and between different projections of high-dimensional datasets [151]–[153].

Each computational result can be thought of as a key frame in an animation and the in-betweened frames are computed by the animation module. Animated transitions are controlled by the first level of operation and are done at 10 Hz or faster. A single animation sequence takes one second. For the sake of simplicity, we focus on animations that display PCA results and we start with a view V that shows the PCA projection of the data based on all the dimensions.

Immediate response animations—In this setting, the visualization responds immediately after one second. and in order to maintain the human-computer dialogue also in this case, the visualization V is fed with new, more accurate computation results every second. As a result, the points in V start animating to their new positions in the newly available PCA projection. However, if there is no apparently interesting structure in the first results, or at any instance, the user can update the selection. In this case, the current animation is stopped immediately and the view animates to the new computation results, instead. The animation ends when all the items are processed.

Keyframed brushing animations—Stolper et al. [131] articulates the challenge of progressive visualization as keeping a balance between showing most up-to-date information and keeping the analyst from being distracted by continuous updates. Accordingly in our case study (Section 4.2.3), the analysts reported that the continuous updates might not be desirable at certain instances and needs to be controlled. In order to accommodate this, we offer an alternative animation modality. These animations are triggered when the user performs a keyframed brush operation. A typical use is as follows: firstly, the user makes a keyframed brush sequence that selects different subsets of dimensions then observe the differences between the PCA computations that are done for each of these selections in the sequence. The system waits for one second before animating to the next result to give the user enough time to observe the results.

DR8: *Support the interpretation of the evolution of the results through suitable visualization techniques.*

Improving animated transitions—In the following, we present selected improvements to animated transitions. The first improvement is related to maintaining the coherence between two key frames (two computational results) of an animation. Such an improvement is important in order to preserve the mental map of the user [154] and similar challenges have been studied in other domains, such as in graph drawing [155]. In the case of PCA, the resulting principal components (PCs) are known to have arbitrary rotations and signs due to the nature of PCA [156]. Due to this fact, although the structure of the point distribution does not change, i.e., item neighborhoods stay the same, the PCs can come out flipped and/or mirrored. This makes it hard to follow the animations and creates arbitrary rotations. We solve this by checking the correlations ρ (using Pearson's correlation measure) of the axes between the first, x_1, y_1, and the second PCs x_2, y_2. If $\rho(x_2, y_1) > \rho(x_2, x_1)$, we flip the axes, and if $\rho(x_1, x_2) < 0$ (negatively correlated) we mirror the axis (mir-

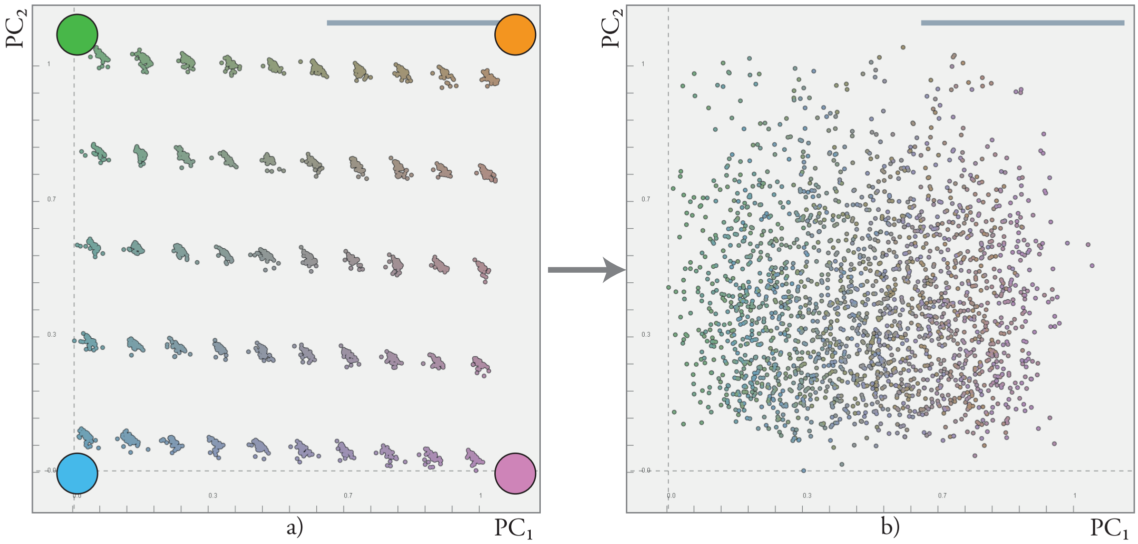

Figure 4.17: Coloring to enhance the communication of change between animation frames. A 2D color map from the CIELUV color space using a fixed lightness value is used (corners of the 2D map are shown). Points are colored according to their position at the beginning of the animation sequence (a) and this color stays the same for the whole animation (b).

roring check is done also for y). Similarly, in the case of cluster computations, the resulting cluster labels are in principle arbitrary. In order to make coherent transitions between key frames (two clustering results), we find a mapping between two consecutive labellings. We use a metric called Jaccard coefficient [157], which measures the overlap between two sets. For each cluster c_i in the current result, we compute the Jaccard values with all the clusters in the next frame c' and find the corresponding cluster c'_j with the highest $Jaccard(c_i, c'_j)$ to update the mapping accordingly.

Coloring is also used to support the tracking of changes in the animations in scatter plots. We map the color of each point based on their x, y coordinates in the beginning of an animation sequence (Figure 4.17-a) and the coloring stays constant for all points through the animation. The corners of the 2D color map and the resulting colors can be seen in Figure 4.17. The color map is constructed using an isoluminant slice of the CIELUV color space [158]. With this approach, we utilize a mixture of dynamic and static techniques in the visualization of change.

Communicating Progress & Uncertainty—The communication of progress and the uncertainty (or error) in the computational results are key aspects when using approaches that present approximate results such as ours. Both Stolper et al. [131] and Schulz et al. [132] refer to this as an important element of progressive systems. We offer a number of channels to support the users in analysis sequences. Firstly, whenever online computation results are visualized, we display a simple progress bar to inform on what percentage of the data has been consumed (see Fig. 4.18, the blue bar on top).

DR9: *Inform analysts on the progress of computations and indications of time-to-completion.*

Secondly, we suggest alternative visual representations that communicate the inherent uncertainty in the computations and also makes it harder to make micro readings (i.e., at observation level) when the presented results are more uncertain. In this visualization option (Fig. 4.18), we switch from a scatter plot to a binned representation where the bin size is adaptively adjusted according to the percent of the data seen in the result – the bin sizes are larger when the sample percentage is low, i.e., coarser grid at t_1, and gets smaller as the computations progress to, i.e., the finer resolution grid at t_4.

One aspect we do not present here is a quality metric as also suggested by Schulz et al. [132]. Any of these views are better supported if domain specific metrics to inform users with quantified measures of success are included. One possibility could be to also incorporate both visualization and data level metrics [159] as alternative heuristics of change towards a "good" solution. One alternative here is to use the *neighborhood preservation ratio* [160] as a measure of change between consecutive results.

DR10: *Inform analysts on the uncertainty in the computations and the way the computations develop.*

4.2.2 Interactive Progressive Analysis System Prototype

We realized our approaches and techniques within the context of a linked view system where several selections on multiple visual representations can be combined using Boolean operations. In order to have the variables as our main visual entities in some of the graphics, we employ a technique called dual-analysis [127]. From here onwards, we refer to this prototype as *DimXplorer* and we used this tool to visually explore the credit card transaction data and locate credit card expense segmentation (Figure 4.1) during our case study. The tool incorporates a map view on which the transaction data is geographically plotted, a difference view that displays the difference (using Cohen's D) between the selected and not-selected items (which can be likened to the deviation view as presented in Turkay et al. [127]) which also enables the interactive selection of features. As detailed earlier, a progressive PCA View visualizes PCA results (projected to the first two components) through animation and a progressive cluster view visualizes the cluster results where each cluster is represented as a small multiple and enables the selection of distinct clusters.

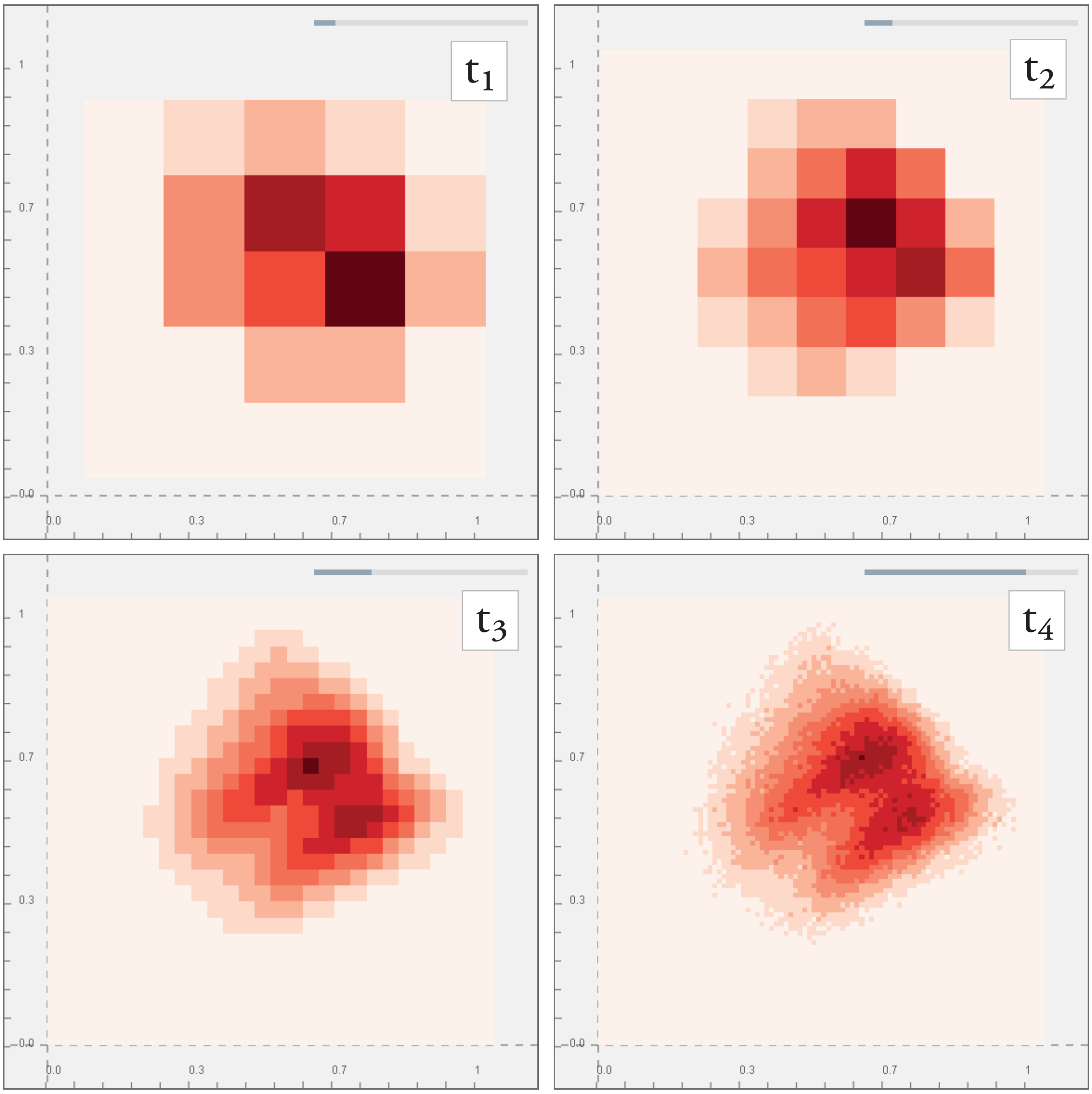

Figure 4.18: An alternative representation to communicate the uncertainty in PCA computation results. Here we inform users on the converge of the computations (thus the inherent uncertainty) through two channels: i) the progress bar to indicate the amount of data consumed – gray bar on top of all four ii) a density estimate representation of the points with a varying kernel size that makes micro readings from the plot harder where results are more uncertain, e.g., from t_1 to t_4.

4.2.3 Evaluating Progressive Analytics Methods

To inform the development and to evaluate the validity of our resulting solutions, we conducted a two-month-long case study during which we observed a group of analysts of a national bank from the EMEA (Europe, the Middle East, and Africa) region performing credit card expenditure segmentation analysis.

At the beginning of the study, we conducted *fly-on-the-wall* observations [73] and semi-structured interviews with *four* analysts to gain insight on the problem domain and their approaches to the problem. We were provided with an anonymized credit card transaction dataset collected between July 2014 and June 2015. We carried out four analysis sessions in the work environment of the domain experts. We introduced DimXplorer during the first session whereas the rest of the sessions were on credit card expenditure segmentation analysis. The last session, however, was devoted to analysis activity with *non-progressive* visual analytics to observe for differences.

Our methodology follows the insight-based evaluation methodology suggested by North [30] where we try to identify insight types and partially adopted an approach similar to Guo et al.'s work [161] that suggests the use of interaction logs together with insight-related investigations. We recorded all the session and transcribed the inferences made and questions asked during the analyses, and described the relationship between the applied guideline and the relevant actions taken during the course of data exploration.

4.2.3.1 Credit Card Expenditure Segmentation Problem

Credit card transaction segmentation is the problem of grouping credit card transactions based on the similar demographics and financial metrics of customers. Location, amount, and frequency of the credit card transactions, demographics and financial well-being of the customers are important factors that might form expenditure profiles. Crucial to the decision-making and long-term strategy-making, the typical outcome of such an analysis is a list of customer or expenditure profiles with subcategories representing customer groups having similar spending behaviors and demographics.

Segmentation analysis takes many aspects of financial metrics and demographics as input, providing a practically inexhaustible hypothesis space. The analysts reported that they form the questions as a starting point for their analysis by carrying out brainstorming sessions or consulting to the experience of senior analysts.

Transaction and customer profile segmentation tasks typically involve clustering and classifying operations on large, high dimensional datasets. Such an endeavor strives to locate interesting patterns, and derive research questions and hypotheses from a large selection of underlying phenomena. We argue that the application of progressive analytics in this context can enable quick hypothesis generation and iterative fact mining.

Credit Card Transaction Data—The data analyzed during the case study involved more than 300.000 credit card transactions of more than 5.000 customers. Each transaction included location, amount, demographic information (e.g. job type, marital and education status, income, age), and financial metrics (e.g. mean transfer and deposit, risk and response score, number of credit cards, entropy of transfers) of the customers. See Appendix-B for the full list.

Inference	**Moment**	**Explanation & Quote**
Quote (Qu-1)	Session 2-1 (7:01)	A1: *"Let's try some other demographic features as this selection seems like not going to bring new patterns. We can generate so many new hypotheses in a very short time without waiting for the whole calculation to end."* A1: *"...visualization is quite engaging as we don't have to wait for even a moment to get some initial results."*
Quote (Qu-3)	Session 2-2 (13:35)	A1: *"It seems like the clustering will not change. Almost all of the data has been calculated, let's switch to some other set."*
Quote (Qu-7)	Session 3-1 (3:10)	The team tried a new feature set and immediately observed a *good* separation of data points. However, after only 15-20 seconds, the separation dramatically changed. A2: *"Well, I think waiting for a while might be a good thing."*
Quote (Qu-9)	Session 3-1 (51:03)	During the analysis, A3: *"I've just seen a high response score for the selected cluster, but it has just gone away."* As the clustering algorithm continued to its calculations, the data points moved to other clusters changing the pattern A3 previously discovered: *"Wouldn't it be nice to have a button that pauses the progressive visualization?"*
Insight (I-3)	Session 2-2 (24:45)	A1: *"The customers working with other banks seem to be more profitable ones as their financial metrics draws a better picture."* (higher transfers, EFT, higher response score, etc.).
Testing (Te-1)	Session 2-3 (00:47)	Hypothesis (insight) I-3 has been rejected. A2: *"Customers with 2 or 3 credit cards and low credit card limits seem to represent low financial profile."*

Table 4.2: A selection of inferences made during and quotes taken away from the case study. Abbreviations *A* and *R* stands for *Analyst* and *Researcher*, respectively. A comprehensive list of all inferences, quotes, questions, and hypotheses can be found in Appendix-B.

Analysis Tasks—During the *precondition phase* [162] of our study, we considered the obvious opportunity for exploratory visual analysis of the data where they could form data-driven research questions and hypotheses for further analysis. On the other hand, exploratory analysis requires high interactivity with analytic systems [14]. When the data is big, such an exploration could be, if not impossible, impractical without progressive analytics where the data is processed in chunks.

To facilitate the exploratory analysis environment, we considered the generation of credit card transaction *subsegments* both automatically and manually so that the analysts could form a consolidated segment from those subsegments — an effective approach enabling human intervention. The high-level tasks can be designed in a way that subsegments are generated at the end of a clustering process and/or through human vision system, combined together by applying union or intersection operations, and finally fine-tuned by filtering and further selections. The final product of such a workflow could be a user-defined segment with a description that can be compared with the rest of the data. For example, an automated clustering algorithm could form a number of segments that might need to be alleviated by the analyst's own mental model of the phenomenon and experience. Not all features (e.g. categorical, location, time stamp) of the dataset could be directly used in the clustering algorithms. This constraint led to a solution with a selectable set of features based on which *automated* and *user-defined* subsegment generation could be performed. With these ideas in mind, we identified the tasks for transaction data exploration as follows. We provide mappings from our tasks to the previously published task taxonomies by Yi et al. [9] in Appendix-B.

T1: *Automatic Feature-based Subsegment (AFS) Generation.* Automation of subsegment generation based on selected features helps to identify underlying groupings in a dataset. We facilitate this through the online clustering algorithm (Section 4.2.1.2) and users can select the elements of the cluster by simply highlighting one of the small multiples (Figure 4.19-b). To support the characterization of the selected subset of items users refer to the *difference view* (Figure 4.19-g).

T2: *User-defined Subsegment (UDS) Definition.* Subsegment detection can also be performed by analyst intervention. To support this, we employ the online PCA computation module from Section 4.2.1.2. Depending on the discriminating power of the selected features, PCA results could lead to visually identifiable groupings (Figure 4.19-c).

T3: *Segment Composition.* Composed segments could be formed through combination (e.g. union or intersection) and refinement of subsegments (Figure 4.19-d).

T4: *Segment Fine-tuning.* Interactive brushing enables further modifications on the combined segment. Such selections could be performed at any level of data granularity from single transactions to a set points forming a cluster (Figure 4.19-e). For instance, composed segments could be modified by making further selections on a map, or a histogram (Figure 4.19(f)).

T5: *Composed Segment Description.* Composed segments can be further modified through peripheral visualizations by applying selection and filtering operations. Composed and fine-tuned segments, from a theoretical point of view, should represent a particular group of transactions with properties that the user (i.e. analyst) has in his/her mental model of the corresponding subset of customers. For example, a segment of transactions mainly generated around the capital city might also be the one generated by young and high-salaried customers. However, such inferences can be made via highly interactive visualizations that can compare the selection and the rest of the data immediately, hence the need for progressive visualization.

4.2.3.2 Exploratory Analysis Workflow

An example workflow (from one of our sessions) with the aforementioned tasks is shown in Figure 4.19. In this particular instance, a progressive cluster small multiples view for AFS generation (**T1**), a progressive PCA view for UDS definition (**T2**) were utilized. Here we start the analysis with a manually determined subset of the features (Fig. 4.19-a) — EFT features, mean transfer, and acceptance rate) to trigger PCA and clustering computations (Fig. 4.19-b,c). After a few iterations, analysts were able to view the first approximate clusters and PCA results. As these views were progressively being updated at each step, analysts made numerous subsegment combinations through selection. In response to such subsegment selections (i.e. clusters or selections on PCA plot), DimXplorer seamlessly composes (**T3**) (Fig. 4.19-d) and presents the combined segment (Fig. 4.19-f) and the segment description (Fig. 4.19-g) (**T5**).

At a particular moment, analysts noticed that the combination of two clusters presented a particular customer segment with customers working with very few other banks (i.e., low EFT entropy) and had low credit card limits. Those customers were also interested in offers and campaigns as their acceptance rates were significantly higher than the rest of the customers (Fig. 4.19-g). Analysts further noticed that the transactions belonging to the segment were also plotted mainly as outliers (Fig. 4.19-c). In order to base the model to a larger set of transactions, they formed two additional subsegments by selecting all the outliers on the PCA view. New selections slightly moderated the EFT entropy and credit card limits resulting in a less clear composite segment picture. After a number of fine-tuning operations (**T4**), they discovered that adding the transactions made around the airport (Fig. 4.19-e) clarified the description of the segment considerably different from rest of the data. The updated version of the composite segment was including the customers that were not only different with respect to acceptance rate, but also associated with a low profile in the mobile banking service usage — making them good candidates for further credit card offers.

4.2.3.3 Observations and Discussions

The case study was designed with the aim of validating the progressive design guidelines rather than the capability of the tool in segmentation. Hence, more emphasis was put on the behavioral observations in relation to the design recom-

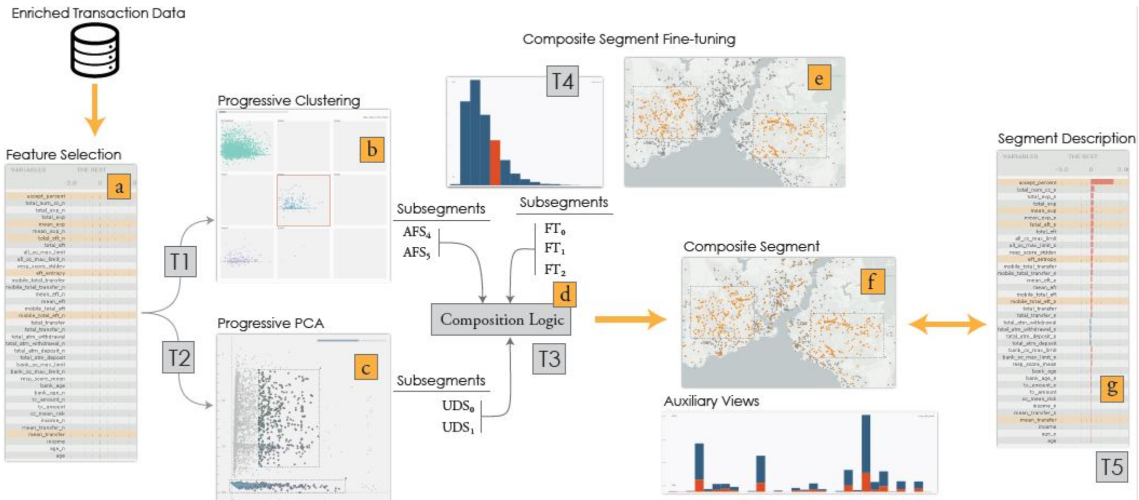

Figure 4.19: An example analysis workflow. In response to a selection of features (a), the *Clustering Small Multiples View* (b) and the *PCA View* (c) progressively visualize the results at each iteration of the calculations while users can renounce any ongoing calculation by selecting a new feature set from the difference view (a). During this process, subsegments are generated in automatic feature-based (AFS) or user-defined (UDS) manner (e.g. AFS_4, AFS_5, UDS_0, and AFS_1). Subsegments are combined (d) based on the composition logic defined by the analysts. Further refinements are applied to the combined segment by brushing on auxiliary views (e) (e.g., exclude the smaller transactions). The resulting segment can be viewed over a map (f), and the data subset corresponding to the composite segment is compared with the whole dataset in the difference view (g).

mendations (DR) discussed in Section 4.2.1. A selection of relevant observations are listed in Table 4.2 and the text refers to them accordingly (e.g. Qu-1).

Human-time constants—Here we report the relevant statements on how the human-time constants (DR1) as an underlying temporal control mechanism were received. Analyst-1 stated that they can "*generate so many new hypotheses in a very short time without waiting for the whole calculation to end,*" and further stated that the "*visualization is quite engaging as we don't have to wait for even a moment to get some initial results*" (Qu-1). As listed in Table 3 in the Appendix, we observed an overall increase in the number of insights made and questions asked per session during progressive analysis sessions compared to the non-progressive one (i.e., the final session).

Update on demand— The update rate of the analytic model and visualization is an important parameter that needs to be adjusted according to the requirements of the application. Analyst-3 stated that ever-changing cluster view frequently interrupted his ongoing explanations about the findings during the course of the analysis (Qu-9). Particularly, fluctuations that occasionally happened during the calculations caused confusion and hesitation, which led them to offer a new feature to DimXplorer (Qu-11) so that they were able to modify the size of the data chunks processed at each step. This was mainly due to the fact that they found it quite cumbersome to discuss on a non-steady visualization (Qu-12). Clearly, we were experiencing the trade-off between *quality-* and *quantity-first strategies* [132].

Progression—We observed that the availability of the progress made during the calculations impacts the renounce decisions of analysts. During the second analysis session, Analyst-3 warned his colleagues that the progressive clustering algorithm was about to consume all the data while they were arguing about a spending pattern of a customer segment (Qu-3). He suggested starting a new round of analysis with a new set of attributes as he was quite confident that the result of the ongoing calculation would not significantly change upon the consumption of the whole data (DR8). During the second and third sessions, we observed that the analysts tried 14 different feature combinations for clustering, and 13 of the times they did not wait until to the end of the calculations (Table 4.3).

The use of data completion rate as an indicator of the progression occasionally caused confusion while making renounce decisions (DR9). Not surprisingly, while deciding to give up on a feature set, the analysts relied mostly on the stabilization of the progressive views (i.e. clustering and PCA). Early and temporary stabilization could also be another problem as it can beguile the analysts into making premature judgments. We observed that analysts made early decisions about the patterns in the data and behaved more conservative in making judgments towards the end of the case study (Qu-7 and Qu-9). Occasionally, they suggested each other to wait until the computations *"settle down"* to some extent. Another interesting observation relates to the cases where there are multiple active progressive views. In certain cases, it is likely that different views reach stability at different progression levels, which made the analysts hesitant at early convergence levels. Such visual and computational fluctuations can lead to loss of trust, however, showing the likelihood of error or a quality metric could mitigate the effects of such fluctuations.

Time vs. uncertainty—The level of the confidence with the current *unfinished* result depends partially on analysts' judgments, since the current progress indication is purely data based. Moreover, momentary state of the analytic process and the sampled data (e.g. the discrimination capability of the selected feature set) seem to be additional factors on decision of when to renounce an ongoing analytic process. This was inline with our observation that analysts gave up on feature sets at various moments of the progression. Even with the same selected feature set, the analysts halted the ongoing calculations at different data completion levels. Table 4.3 shows

Data Contributed	**%0-25**	**%26-50**	**%51-75**	**%76-100**
Renounce Counts	6	5	2	1

Table 4.3: Number of renounces made by the analysts with respect to the percentage of data contributed to the calculations.

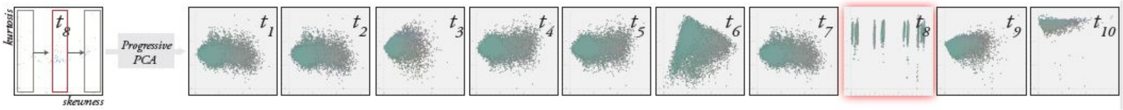

Figure 4.20: An analysis of the 77 dimensional protein homology dataset. The aim is to find dimensions that carry structured information. A keyframed brush is generated through a *skewness* vs. *kurtosis* plot (the dimensions have similar kurtosis but varying skewness). The resulting brush sequence traverses from the left-skewed dimensions to the right-skewed ones. When the resulting animation of PCA results is observed (only the key frames here), a strong structure is spotted at key frame t_8. On close inspection, the structures are found to be due to a categorical dimension.

the renounce counts with respect to the moments they abandoned the continuing calculations.

During the analysis sessions we observed that analysts made use of every moment of the progression by discussing and reasoning, instead of waiting idly. There have been cases where analysts made claims, and, during the same course of calculation, rejected them. For example, Analyst-3 pointed to a relationship between the total number of credit cards of the customers and the EFT entropy after roughly 30% of the data had been consumed by the clustering algorithm (I-3). However, after a period, he rejected his own claim due to the new shape of the clustering calculated with more than 60% of the data (Te-1).

Interactions supporting progressiveness—We observed that there seems to be several design decisions to better support progressive visual analysis. Most of the interaction issues we observed seem to be due to the longstanding interaction habits acquired from traditional non-progressive tools, i.e., the typical linking & brushing behavior. Analysts frequently asked whether changing an aspect in one of the views will affect the calculations (i.e., restart or continue), we postulate that visualization systems need to incorporate effective affordances for guiding the users in complex use scenarios such as the ones described here.

Interaction design decisions seem to be of great importance while orchestrating the progressions of multiple views, and while providing intuitive usability to guarantee progression abstraction from various tasks such as selection and filtering (DR6).

4.2.4 Discussions and Limitations

Due to the conflicting reports on the successful utilization of animations [123], [150], we carefully consider our design choice related to the use of animated transitions. In our visualization approach, PCA projection results carry spatial characteristics, i.e., have a meaningful mapping to the x and y coordinates and all the changes between different projections happen within this spatial mapping — thus meeting the *congruence* principle [123]. Also through functionalities such as pausing or looping we aim to satisfy the *apprehension* principle [123].

Although online learning approaches demonstrate time and memory efficiency when dealing with large datasets, compared to offline versions, i.e., algorithms that process the whole data in a complete *batch*, online algorithms can lead to inaccurate results and might suffer from overfitting to the data that has been processed. Such problems can be tackled via error-bounded methods [133] and as touched upon in several parts of the text through more effective sampling strategies. In our current approach, we employ a random sampling strategy with a focus on maintaining the pace of interaction. This approach can potentially lead to further fluctuations (which are observed to be disruptive for analysts) in the computations since the underlying structure of the data is not considered. Our numerical evaluation also revealed that for datasets with strong structures tend to consolidate with larger portions of the data, further supporting this issue. Such limitations can be addressed by incorporating sampling methods that are not "blind" to the structures in the underlying data. Related literature from data mining and machine learning domains offer advanced alternatives [163], [164]. One promising future work to mention here is investigating the impacts of such sampling methods on progressive analysis sessions.

Depending on the task and type of computational tool that needs to be employed, there are incremental versions of different algorithms in literature, e.g., classification, regression [163]. However, there are tasks not suitable to approach with online algorithms.

In such cases, it is advisable for visualization designers to focus on improving the performance of the system, using methods such as pre-computing or caching, to maintain interactivity within the limits of human time constants as also evidenced in recent work focused on such capability [130], [147].

We adhere mainly to the temporal constraints of the users. However, there might be cases where quality of the results are of critical importance for the analysis scenario. In such analysis situations, one might also consider accuracy/confidence thresholds that can be incorporated via building *error profiles* (similar to the "error latency profiles" generated for database query accuracy [133]) for computational methods to estimate accuracy boundaries.

During the case study sessions, we observed that when different online computational methods are incorporated concurrently, it is likely that these methods have variations in their progression. This was, at times, confusing for the analysts. Our current prototype had no mechanism to check and correct for that. However, we consider this observation as a pointer for future work and a sign that there are several aspects to investigate in the use of these tools.

At the end of the case study, analysts reported that they had experienced higher level of engagement during the analysis compared to their existing working setup. With the leverage of progressiveness, high engagement with, and escalated throughput of the analytical processes seem to require new ways of thinking about the usability of the exploratory analysis tools. High number of insights and interesting clues extracted from the data might be lost during the analysis sessions – calling for advanced provenance capabilities. One of our interesting observations involved an inference and rejection of an insight that was done in almost 20 minutes (I-3 and Te-1). Due to the lack of provenance features, they were not able to regenerate the case.

One important limitation in progressive analysis sessions is to determine when to renounce and pave the analysis to a new direction. The observations that analysts can occasionally draw insights and reject them later on signals a drawback: the time saved by the application of progressive analytics could be lost while trying to amend the consequences of incorrect inferences. This risk could be mitigated by employing more appropriate data sampling mechanisms and progression indicators. However, the renounce decision still requires human intervention, which is likely to require the development of domain-specific heuristics to support more effective decision making. One additional perspective to mention is the need for appropriate training and familiarity with progressive methods. Although we observed that fluctuations in the results can lead to confusions, the analysts can learn to interpret those as pointers to underlying structures, however this requires the prolonged use of such tools.

Traditional mechanisms such as "undo" or "redo" should also be considered in the design of progressive visual analytic tools together with mechanisms to support provenance. Such a mechanism might utilize the notion of *checkpoints* which could be considered as screenshot of particular states of the analysis so that interesting findings could be re-generated for secondary discussions or presentations. One of our interesting observations involved an inference and rejection of an insight that was done in almost 20 minutes (I-3 and T-1). Due to the lack of the aforementioned capability in DimXplorer, they were not able to regenerate context and present how they rejected the insight.

In this study we discuss how an established cognitive model of human-computer interaction [15] can be placed as the underlying mechanism to determine the pace of interaction in approaches where the analysis happens through the successful facilitation of the dialogue between the analyst and the computer, in particular those that involve high-dimensional data sets. We present how suitable computational methods that can perform "progressively" can be integrated to operate at the temporal frame set by these underlying constraints. To better facilitate the dialogue with these computational methods, we suggest a series of interaction and visualization techniques and externalize our reasoning in a series of design recommendations.

In order to understand and evaluate how our approach facilitates better analysis, we carried out a series of analyses with a group of financial data analysts. Upon working on our progressive visual analytics approach through a prototype, analysts reported increased levels of engagement during the analysis sessions, thus leading to a higher number of observations made and hypotheses tested. As inline with the findings of our case study, progressiveness facilitates early acceptance or rejections of hypotheses, and makes the testing of several computational models feasible which would otherwise take long time and effort. We observed that progressive analytics has shown to be a powerful facilitator of engagement within the exploratory analysis of high-dimensional data. However, we further observed that progressiveness could be misleading under particular circumstances and should be utilized carefully. Adoption of domain specific metrics and the design of visualizations to better inform analysts on the progress has the potential to address these limitations calling for further research through targeted design studies involving progressive methods.

4.3 Observation Data

Study 3 and 4 have been evaluated with controlled laboratory experiment and case study, respectively. For both of the studies, Likert-type scales and semi-structured interviews are applied in order to collect subjective measures, and system logs have been used to collect interaction data. Analysis sessions of both studies have been video recorded. Observation data for these studies were extracted from these objective, subjective, and behavioral measures.

4.3.1 Objective Data

Based on the methodologies described in Sections 3.1.4 and 3.2.2, objective data has been extracted from the case study and laboratory experiment sessions. Due to the methodological differences in the evaluation of Study-3 and 4, the processings of the recorded material from the analysis sessions had slight variations. Controlled laboratory experiment material was transcribed for five randomly chosen experiment session. Extracted observation data and their source material are as listed in Table 4.4.

The calculated analytic system properties are shown in Table 4.5. It should be noted that both analytic systems are mainly desktop application with no multiuser support, hence the communication media level is 1. The analytic system reported in Study-4 is a progressive visual analytic system, hence it has a value greater than 0 unlike the case for the analytic system in Study-3.

Table 4.4: Study-wise data sources for interaction characteristic data calculation. Objective data sources for each analytic system study is listed. *Log* corresponds to the interaction logs on the data side, *video* is the records of the sessions, *expert review* is the data extraction process carried out by analytic system's expert, and *heuristics* is a predefined procedure for the calculation of the corresponding data set.

Characteristic	System 3	System 4
Human-side Message Timestamps	Log + Video	Log + Video
Human-side Message Types	Log + Expert Review	Log + Expert Review
Media	Expert Review	Expert Review
Medium Physicality Constants	Expert Review	Expert Review
Data-side Message Timestamps	Log	Log
Data-side Message Types	Log + Heuristics	Log + Heuristics
Data Consumption Rates	Log	Log

Progressive analytic systems process available data in pieces, and such data processing results in regularly sent data-side messages. Each data-side message sends some form of data representation by consuming a fraction of the whole data. Based on the timedelta between those messages and the fraction of consumed data is used to calculate a progressiveness level as explained in Section 2.5.2. The progressive visual analytics system reported in Study-4 employs an adaptive sampling algorithm that enables the update of data representations in one-second intervals. Due to this reason, the progressiveness of the mentioned system is significantly regular, and hence, its progressiveness level is close to 1.

Table 4.5: Analytic system property values for systems 3 and 4. Property values for the analytic systems are calculated based on the interaction characteristics data. These values forms the objective part of the observation data.

Property	System 3	System 4
Responsiveness	0.88	0.81
Communication Media Level	1	1
Unit Task Diversity	0.76	0.95
Human-side Message Closeness Factor	1.34	2.65
Progressiveness Level	0	0.86

4.3.2 Behavioral Data

Evaluation of Study-3 was conducted as a controlled laboratory experiment whose each session was conduct one participant whilst that of Study-4 was a case study with 4 analysis session with a team of analyst. Due to this distinction in the evaluation methodologies, Study-3 and 4 have different behavioral data sets. Behavioral data transcribed during the analysis sessions are as follows:

- *Directed technical communication.* Questions or statements made by the users of the systems that are not directly related to the analytic task at hand. These behaviors involve analytic system interaction requests of peers such as context or mode change, object detection. This behavioral data could be transcribed for only Study-4.

- *Directed task-related communication.* Questions or statements made by the users of the systems that are directly related to the analytic task. This communication pattern involves the statements that could be attributed as task completion efforts. For example, questions asked to other peers such as "Do you think young customers spend more during weekend in region X?" can be considered in this behavior category. This behavioral data could be transcribed for only Study-4.

- *Hypothesis, insight, and question statements.* Any form of statement that could be qualitatively judged as a hypothesis, insight statement or research question are considered in this category.

- *Idling communication.* Any form of communication that is not made for technical or task-related reasons are considered in this category. The moments where the participants checked their personal belongings such as cell phones or when they turned their gaze away from the computer screen in Study-3, this behaviors are also considered as idling communication.

- *Bodily gesture towards peers.* Bodily gestures that are directed toward the analytic system's interface or peers are considered in this category.

4.3.3 Subjective Data

Subjective data which is used for the validation of the findings that are reported in Section 4.4 comprises self-reports of the participants of the data analytic studies and are listed as follows:

- *Pre-study self-reports.* Participants of the Study-3 reported their personal background in the usage of analytic or information visualization systems in order to take the control of the experience as a confounding variable. Users of the analytic system reported in Study-4 were expected to describe their analysis strategies for the given tasks prior to the case study sessions.

- *Post-study self-reports.* Post-study scales and semi-structured interviews was conducted in order to measure the task-tool match quality from the users' perspective. Users were asked to express their own considerations on how successful they felt in completing their tasks, and collaborating with their peers.
- *Assisted self-reports.* Moments of interests have been cropped from the recorded video, and were shown to the users in order to get deeper explanation on the root cause of their behavior or interactions with the tool.

4.4 Realm Characterization

Characterization of the realm benefited from the behavioral and objective data collected during the conduct of the studies reported in this chapter. Except for the pre-study self report for Study-3, subjective data was used to verify the characterization.

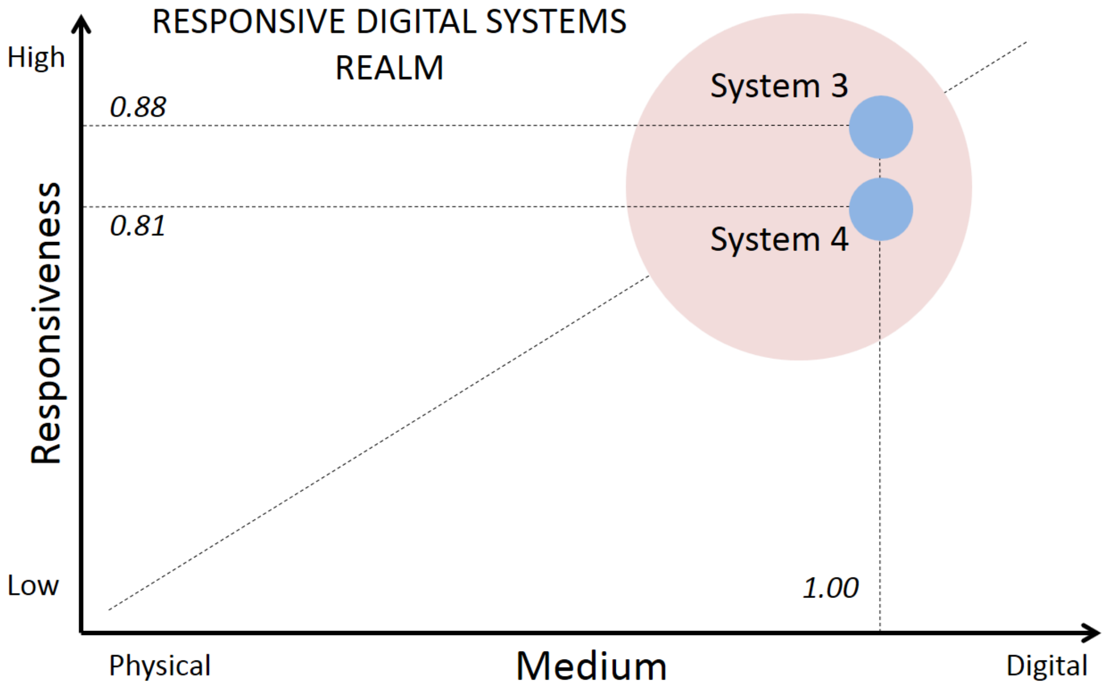

Figure 4.21: Responsive digital systems realm. Plot of system 3 and 4 on the exploration space of HDI is represented.

Properties of analytic systems 3 and 4 are as listed in Table 4.5. Based on the responsiveness and communication media level values, these two analytic systems are positioned in high digital and high responsive region of the exploration space as can be seen in Figure 4.21. As both of the systems are desktop applications, they are considered completely digital and their communication media level are evaluated as 1. These applications are designed in order to facilitate interactive exploratory analysis of spatio-temporal data, and hence their responsiveness scores are evaluated close to 1. Please note that system 4 is evaluated as less responsive. This is mainly due to the fact that system 4 is a progressive analytic system, and the update of the data representations with the data-side messages are tuned so that update happens approximately once in a second.

Unit task diversity for both of the systems is high, as expected, due to the fact that both systems support view of the data from various perspectives such as rotate, zoom in or out, and filter on demand. However, diversity is higher for system 4 which is highly likely due to the system's support for concurrent multiple data views management making the task selection more chaotic.

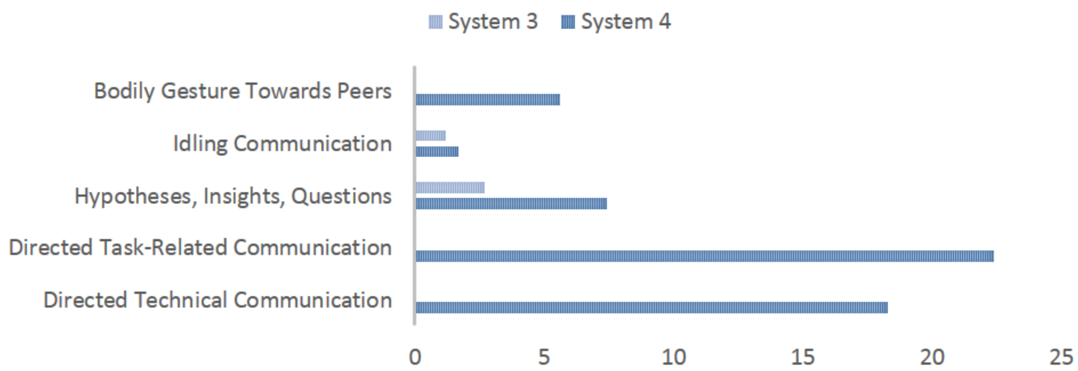

Figure 4.22: Behavioral data summary. Behavioral data collected during the case studies of analytic system 3 and 4. The values represent the mean counts per hour.

On the other hand, human-side message closeness factor is evaluated less for system 3 meaning that the users of this system sent their messages more frequently compared to the ones sent in system 4. High closeness factor value for system 4 stems from the progressive nature of the analytic system: Users typically wait for a while after sending a message, i.e. making interaction, in order to see the change trend in the data representation. In some cases, this contemplation lasts for a long while until the user of the system is persuaded that "the view is settled down". Such waits seem to increase the closeness factor.

Responsive digital systems realm seems to be the set of solutions for the exploration tasks requiring unit tasks to be completed in brief amounts of time. Furthermore, collaborative effort could also be supported in these systems, however, with more effort on the human side. As can be seen in Figure 4.22, users of the system had to communicate with their peers around 18 times for technical reasons on average. In case of technical communication to the peers, users had to request for a particular change on the current view of the data. Similarly, in order to collectively develop better understanding of a particular case of analysis activity, they had to communicate with their peers as many as 22 times per hour on average. I suggest that main reason for high amount of communication is related to the lack of physicality, and hence the need for more verbal conveyance of the ideas to the peers.

Users of the system 4 reported that the analytic system was a good fit for their tasks presented in the analysis sessions. Furthermore, they found the tool engaging and inspiring for new ideas and hypotheses. This was also inline with the high counts for the hypotheses, insights, and questions as shown in Figure 4.22. Low idling communication times implies the high engagement associated with the analytic systems of responsive digital systems realm.

CHAPTER 5

UNRESPONSIVE DIGITAL SYSTEMS REALM

High responsiveness of an analytic system facilitates engagement, catalysts hypothesis and inference generation during the analysis sessions, however, there might be cases where responsiveness is not attributed as the top priority. In such cases, the goal of the analysis could be to achieve some form of result based on specific methodology and data, and systems typically execute batched predefined analytic execution steps without need for human intervention. For such tasks, commonplace solutions adopt highly digital and, depending on the data size, lowly responsive analytic systems.

The fact that these systems are typically lowly responsive does not necessarily mean that none of the human-side messages get data-side response in short time. There are, indeed, responsive moments during the interactions with systems in unresponsive digital systems realm such as data cleaning or similar other preprocessing action. Nevertheless, responsive side of the analysis sessions with these systems are not meant to foster creative collaborative analysis sessions unlike it is the case for other realms.

The systems residing in this realm could be regarded as the test bed for the ideas generated during the highly responsive and interactive sessions performed with systems in other realms. Based on the subjective data collected during the reported study, the major problem with these systems seems to be the low engagement and cost of context switches occurring during the time waited for system response.

5.1 Study-5: Behavioral Attributes and Financial Churn Prediction

Customer retention is crucial in a variety of businesses as acquiring new customers is often more costly than keeping the current ones. As a consequence, churn prediction has attracted great attention from both the business and academic worlds. Traditional efforts mainly focus on domain specific variables, however, without considering behavioral and decision-making patterns of customers. In this study, we attempt to fill in this gap by investigating the spatio-temporal and choice patterns underlying the customers' financial decisions and their relations to customer churning activities. Inspired by previous works in the emerging field of computational social science, we built a prediction model based on spatio-temporal and choice behavioral traits using individual transaction records. Our results show that the behavior-based models could predict churn decisions significantly better than traditionally considered factors such as demographic-based features, and that this effect remains consistent across multiple data sets and various churn definitions. We further study the relative importance of the various behavioral features in churn prediction, and how the predictive power varies across different demographic groups. More generally, the proposed features can also be applied to churn prediction in other domains where spatio-temporal behavioral data are available.

We live in the era of "Big Data". The recent availability of large-scale quantitative behavioral data provides opportunity to study human and social behavior at an unprecedented scale, leading to the emerging field of computational social science [165]. The breadcrumbs that we leave behind with everyday activities seem to reveal more about our conscious behaviors and decisions than our socio-demographic characteristics. For example, only little information about when and where we make purchases can predict our financial well-being [25], and the implicit patterns in our communication networks can determine our performance at both individual and group levels [166]. Despite the promising results of these efforts, however, current business analytics solutions and the related body of research still seem to lack consideration of customer behavior.

In many businesses, predicting the next set of customer actions is of central importance as this capability enables companies to forestall undesirable decisions of the customers. Among those, *churn prediction* has attracted increasing attention from a variety of businesses such as telecommunication and banking industries as well as researchers in academia, as retaining customers is far less costly than acquiring new ones [167]. The cost of making a new customer as profitable as a current customer could be up to 16 times higher than the cost of retaining efforts [168], and decreasing the churn rate by only 5% can increase the profitability by 25-125% [169]. According to a survey carried out with over 24 thousand customers in 33 countries [170], 68% of the churning customers expressed that they would not do business again with the companies that they left. The cost of such provider switches of customers is estimated to be $1.6 trillion for the United States. However, predicting whether the customer will quit his or her contract seems to be a rather daunting task, due to the unpredictable nature of active decisions of customers, such as quitting contract due to unsatisfactory service [171], and incidental or non-voluntary events, such as change of home/work locations or financial troubles [172]. Complexity becomes even worse for the financial churn prediction due to the relative sparsity of the transactions compared to other domains such as telecommunication.

Furthermore, the financial decisions might require longer investigation periods (e.g., loan) leading to the development of heuristics for the churn prevention efforts rather than prediction models based on transactional data.

The problem of churn prediction has been tackled in many different domains such as telecommunication [173]–[178], banking [179], [180], subscription services [181], game businesses [182], and retailing [183]. In general, most of the efforts in churn studies involve prediction with different definitions of the churn event [177], [184], evaluating new data mining algorithms [178], [185], [186], and introducing ways to deal with large volumes of data [187]. In these studies, churn prediction has usually been considered as a benchmark against which novel analytic approaches are evaluated, and domain-specific historical actions of customers are typically involved as features for prediction. For example, such features may include aggregation of service usage rate for a given time frame or static information such as socio-demographical variables. Nevertheless, historical actions in terms of service usage might bear an egg-chicken problem such that the churners are intuitively expected to reduce their service usage and eventually quit towards the time instance when they churn and vice versa, while the limitation of demographic-based variables is that they are static one-time information, and prone to manipulation and error.

Recent studies suggest that behavioral traits in our everyday activities may better explain the phenomenon under investigation. For example, diversity of phone communication or interaction within social networks has been shown as a strong indicator of economic development of communities [188] and financial status of individuals [189]. Behavioral traits in customers' daily purchases, which are computed based on individual financial transaction data, can predict financial well-being of the customers significantly better than demographic features [25]. These findings highlight the potential of patterns behind social interaction (e.g., phone calls, face-to-face meetings) and decision-making (e.g., expenditures) in financial outcome prediction.

In this study, we investigate whether behavioral features could be utilized in the prediction of churn decisions of bank customers. To this end, based on individual credit card transaction records of a large set of customers, we develop spatio-temporal behavioral features, namely diversity, loyalty, and regularity, that were introduced by Singh et al. [25] with minor modifications. We also introduce a set of novel *choice* features which reflect the behavior of customers when they make financial choices such as the merchants to purchase from, shopping categories, and the addressee of the fund transfers.

Our findings suggest that the proposed spatio-temporal and choice behavior features are significantly better than demographic features in bank customer churn prediction. We report that diversity and regularity in customers' spatio-temporal activities and financial choices are more important than other behavioral traits in financial churn prediction. In particular, the exploration levels of the eventual churners, reflected by the diversity patterns, had a decreasing trend during the observation window, while this was not the case for non-churners. This seems to suggest that deviations from a customer's usual spending patterns could be indicators of the presence of the financial stress that leads to his churn decision, in a way similar to the behavioral changes of living organisms as a reaction to persistent stress [190]. Furthermore, we conducted the same study for different demographic groups and found out that churn prediction seems to be relatively easier for the group of younger customers, while gender-based difference is not significant. These findings also remain consistent over a number of different data sets that are generated based on different sampling strategies and observation windows.

This study contributes to a growing body of research on data-driven behavior understanding, and in particular provides novel insights into the understanding of the challenging problem of financial churn prediction. This would help decision-makers to go beyond traditional demographic-based variables and analyze customer churning from a behavioral perspective, not only in banking industries but also in a variety of businesses. For example, the spatio-temporal and choice models utilized in this study can also be applied to churn prediction in other domains, such as telecommunication industries, where similar information on spatial and temporal activities as well as choice decisions are readily available. In summary, the main contributions of this study can be summarized as follows:

- We show that behavioral features are superior to traditionally considered demographic or activity-based features in financial churn decision prediction.
- We demonstrate the performance comparison of demographic and behavioral features based on stratification of demographic groups, which implies that churning decisions of younger people seem to be more easily predicted.
- We introduce choice features characterizing the behavioral patterns in selecting products, merchants or transfer addressees. Moreover, we analyze the relative performance of each behavioral feature to investigate feature importance.
- We contribute by introducing novel financial data and churn definitions.

5.1.1 Materials and Methods

5.1.1.1 Data

A major financial institution in an OECD country donated two de-identified samplings of their data that were collected over the period between July 2014 and July 2015. The samples comprise demographic information, credit card transactions, money transfers, and electronic fund transfers (EFT) of over 100 thousand (Sample-A) and 60 thousand (Sample-B) customers. Sample-A and Sample-B contain in total roughly 45 millions and 22 millions of transactions, respectively. Both samples were drawn from a much larger sampling of 450 thousand customers who were located in a major metropolitan city, updated their home and work addresses since January 2012, and made at least one credit card transaction during the sampling period. Sample-A was drawn randomly from this larger set whereas Sample-B was drawn from the same set such that each customer has at least 10 credit card transactions, and in total around 60% of all the credit card transactions were performed with point of sale (POS) machines of the bank donating the data. Customers may prefer to use their credit cards on the POS machines of other banks. In that case, some part of the transaction information such as location cannot be collected by the bank issuing the credit card. The bank also donated monthly segmentation information for each customer. We further elaborate on the segmentation information and the way we utilize it for label generation in the Labeling subsection.

The customer transactions and demographic information were anonymized by the bank officials by masking the unique identifier and names of the customers. For each customer, a pseudo-unique identifier has been generated for cross-referencing of different transaction sets. The credit card transactions of customers with missing home and work location information, and the transactions without merchant or location information were also not included in the calculations.

While the samplings (A and B) of the customer transactions were collected over a 12-month period, the customer segment information covers a 23-month period including the sampling period such that the segmentation information of the customers were available for an additional five-month period following the sampling window. This additional information enabled us to generate two variants of each of the samplings by defining 12- and 9-month observation windows for feature extraction, and 5- and 3-month churn decision windows for label (churner or non-churner) extraction. This translates into four distinct data sets, namely data sets A1 and A2 generated from Sample-A, and data sets B1 and B2 derived from Sample-B. Please see Table 5.1 for the characteristics of the data sets.

5.1.1.2 Features

In order to establish the relationship between customers' churn decisions and their behavior patterns, we have prepared demographic and behavioral features from the donated data set. For the characterization of the behavioral features, we employ a slight variant of the pattern extraction technique described in [25], and in addition, we introduce a new set of features, namely *choice* behaviors, that explains the expenditure and transfer tendencies of the customers.

Table 5.1: Data set characteristics.

Data Set	Source	# of TXs	# of Cust.	Label Sets	Churn(%)
$A1$	Sample A	8.5M / 3.3M	55K	SB	1.97
$A2$	Sample A	6.3M / 2.4M	53K	SB, CC, CA	0.99
$B1$	Sample B	4.2M / 2.6M	43K	SB	2.27
$B2$	Sample B	3.1M / 1.9M	42K	SB, CC, CA	1.42

Based on samples A and B, four data sets with different characteristics have been generated. The summary includes the sampling source of the data set, count of all transactions and the transactions with POS location information (# of TXs), number of customers (# of Cust.), the label sets generated for the related data set, where SB, CC, and CA stand for segmentation-based, credit card usage-based, and checking account usage-based labeling, and finally percentage of churning customers in the data set (Churn(%)) according to label *inac-full*. The transaction and customer counts represent the state after the data filtering process.

Table 5.2: Data set characteristics.

Data Set	Source	Observation Win.	Labeling Win.
$A1$	Sample A	07/2014 - 06/2015	07/2015 - 11/2015
$A2$	Sample A	07/2014 - 03/2015	04/2015 - 06/2015
$B1$	Sample B	07/2014 - 06/2015	07/2015 - 11/2015
$B2$	Sample B	07/2014 - 03/2015	04/2015 - 06/2015

Observation and labeling windows for feature and label generation for each data sets are presented.

Demographic Features—The demographic features of the customers included gender, marital status, educational status, job type, income, and age of the customers. Except for income, all the demographic information of the customers were available. The missing part of the income information (less than 2% of the customers) were filled with the mean income of the rest of the customers.

Spatio-temporal and Choice Patterns—The behavioral features comprise implicit spatio-temporal expenditure patterns and financial choice patterns. Spatio-temporal expenditure patterns, namely *diversity*, *loyalty*, and *regularity*, refer to the measures of how diverse or loyal customers are in their spending patterns from time and location perspectives, whereas financial choice patterns indicate how customers distributed their financial activities (i.e. online/offline credit card purchases, fund transfers) with respect to merchants, spending categories, and the addressees of the fund transfers.

For the spatio-temporal expenditure patterns, our study benefited from the formulations introduced in [25], however, with minor necessary modifications. First, *Diversity* represents the extent of the customers' tendency to make purchases at different locations or times. A high score of diversity means that customer spreads his or her transactions to a large number of *bins*, which can be considered as the slices of time or space, and will be further discussed below. Mathematically, diversity D_i is the normalized entropy of the transactions of customer i with respect to space and time slots (i.e., bins):

$$D_i = -\sum\nolimits_{j=1}^{N} p_{ij} \log_M p_{ij}$$

where p_{ij} is the probability of customer i having transaction in bin j, N is the total number of bins, and M is the number of non-empty bins. Our modification to this calculation was that, as the normalization factor of the entropy, we used the *total* number of bins, rather than the number of *non-empty* bins. The downside of using the number of *non-empty* bins is that it would calculate the same diversity scores, for example, for two customers evenly distributing their transactions into *different* number of bins. With our approach, for the same scenario, the customer evenly spreading her transactions into larger number of bins can get a higher diversity score, as expected.

Second, *Loyalty* is the fraction of the customers' transactions in their k-most frequented bins. If f_i is the total number of the expenditures that happened in the top three bins of the customer, then the loyalty for customer i is calculated as

$$L_i = f_i / \sum\nolimits_{j=1}^{N} p_{ij}$$

Finally, *Regularity* represents the level of the similarity of the customers' diversity and loyalty scores over shorter and longer terms. In principle, it is one-complement of the mean Euclidean distances between shorter and longer term score vectors for diversity and loyalty. Regularity is calculated as

$$R_i = 1 - \sqrt{((D_i^S - D_i^L)^2 + (L_i^S - L_i^L)^2)/2}$$

where D_i^S and D_i^L stand for shorter and longer term diversity, respectively. Likewise, L_i^S and L_i^L stand for shorter and longer term loyalty. Regularity scores closer to 1 represent higher regularity indicating having similar diversity and loyalty scores in shorter and longer term periods. For our study, the duration of the shorter term is selected as the one third of the observation window (Please see Table 5.2 for the observation windows.).

The *bins* mentioned in the calculation of spatio-temporal features can be considered as the slices of time or space. In this study, the space is organized as a collection of square grids (grid bins), and concentric annular areas centered at home and work addresses of the customers (radial-home and radial-work bins). We selected the edge size of the square bins to be 0.1 degree units, and radii of the annular areas to be 0.5, 1, 2, 3, 4, 5, 10, 15, 30, 50, 100, 150, 300, and 500 kilometers. Temporal hourly and weekly bins are the hour of the day and day of the week of a given transaction, respectively. Hence, there are in total 24 temporal hourly and 7 weekly bins.

Three spatio-temporal behavioral traits with five different spatial and temporal variants translates into a set of 15 behavioral features whose names we abbreviated for better readability as follows. For the three behavioral traits diversity, loyalty, and regularity, we used the prefixes *div-*, *loy-*, and *reg-*, respectively. Similarly, for the five bin variants grid, radial-home, radial-work, hourly, and weekly, we considered the suffixes *-g*, *-rh*, *-rw*, *-ho*, and *-we*, respectively. For example, given the *loyalty* trait and the *grid* variant, we abbreviated the feature *grid-based loyalty* as *loyg*.

In addition to the features merely based on temporal and spatial patterns, we introduce the *choice* behavioral trait representing the variety of the selections that the customers make in their purchase or fund transfer behavior. In other words, it is the entropy of their choices of *products* when shopping or *peers* when making transfers, and calculated as follows:

$$C_i = -\sum_{j=1}^{N} p_{ij} \log_M p_{ij}$$

where p_{ij} is the probability of customer i making selection j, N is the number of all possible distinct selections, and M is the number of unique selections customer i made. For example, the set of the selections for a particular customer would be all other banks that she has ever made money transfers to, or all the merchants from which she has ever made purchases. Choice behavior features are money transfer entropy (*transe*), EFT entropy (*efte*), entropy of credit card transactions with respect to merchants (*ecctmer*) and merchant category code (MCC) (*ecctmcc*), and entropy of offline credit card transactions with respect to merchants (*efcctmer*) and MCC (*efcctmcc*).

5.1.1.3 Labeling

There exists numerous definitions of customer churn in the literature, and most of the definitions represent very specific aspect of the customers' activity such as whether they used their credit card for a specific period of time [191], whether they made a call during an observation window [178], as a fuzzy concept description [177], or some other metrics based on sliding window methodology [175]. However, the decision of whether a customer churned or not is quite subjective and is usually defined based on heuristic rules set by the industry officials. We followed the suggestions of the bank officials, and developed a set of churn decision models based on the segmentation information, credit card expenditure patterns, and checking account activities.

The bank donated monthly segmentation information of each customer for a 23-month period, which also covers the time window for the transaction data provided. Monthly customer segmentation is generated by the bank in order to facilitate productive and convenient management of business processes such as advertising and churn prevention. Being one of the more than a dozen of such segments, the segment *inactive* is applied to the months of a customer for which he/she owns no bank products or utilizes his/her products under some predefined aggregated activity level. From a practical point of view and as indicated by the bank officials, the inactive customers represent the set of the least profitable customers of the bank.

By adopting the bank's approach, we developed a set of churn definitions and corresponding labels (*churner* and *non-churner*) for each customer based on a set of rules pertaining to the order of the *inactive* months of each customer. For example, a customer who was tagged as *inactive* for all of the months in labeling window was considered as churner based on the churn definition that we named as *inac-full*. For the rest of such segmentation-based (SB) labels, please refer to Section A.3 of Appendix-A. The segmentation-based labels were generated for all the data sets listed in Table 5.1.

Unlike for data sets A1 and B1, in addition to segmentation information, we were also provided credit card transaction and checking account balance data for the labeling windows of data sets A2 and B2. This enabled us to develop several additional churn definitions based on credit card and checking account usage patterns forming the label sets *credit card usage-* and *checking account usage-based* labels (CC and CA in Table 5.1).

The results reported in this study are based on the churn definition *inac-full*. Other segmentation-based definitions along with the credit card- and checking account usage-based churn definitions were also considered in the context of the study. However, we find that the results do not significantly change as can be seen in the results listed in Appendix-A.

5.1.1.4 Experimental Settings

In order to evaluate the performance of our behavioral features, we performed analyses with four different data sets as discussed in the previous section (Table 5.1). The data sets B1 and B2 are biased more towards higher credit card usage whereas data sets A2 and B2 are generated for shorter term prediction. The same methodology independently applies to each of these data sets. Prior to the generation of predictive models, the missing values for each feature are assigned the mean values, and all numerical features are standardized by removing the mean and scaling the values to unit variance. The categorical variables are one-hot encoded.

We adopted Random Forests [192] as the classification training technique for our study. We trained our classification models with 500 trees and maximum two features per tree. We evaluated our models with stratified 8-fold cross-validation. In order to mitigate the risk of imbalance between the number of churners and non-churners in our data sets (Table 5.1), we applied SVM-SMOTE [193] with the ratio of 0.25, meaning that the minority class is oversampled until its cardinality reached to a quarter of that of the majority class. All the preprocessing and classification implementations were done in Python language mainly with Scikit-learn package [194], and for SVM-SMOTE, the open source package Imbalanced-learn by Lemaître et al. [195] was adopted.

5.1.2 Results

We analyzed two anonymous credit card transaction samplings collected over a one-year period by a major financial institution. We treated each sampling separately to generate four data sets by applying different observation and labeling windows. Around 40-50 thousand customers along with 1.9 to 3.3 million transactions have been considered. The results show that the spatio-temporal and choice behaviors are significantly related with customers' churn decisions. The results remain significant even for 11 different definitions of churn, and various versions of the data sets (e.g., using just weekend transaction data).

Figure 5.1 illustrates the cumulative density functions (CDF) of the diversity, loyalty, regularity, and choice features of the customers for data set A1. The CDF graphs for other data sets are graphs with similar curves as can be seen in Figure A.3.

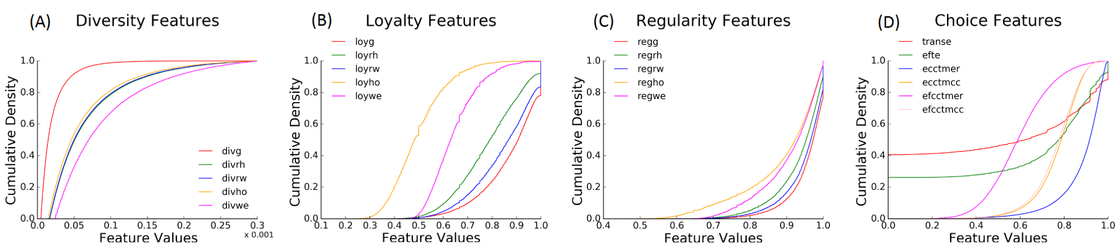

Figure 5.1: Cumulative density functions of the spatio-temporal and choice behavior features.

The noticeable steepness of diversity CDFs (Figure 5.1A) is due to the high normalization denominators that are based on the number of all bins rather than the non-empty ones. It shows that customers were a lot more diverse about where they shop based on grid-based characterization. Figure 5.1B shows that the customers' three most preferred locations account for a large population ($\sim 75-85\%$) of all their shopping. While less than 65% of the purchases were made on the preferred days, around 50% of the transactions were made during the top three time slots. Figure 5.1C shows that customers have similar regularity patterns in terms of both location and time. More than around 80% of the customers seem to have high regularity scores meaning that, both temporally and spatially, they presented similar behaviors in shorter (3-4 months) and longer (9-12 months) terms. The CDFs of choice features are illustrated in Figure 5.1D. Almost identical curves of *ecctmcc* and *efcctmcc* imply that the customers distributed their transactions into different shopping categories very similarly regardless of whether they made their purchases online or in person. More than 80% of the customers distributed their purchases to different merchants with high entropy value as high as 0.8 (*ecctmer*), whereas they showed relatively more deterministic pattern in their offline purchases (*efcctmer*). The money and EFT transactions were not as frequent as credit card transactions in the data sets leading to 0 entropy for around 25% and 40% of customers having very low number of transactions.

Prediction performances of demographic, spatio-temporal and choice behavior features and their combinations are compared in Figure 5.2. Due to high level of imbalance of labels (churners vs. non-churners) in our data sets, we adopted the area under ROC curve metric (AUROC score hereinafter) for the evaluation of our models as suggested by [196], [197]. For all data sets employed in the study, behavioral features were significantly better than the demographic features in terms of area under the ROC curve metric. The results were very similar for the other 10 different definitions of the churn (Please see Figure A.1 and A.2, and Appendix-A for the accuracy scores and significance test results.). The combination of all models were slightly better than behavioral features for each of the data set, however, this superiority was not significant. Similarly, demographic model was better than the baseline model without any significance. For example, for data set A1 (not biased towards credit card users and usage of longer observation window), the behavioral model predicting the churners reached to AUROC score of 77.9% as compared to demographic model which obtained 51.3%, and the baseline model of 50.0%. Hence, the behavioral model performed 55.8% better than the baseline, and 51.9% better than the demographic model for churn prediction in the data set A1. For the data set A2 (not biased towards credit card users and usage of shorter observation window), similar results with slightly higher AUROC scores were observed: Behavioral features were 58.0% and 51.6% better than baseline and demographic models, respectively.

For data sets B1 and B2 (biased towards credit card users), the AUROC scores were higher around 1-2% compared to those of data sets A1 and A2, as expected due to the possible high resolution of the behavioral models that were generated

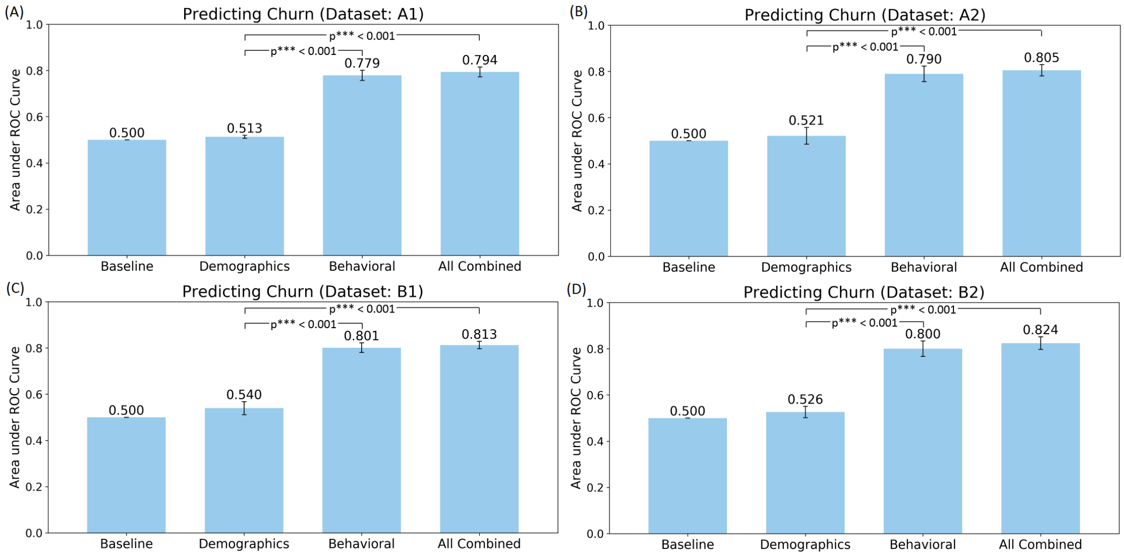

Figure 5.2: Area under ROC curve metric comparison of *demographic* features, *spatio-temporal and choice behavior* features, and the combination of both feature sets for each of the data set versions. The comparison is performed for the label *inac-full*. The length of the error bar corresponds to 1.0 standard deviation. $p***$ denotes the significance level with p value smaller than 0.001.

with higher number of transactions. For all data sets, the AUROC score of the combined model was 1.3-2.4% higher than the behavioral models showing that the demographic features may have a positive effect on the predictive power of the behavioral features. The AUROC scores of the models did not significantly change when the same analyses were conducted with only weekend transaction data.

To understand the individual contributions of the features to the prediction power of the combined model (demographic and behavioral features), we applied mean decrease accuracy analysis for each of the features in each of the data sets. The importance of the features are calculated as the negation of the AUROC score difference of the combined model after having permuted the values of each feature. In other words, for each iteration of the cross validation, the combined model has been trained and tested with the current bag, the AUROC score A_g for the current bag is saved, and then for each feature f_i of the combined model, the column c_i of the test portion of the current bag is randomly permuted, and AUROC score A_i for the modified test data set is re-calculated. Importance I_i for feature f_i is calculated as $(A_g - A_i)/A_g$.

The importance values calculated for each feature in each of the data sets and data set-wise mean of the scores are plotted in Figure 5.3. Sorted by the mean importance values, the plot shows that the importance values vary roughly in the range of -0.01 and 0.03, and the importance of the features follows similar patterns in each of the data sets meaning that high-scored features in the mean score column tend to be high in other data sets as well and vice versa. Clearly, all of the diversity features are more important than the rest of the features, and except for a few demographic features in between, regularity and loyalty features are following. The choice behavior features are distributed to the different positions along the ordered list. It is notable that the educational statuses *bachelors*, *high*, and *middle school* were found to be important in this respective order, implying that as the customers' education level increase their churn decision might be more easily predicted. However, this inference is valid only when the customers with educational statuses *illiteral*, *college*, and *masters*, comprising about 12-13% of all the customers, are not considered.

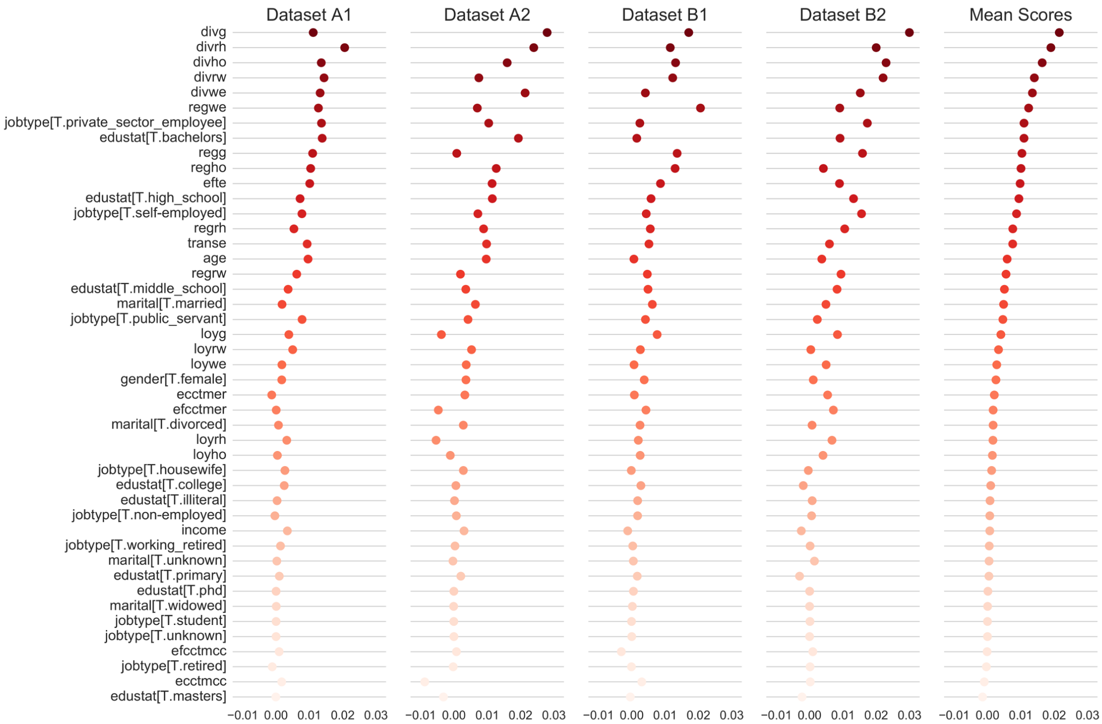

Figure 5.3: Feature analysis. Illustrated is the importance of the features which were calculated based on the mean decreasing values of area under ROC curve after randomly permuting the relevant feature. Higher decreasing value of a feature indicates more contribution.

To identify the effect of gender and age on the predictability of the churners, we prepared separate models for age (under 30, between 30 and 50, above 50) and gender (males and females) groups. In doing so, we divided the data into subgroups each of which have the data for the particular group members (e.g. males), and built the prediction models as described previously. The plot of the evaluation of the age group models based on data sets A1 and B2, and the gender group models based on data sets A2 and B2 are shown in Figure 5.4. Please see Figures A.4 and A.5 for the results generated with other data sets.

As shown in Figure 5.4A, for the unbiased data set, male and female customer groups have no significant superiority over each other in terms of predictability.

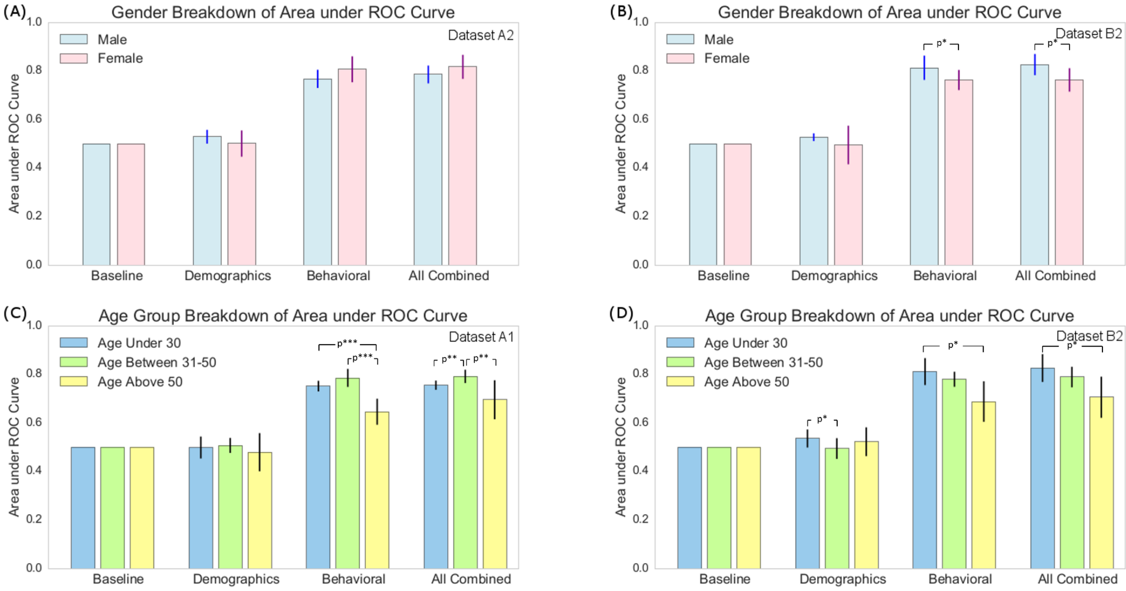

Figure 5.4: The breakdown of churners based on *gender* and *age groups*. The comparison is performed for the label *inac-full*. The length of the error bar corresponds to 1.0 standard deviation. $p*$, $p**$, and $p***$ denote the significance levels with p values smaller than 0.05, 0.01, and 0.001, respectively.

On the other hand, for the data set B2 male customers seem to be more easily predicted with both behavioral ($t(7) = 2.19, p < .05$) and combined models ($t(7) = 2.74, p < .05$) as shown in Figure 5.4B. For the age group analysis, we observe that the elderly customers (above age 50) seem to be significantly more difficult to predict (all p's$< .05$, except for data set A2). In particular, this finding is more significant in the analysis with behavioral models based on data set A1, as shown in Figure 5.4C.

In order to observe the diversity and loyalty trends of churners and non-churners, we plotted aggregated diversity and loyalty feature values with respect to 3-month periods as shown in Figure A.6. The churners seem to have increasing loyalty and decreasing diversity trend towards the time they decide to churn whereas the non-churners have flat trends for the same feature sets. This finding might lead to an intuition such that the descending diversity trend of churners could be explained with the general financial activity decrease of the churners.

In order to differentiate the effects of diversity decrease from the diminishing financial activity, we carried out a study on the comparison of the financial activity trends of churners and non-churners. To this end, we aggregated overall financial activity of each customer into monthly bins and applied a linear fitting to his/her monthly financial activity levels (in terms of both number of transactions and spending amount). We then considered the slope of the fitted line as a feature for predicting the churning decision of each customer. The slope values for churners and non-churners are shown in Figure A.7. Based on the prediction tests that we performed with a single-node Decision Tree classifier trained with financial activity trends of the customers, we found out that the predictive performance of financial activity trends in terms of expenditure amounts was similar to that of demographic features (0.52 AUROC score), and it was statistically significantly worse than the performance of behavioral features in general ($t(7) = 28.02, p *** < 0.001$). The performance based on activity trends in terms of transaction counts was relatively better than the demographic features (0.65 AUROC score); however, it was also significantly worse than the performance of the behavioral features ($t(7) = 13.73, p*** < 0.001$). From the results, we confirm that our behavioral features capture the signal of customer's churning behavior more precisely than a simple feature of inactivation trend of credit card usage.

5.1.3 Discussion

In this study, we have shown that churning behavior can be predicted to a large extent by analyzing behavioral patterns of bank customers in several dimensions. This result not only solidifies previous results in the literature about the relationship between spatio-temporal mobility patterns and individual financial well-being [25], but also serves as a first step towards effective modeling of churning behavior using large-scale financial transaction data.

The diversity and regularity features seem to have a systematic and large effect on the prediction performance. As shown in Figure A.6, the diversity scores of churners have a decreasing trend while this is not the case for non-churners. This is in one sense analogous to Selye's characterization of the response to stress of high-level organisms, according to which the body of the organism gets exhausted and vulnerable at the later stages of the persistent stress [190]. In other words, the main factors affecting the customers and pushing them to churn might also limit their spending power and make them become more cautious and conservative in terms of financial decision-making, which, in our case, is reflected by an irregular shopping behavior, i.e., a decreasing exploration level towards the moment they become inactive. This suggests that, in general, diversity and regularity could be strong predictors of financial stress.

Our results also suggest that younger people seem to be relatively more predictable compared to elderly people as far as the behavioral patterns are concerned (Figure 5.4), an observation consistent across several data sets that are considered. This difference in prediction performance for different age groups is consistent with the relatively high ranking of the feature *age* as shown in Figure 5.3. Compared to age, however, we find out that there seems no significant difference between the predictive power of behavioral features for male and female customers except for the weak significance signals in the analyses made with data sets B1 and B2 (Figure A.5). We also performed the analyses for various data sets generated with different churn definitions (based on credit card and checking account usage), different time frames (with only weekend data and customer-specific observation window), and different parameters for behavioral feature extraction (e.g., edge size of grid bins and radii of annular areas for radial bins); however, the reported results did not significantly change.

It might be suggested that the usage of online and offline credit card transactions should be evaluated differently in the choice behavioral trait. Nevertheless, as can be seen in the importance ranking of the features, merchant-wise and merchant type-wise entropy (*ecctmer* and *ecctmcc*) and their offline variants (*efcctmer* and *efcctmcc*) are ranked close to each other (Figure 5.3). This implies that the distinction between online and offline transactions for the choice behavior does not seem to be important in the present data set.

Understanding reasons behind churning activities is extremely important for financial institutions to accurately deliver more engaging and rewarding experience for customers. However, finding causal relations between certain factors and churning decisions could be a very difficult task, as churning might be due to a wide range of personal circumstances such as job loss or unsatisfying customer service. It is therefore worth noting that the results presented in this study reveal statistical correlations between behavioral patterns and churning activities, and do not support a causal relation between the two. More research needs to be done, potentially by combining data-driven approaches presented in this study and traditional methodologies based on surveys and questionnaires, to fully understand why and when people decide to churn away from banking products and services. However, our results at least suggest that, even without taking into account the actual financial activities of the customers, such as monthly spendings or savings, greater accuracy in churn prediction may still be achieved merely based on customers' behavioral patterns.

Our study has several limitations. The data sets of credit card transaction records used in this study are based on samples of the full customer set of the financial institution. Therefore, sampling bias could exist and potentially influence the results. Another limitation of using credit card transaction data is that, credit card holders may only represent a certain fraction of the population, and people may choose to pay by cash under certain circumstances. However, our data sets do cover a relatively large period in time, which makes our results robust against external factors that might influence customers' financial activities such as seasonality and economic instability.

Spatial and temporal signatures in customers' mobility patterns seem to be promising features for predicting financial outcome and decisions, as evidenced by our results and those of [25]. Fortunately, behavioral traits employed in these studies depend only on the data that have references to time and space, thus making them independent of the particular application domains. We therefore expect that the generality of these features enables the applicability of our churn prediction methodology in other business domains or industries, such as telecommunication or insurance services. On the other hand, for our methodology to work efficiently and effectively, especially from a streaming data point of view, organizations must employ mechanisms and technological frameworks to streamline the applicability of our approach on a daily basis, given the constraints on data availability, data quality and data confidentiality.

There is a growing research community working on utilizing quantitative data for understanding human and social behavior, with several dedicated academic venues [198], [199]. However, many works have focused on problems with behaviors that are correlated with poverty and changes in behavior as a signal for future financial problems. This may be in part because it is still rare to have both financial information and detailed mobility information, and in part because churn is more a commercial problem than a social problem. Consequently, the current study is unique within the computational social science literature. It is also significant because, to our knowledge, it is the first study to find behavioral signals that predict significant changes in a daily habitual behavior. It may be, for instance, that these or other related behavioral signals can also predict changes in other daily habitual behaviors such as shopping, dining, or work-related patterns.

5.2 Observation Data

The analytic system reported in this chapter is basically a set of procedures that had to be followed by the data scientists. During the course of the data cleaning and predictive model generation as well as its evaluation, data scientists issued various data manipulation and processing operations through command line interface of an open source program. Development and evaluation of the analytic system required five moths of research effort, due to this reason, the observation data has been used by applying methodologies from ethnography such as fly-on-the-wall observations. During this longitudinal case study, objective and behavioral data have been collected only for the sessions during which the scientists interacted with tool for data processing purposes. Subjective data has also been collected via structured interviews in order to better understand the analytic processes occurring in the unresponsive environments.

5.2.1 Objective Data

Objective data collection for this analytic system necessitated employment of methodologies that are different from those explained in Chapters 3 and 4. Data collection was performed as a longitudinal case study during which a series of interviews and ethnographic observations have been performed. Data scientists were also asked to develop their analysis procedures so that the analytic system logs the time and task type. During the observations, notes have been taken to keep track the timing and type of human-side messages. The data sources for the objective data calculation are as shown in Table 5.3.

Table 5.3: Data sources for interaction characteristic data calculation. Objective data sources for the analytic system study is listed. *Log* corresponds to the interaction logs on the data side, *video* is the records of the sessions, *expert review* is the data extraction process carried out by analytic system's expert, and *heuristics* is a predefined procedure for the calculation of the corresponding data set.

Characteristic	System 5
Human-side Message Timestamps	Notes
Human-side Message Types	Notes
Media	Expert Review
Medium Physicality Constants	Expert Review
Data-side Message Timestamps	Log + Notes
Data-side Message Types	Log + Heuristics
Data Consumption Rates	Notes

The calculated analytic system properties are shown in Table 5.4. It should be noted that the calculations are based on the data collected during the active analysis sessions. The time elapsed between the analysis sessions was disregarded as those portions of time are not representative for the purposes of my thesis research. For example, elapsed time between the initiation of a clustering operation and on-system presentation of the results are counted towards the calculation of the objective data. Similarly, the discussion periods between two analytic commands given on the system were considered in the objective data calculation. However, the times when the analysts worked on different projects were not considered in the scope of my research.

Much of the data cleaning, manipulation and preprocessing operations (i.e., commands) were less costly compared to those such as predictive model training, oversampling, cross-validation, and clustering. In some cases, model training and clustering operations could take several hours leading to very low responsiveness value for the analytic system. As all the human- and data-side messages are sent over the computer screen, the system was evaluated as being completely digital as verified by the domain experts. Except for few kinds of unit task types such as data retrieval, almost all of the task types were equally employed, however, repeated filtering operations performed during the data cleaning operations increased the unit task diversity.

Due to the long time deltas between the costly analytic tasks such as clustering and model training, there has been cases where human-side messages are sparsely distributed resulting in very high human-side message closeness factor. All the analytic system operations used the data as a single chunk, hence the reason of the 0 value for progressiveness level.

Table 5.4: Analytic system property values for system 5. Property values for the analytic system are calculated based on the interaction characteristics data. These values form the objective part of the observation data.

Property	System 5
Responsiveness	0.03
Communication Media Level	1.0
Unit Task Diversity	0.43
Human-side Message Closeness Factor	1.12×10^3
Progressiveness Level	0

5.2.2 Behavioral Data

Much of the data analysis sessions were conducted as single-user interaction with the analytic system. Most of the collaboration occurred during the pre- and post-study meetings, and consequently no behavioral data pertaining to the collaboration and communication aspects could be collected. Data about the hypothesis, insight and question formation, the extent of feeling distracted from the ongoing analysis activity, and the relationship between the length of the analysis sessions and the frequency of the task-related communications are investigated with mainly subjective data.

5.2.3 Subjective Data

In order to compensate the lack of behavioral data, self-reported measures pertaining to the efficiency of the analysis sessions in terms of insight generation, and idling cases, interviews are performed before and after the analysis sessions.

- *Pre-study self-reports.* Users of the analytic systems were expected to describe their analysis strategies for the given tasks prior to the case study sessions. This subjective data was employed in order to measure the extent that the analytic system affords itself as for its capabilities.

- *Post-study self-reports.* Semi-structured interviews was conducted in order to gain insight on the effect of unresponsiveness on the efficiency of the analysis sessions.

5.3 Realm Characterization

The reported analytic system has been plotted on the subregion of the HDI exploration space where the responsiveness level is low and the communication is made mainly on the digital media. Currently available observation data is based on only a single analytic systems research, and due to this reason, characterization for this realm should be considered as mere descriptions based on a *probe* into this realm of the space rather than as overall statements.

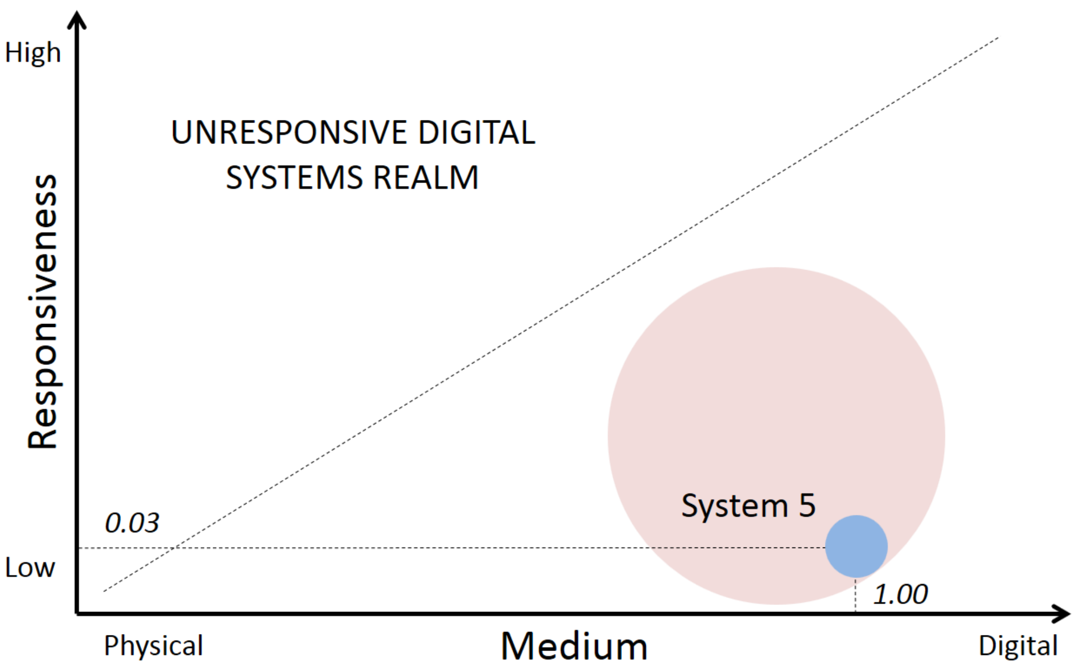

Figure 5.5: Unresponsive digital systems realm. Plot of system 5 on the exploration space of HDI is represented.

Analysts participated in the study reported that they typically have some form of hypothesis or research question to investigate at the beginning of the analysis processes. They stated that such approach is mandated by the fact that the analyses they perform usually takes a day or two, and due to this reason, they need to design their methodology carefully, typically in group meetings, and apply their decisions. According to the pre-study self reports, prior to starting long lasting batch operations, they try many distinct ideas on small portions of the data set in order to acquire initial results in short periods of time (high responsiveness). This is inline with my finding that users tend to utilize responsive systems in order to explore the data and generate as many initial insights and hypotheses as possible as reported in Chapters 3 and 4. In this regard, unresponsive systems realm seems to be hosting systems with which the users expect to get final results with predefined sets of methods and particular data sets.

Lack of support in exploratory data analysis and responsiveness lead the analyst individually work on the problem rather than setting up concurrent collaborative environment. Based on the post-study interviews I assert that they tend to collaborate in the meetings, and in less frequent circumstances, conduct focus groups around analysis tools in order to test a number of ideas and reflect on them in relatively short periods of time.

In case of extremely long response times to human-side messages, analysts report that the analysis sessions become far from engaging, and switch to other tasks until getting response from the system. Such context switches necessitates some form of book keeping of the analysis states and next action steps.

Contrary to what is expected, the reported unresponsive system has relatively high unit task diversity value compared to the average of those scores reported in other realms. Closer investigation of the interactions performed during the data cleaning and preprocessing steps reveals that these operations require trial of many computationally inexpensive steps resulting in highly interactive sessions. These phases of the data analyses seem to increase unit task diversity score.

CHAPTER 6

DISCUSSION & CONCLUSION

6.1 HDI Exploration Space

As explained in Section 2.4, maintaining HDI exploration space could bear potential benefits to the community. For example, given a task specification of an analytic problem, we can refer to the space with task properties, determine the realm in which the proposed system reside, have recommendations on the properties of the analytic system that we will need to design, and even search through example systems along with their user study data. Existence of such analytic system space could also foster usage of a shared terminology for analytic systems.

Nevertheless, maintaining HDI exploration space and associated procedures would require a community-wide support rather than individual efforts. The direction of such community support should consider topics ranging from property definitions to databases that will keep data about analytic systems, which I explain in the following subsection. As could be guessed, objective data matrix could develop as a sparse matrix along with the definition of new system properties, and I explain the issue in detail in Section 6.1.2.

6.1.1 Maintaining the Exploration Space

Potential usefulness of exploration space could be realized by maintaining currency of the objective data set, realm characterization and data extraction methodologies, and set of properties along with their definitions.

Only five of the analytic system properties could be introduced within the scope of my thesis, and infinitely many of them could be identified by others. Such identification should also involve their calculation methodology based on the concepts of human-data interaction conceptualization. In my conceptualization, interaction between human and data could be boiled down to primitives such as human- and data-side messages, their types, timestamps, and medium constant. Based on such primitives, one could generate various properties for analytic systems. Conceptualization of the human-data interaction should also be maintained as open to community improvement.

Analytic system properties calculation efforts should be conducted in the frame of a certain set of standards that could be designed by the community effort. For each human-data interaction conceptualization, different standards could be formed, and so that objective data sets could be referred to with respect to the standards based on which they are calculated.

Any change in the analytic system property set will have a reflection on the metadata of the objective data set. Due to this reason, versioning of the metadata of the objective data set should be tied to that of analytic system properties calculation standards.

Methodology for observation data extraction and realm characterization for both future and past analytic systems should also be determined by certain authorities managed by communities. As suggested by Strauss et al. [33], main idea behind the formality of qualitative data analysis is the application of the methodologies according to the publicly accepted guidelines. Similarly, extraction of behavioral data from the analytic system study observation data is crucial for the reliability of the findings that will be acquired with exploration space of HDI.

6.1.2 Sparsity of Objective Data Matrix

Not all different kinds of analytic systems could be explained with a unique set of properties as explained in Section 2.5.2, and instead, a union of property sets of the analytic systems could be used where some properties could get special *not applicable* value, abbreviated as *NA*. Due to this reason, the sparsityof the objective data matrix is inevitable as the set of system properties will expand.

As a generally accepted mantra in the data science community, missing or non-existing values are also important piece of information, and they should also be contributed to the sensemaking processes. In this regard, applicability of a property to a given analytic system can help better in differentiating the systems from each other. Application of clustering operations on the objective data set should also be dominated by the existence of *NA* values while determining realms.

6.2 Improvement on the Methodology

6.2.1 Backward Application of the Methodologies

Maintaining exploration space efforts encompass forming a set of standardized procedures to extract observation data from future data analytic systems as explained in Section 6.1.1. Unfortunately, mapping old systems to the exploration space is a more challenging problem as most of the concurrent or legacy analytic systems are not distributed or published as academic work along with their usability test data. Except for rare cases, none of them logs interaction patterns of their users. Even if the research data of an open analytic system academic project is available, the chances will be that the collected data would not be a good fit for the kind of data what we call "observation data". As consequence of all these drawbacks, contributing old studies to the exploration space and extracting their objective and behavioral requires conduct of case or usability studies with those old systems.

As I explained in realm characterization sections, formation of observation data may require different approaches. For example, for the unresponsive digital systems realm, one would need to conduct a longitudinal case study in order to comprehend the relationship between long response times and engagement. For highly responsive systems, a few sessions of exploratory analysis would be adequate to collect behavioral data and interaction characterization logs. For the unresponsive and highly physical analytic systems such as data sculptures, controlled laboratory experiments could be performed with considerable number of repetitions. In all of the cases, users of the system are expected to be familiar with the data that will be investigated to a specific extent. Due to this reason, for case studies or usability testing, domain experts should participate to the study. For the case of controlled laboratory experiment, data of well-known systems, such as Twitter data, may be adopted in order to be able to hire high number of novice participants.

6.2.2 Role of Expert Reviews

There seems to be a discussion on whether the expert reviews have a significant role in the evaluation of the information visualization systems [28]. Aside from this discussion, I employed the expert reviews as a methodology to partially support observation data collection rather than as an evaluation methodology.

During the course of human-data interaction, decision on whether each individual human- or data-side message is sent on physical or digital media (a and b values in Section 2.5.2.2) necessitates human judgment. In such cases, decision of which type of messages are sent over which media could only be made with careful observation of a user who is familiar with both data and problem. From statistical point of view, investigation of a single expert would not be reliable and at least two or ideally three or more expert should evaluate physicality constants. The reliability of their judgments are then determined based on the intercoder reliability checks [200].

6.2.3 Data Collection Issues

Main goal of the case studies is to investigate the phenomenon in question in its real world with real users. In this regard, effort to control any other factors that are not in the nature of a topic that we would like to investigate may lead to artificial actions of the users. Due to this reason, adoption of data collection techniques requiring usage of wearable equipment might affect the reliability of the collected data. As I have observed, even bare video recording of the analysis sessions seem to lead to self-restraint of the users while communicating or interacting with the tools.

On the other hand, employment of ubiquitous data collection methods seems to be a better approach as they are less disruptive. Such techniques involve gaze and bodily gesture detection from the video recordings of the analysis sessions.

Interaction characteristics data could be collected with plugins for human- and data-side message traffic. Such plugins could detect the message type and timestamps and how much of the data is consumed, and record such data per user basis. With simple application programmer interfaces and wide support for various technologies could foster usage the library and objective data collection.

Hiring adequate number of domain experts could be a compelling as they tend to be busy and their time might be constraint. At least for the cases where the task of the domain expert could be simplified so that relatively high number of novice participants can perform the task of domain experts. For example, for multiuser collborative tabletop systems, domain experts' investigation could be an obligation whereas a rather simple system with a single user and standalone desktop application could be evaluated with a group of novice users. Platforms for hiring such novice users could be established.

6.2.4 Application of HCI Methods

I suggest that findings from the literature about gaze and gesture could be employed in order to better characterize HDI space realms. Gaze cues in social interaction are highly related to attention [201], learning [202], and recall [203], [204]. Particularly in the collaborative analysis setups, I suggest that, by collecting gaze and gesture direction and duration, we can study the correlation between particular patterns in the gaze and gesture data and the task-tool match quality. Such social behaviors of users could have specific distinct patterns for the analytic systems in different realms.

I investigated co-located collaborative analytic systems where face-to-face communication was employed. Body of research in the computer-mediated communication domain report the issues arise during the communication carried out over computers or similar online systems [205], [206]. Chances are likely that new ways of participant behavior collection would need to be employed for the analytic systems supporting computer-mediated communication.

Having not been considered in the scope of my thesis work, biological and social interaction data could also be collected from the users of the analytic systems. Such data of face-to-face social interaction data can then be employed in order to measure the productivity of the analyst group [207], [208].

6.3 Conceptual Matters

6.3.1 Other Media Types

Medium types employed in my thesis work have been decided based on the observation that most of the analytic systems reported in the literature were adopting digital physical medium. By the term physical, I indicate the medium that is sensible by the human touching perceptual capability. From technical point of view, sound or light could also be considered as physical medium if they were to be used as part of analytic system. For example, Smith et al. developed several auditory data representation techniques and generated stereophonic sounds emanating from two-dimensional space [209]. Nevertheless, adoption of such approaches is significantly rare compared to the typical media employed in the concurrent analytic systems, namely, physical and digital media.

6.3.2 Other Dimensions

Two-dimensional space represented in this thesis was merely one of the many possible combinations of analytic system properties. Put differently, various combinations of properties or different number of properties could be used in order to explain the HDI space. For example, I could have explored three-dimensional space of HDI whose axes, for example, would have been unit task diversity, human-side message closeness factor, and progressiveness level. All different exploration efforts would be investigations of the same HDI space from distinct perspectives, and each individual investigation would lead to new realms of that particular investigation.

Characterization of realms for a selected set of analytic system properties requires conduct of all the methodologies explained in my thesis research. However, distributions of the analytic systems reported in this thesis could be made based on the objective data as shown in Figures 6.1 and 6.2. As can be seen in the plots, new clusters are formed in various parts of the *new perspective* of the HDI space. Characterization of new realms that could be formed on this new perspective necessitates new grounded theory analyses.

6.4 Projections of Previous Analytic Systems Research on HDI Space

Unfortunately, currently available analytic systems or application can not be readily placed in the HDI exploration. Adequate amount of observation data should be collected and analyzed based on the methodologies discussed in Section 2.6.

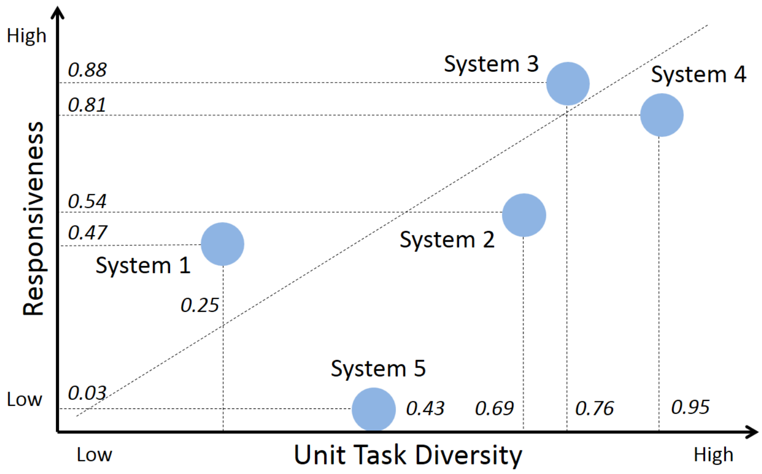

Figure 6.1: Plot of analytic systems on exploration space formed with unit task diversity and responsiveness properties.

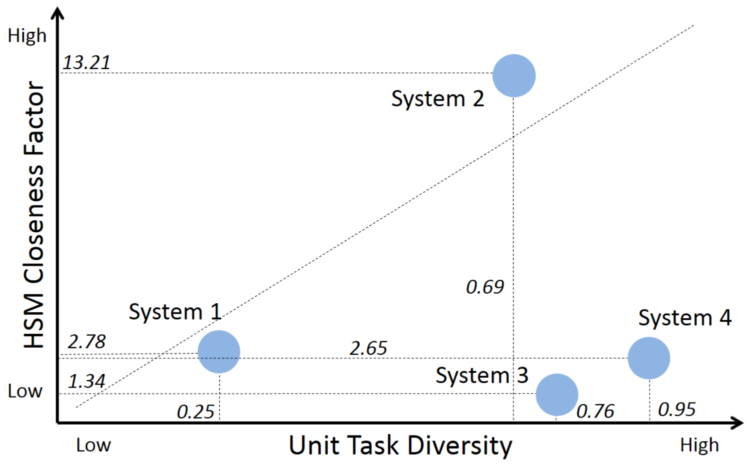

Figure 6.2: Plot of analytic systems on exploration space formed with unit task diversity and human-side message closeness factor properties. Due to the extremely high closeness factor value, Study-5 is not shown in the figure.

However, I will attempt to project some currently existing analytic systems to HDI exploration space formed with communication media level and responsiveness mainly based on my own informal observations. It should be noted that the projections that I report in this section will barely be made on my personal judgment and for the exact results, formal case studies or longitudinal observations should be performed.

6.4.1 Data Physicalization

Physical representations of data has attracted a growing interest from the information visualization community [210]. Published material on data physicalization is currently been evaluated in information visualization venue as they are considered as lacking interactive elements, hence, none of the existing physicalization is attributed

Figure 6.3: Financial data physicalization generated with PhysVis.

as a visual data analytic system. Nevertheless, I suggest that, even practically they do not afford interaction in the practical sense, there is still a communication between human and data when a human analyst tries to make sense of the physicalized phenomenon.

In Figure 6.3, examples of data physicalization designed by Dogacan Bilgili and Cem Batmansuyu, and created by PhysVis, a web-based application for data sculpturing [211]. PhysVis is also attributed as one of the first application known to implement whole physicalization pipeline. Both physicalizations represent a summary of monthly aggregated credit card transaction. Brief informal case studies have been performed with a group of participants, and during the case studies, participants were expected to think aloud as they look at the physicalizations in their hands. Each deliberate movement done by the participants were considered as human actions, i.e., human-side messages, such as zoom in or out, rotate, information retrieval, and even comparison. However, except for rotate, zoom in or out, information guess by bare human vision system, user actions were considered idempotent as their actions did not change the appearance of the physical model, and hence, considered as the system was not responsive for those actions. Contrarily, other actions were considered as highly responsive: When user turns the physicalization in his or her hand, the physicalization responds immediately to physical action of the users. Based on this interaction conceptualization, human- and data-side messages are extracted and the objective data shown in Table 6.1 is calculated.

Developed by Bilgili et al. [211], in Figure 6.4 data physicalizations enhanced with augmented reality applications is demonstrated. Augmented reality improves

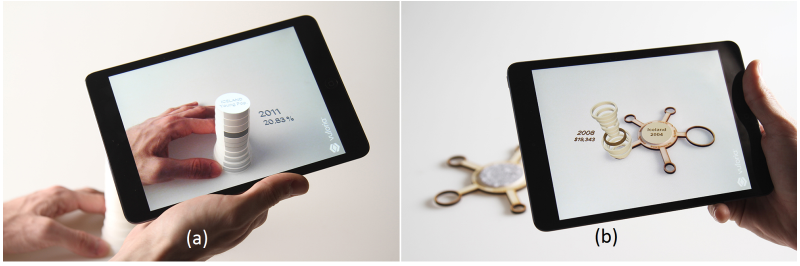

Figure 6.4: Financial data physicalization enhanced with augmented reality. Augmented reality facilitates an interaction mechanism in which users' physical actions are captured from the camera of the computing device, and change on the data representation is tagged on the reality as virtual extensions of the data physicalization.

the physicalization by providing partial interaction functionality with which users retrieve or filter information, and acquire detail on demand through the display of a tablet computer, however, while interacting with the physical *or virtual* part of the system. For example, as shown in Figure 6.4(b), user could select one of the five features from the physicalization, and upon selection, the application generate the virtual counterpart of the visualization on the screen. And then, user could further select the virtual part of the physicalization with bare hand. Hence, physical actions of user cause representation change in the virtual part of the analytic system[7]. Interactions of the participants with this system has also been investigated in informal case studies, and the objective data in Table 6.1 has been calculated. As is expected, due to augmented reality capability, responsiveness and media constant of the analytic system has increased.

Responsiveness and communication media level of the data physicalizations position these systems on the low responsive physical side of the exploration space, and closeness of their plots leads to a new cluster, however, this result should be verified with formal case studies.

Table 6.1: Analytic system property values for data physicalization system and GNU PSPP software. Property values for the analytic systems are calculated based on the measured interactions performed during the informal usability test.

Property	Data Phys.	Data Phys. + AR	GNU PSPP
Responsiveness	0.25	0.43	0.69
Communication Media Level	0	0.22	1
Unit Task Diversity	0.55	0.72	0.87
HSM Closeness Factor	16.4	12.8	3.1
Progressiveness Level	0	0	0

[7]https://vimeo.com/227340532

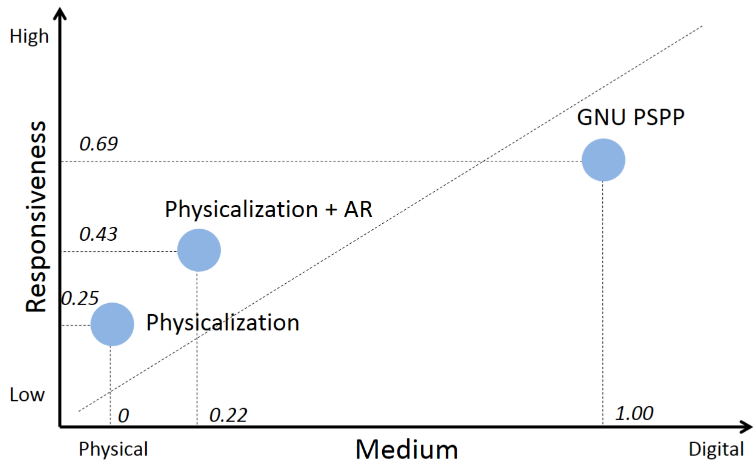

Figure 6.5: Plot of physicalization and GNU PSPP analytic systems on exploration space. Even enhanced with augmented reality capabilities, data physicalizations seem to form a distinct realm in the lower left side of the exploration space. GNU PSPP seems to reside in responsive digital systems realm.

6.4.2 GNU PSPP

GNU PSPP [212] is a statistical analysis tool developed as a free replacement for the proprietary program IBM SPSS. Being a standalone desktop application, PSPP supports most of the commonplace and advanced descriptive and inferential statistical functionalities. The usability test that I have performed with the software involves credit card transaction data set introduced in Section 4.2.3. Five statistical summary operations, five two-factor analysis of variance (ANOVA) tests, and 5 clustering analyses have been performed, and the response times of the application was measured with a stopwatch. The calculated analytic system properties are as shown in Table 6.1, and plotted to exploration space as shown in Figure 6.5. GNU PSPP seems to reside in responsive digital systems realm.

6.5 Concluding Remarks

Astonishingly wide variety of the analytic systems complicate the task-tool match problem leading waste of resources, or worse, to the wrong decisions due to the wrong way of handling the data. Even more, this variety seems to affect the academia which investigate the analytic systems in their own venues such as visual analytics and data analytics. In this thesis work, I suggested an approach in which all the interaction occurring between human and analytic systems are reduced to a simple human-data interaction concept. While using an analytic system, users are considered as communicating with data over a set of media. In this communication scheme, both human and data sends messages to each other affecting the users' mental models and representations of the data. Timings, contents, and types of those messages could then be employed to evaluate general properties of data analytic systems regardless of the kind of the analytic system.

Availability of objectively calculated properties of analytic systems can assist us in investigating in a quantitative way from different perspectives. For example, I introduced a two-dimensional exploration space by taking two of those properties as axes, and plot five analytic systems onto this space for which I had enough usability observation data. Visually identifiable clusters of those analytic systems on the exploration space serve as representatives of all existent analytic systems that could be plotted on the subregion on which they reside, which I called realms. Identification of the properties of the realms has been a beneficent effort as such characterization would be adopted in the design of new analytic systems whose purposes are known, however, analytic properties are not decided yet.

Existence of an actively updated objective data regarding the analytic systems could serve as a reference system for all analytic systems. By positioning into the exploration space, designers, customers, and other stakeholders of a proposed analytic system can explain the system with a set of parameters. Such a referencing system will also foster usage of a share terminology among the data scientists and analysts from different venues unlike the current situation today. Moreover, exploration spaces can also assist the envisioning future analytic systems. For example, for the empty regions of given exploration spaces, we can extrapolate the features of the closest analytic systems and employ these extrapolations as a starting base for developing visions.

Nevertheless, characterization of realms necessitates availability of observation data collected during the formal case studies, controlled experiments, and longitudinal observations. Few of the existing published research seem to report their experimental data which, unfortunately, are not in the processable format enabling calculation of system properties. Due to this reason, it is of importance to develop guidelines and frameworks facilitating collection of observation data with less effort.

Analytic system properties and observation collection methodologies are not in their final mature form, I assert their development should be continued by the community. As the set of system properties expand, other forms of explorations spaces should be investigated, and as its consequence, multiple referencing systems for different goals should be developed. The success of the exploration space idea that I introduced in this thesis will mainly rely on the strong adoption of the data set and methodologies by the community.

Following the invention of computers and digital systems, the way we interact with those systems have been the focus of interest for the last three decades. As is well known and widely accepted, the way we manage and make use of the tremendously increasing data will be one of the main issues of the computing world. In this regard, systematic characterization of the way we interact with data and formation of analytic systems database can serve as an important guidance mechanism in struggling with this ever-growing complexity.

APPENDIX A

SUPPLEMENTARY RESULTS OF CHURN STUDY

A.1 Performance Comparison of Models Trained with Various Churn Definitions

Various churn definition variants are listed and performance summaries of models built with these churn definitions are provided. The significance test results of the pair-wise comparisons of feature sets as well as a detailed list of parameters applied in the prediction models are included.

The area under ROC curve scores of the churn prediction models generated with respect to churn definitions are listed in Table A.1.

Table A.1: Predictive model scores for various churn definitions.

Label	Label Set	Data Set A1			Data Set A2		
		DMG	STC	ALL	DMG	STC	ALL
inac-full	SB	0.51	0.78	0.79	0.52	0.79	0.81
inac-3m	SB	0.53	0.76	0.78			
inac-l1m	SB	0.54	0.74	0.76	0.53	0.75	0.77
inac-l2m	SB	0.54	0.75	0.77	0.53	0.77	0.78
inac-l3m	SB	0.53	0.76	0.77			
inac-1m	SB				0.53	0.74	0.77
inac-2m	SB				0.54	0.77	0.78
che-lowl3m	CA				0.59	0.74	0.75
cc-zerostmtl3m	CC				0.53	0.67	0.69
cc-ltal3m-ccowned	CC				0.54	0.71	0.72
cc-lowstmtl3m	CC				0.55	0.67	0.69
cc-inacl3m	CC				0.54	0.76	0.76
cc-inacl3m-ccowned	CC				0.53	0.73	0.73
Label	**Label Set**	**Data Set B1**			**Data Set B2**		
		DMG	STC	ALL	DMG	STC	ALL
inac-full	SB	0.54	0.80	0.81	0.53	0.80	0.82
inac-3m	SB	0.54	0.77	0.78			
inac-l1m	SB	0.54	0.74	0.75	0.51	0.77	0.79
inac-l2m	SB	0.54	0.76	0.77	0.51	0.77	0.79
inac-l3m	SB	0.54	0.76	0.79			
inac-1m	SB				0.51	0.76	0.77
inac-2m	SB				0.50	0.78	0.79
che-lowl3m	CA				0.57	0.73	0.74
cc-zerostmtl3m	CC				0.54	0.67	0.69
cc-ltal3m-ccowned	CC				0.53	0.71	0.72
cc-lowstmtl3m	CC				0.55	0.69	0.71
cc-inacl3m	CC				0.53	0.77	0.78
cc-inacl3m-ccowned	CC				0.51	0.74	0.75

Predictive model scores (area under ROC curve) for various churn definitions and data sets are presented. DMG, STC, and ALL stand for demographic feature set, spatio-temporal and choice behavior feature set, and combination of all the sets, respectively. Labels are grouped by the data source based on which they are generated. SB, CA, and CC stand for segmentation-based, checking account usage-based, and credit card usage-based label sets. Listed under the *Label* column are the abbreviations for the churn definitions. Not all the labels could be generated for each of the data sets due to different labeling window sizes of the data sets. For such label-data set combinations, the score fields are left blank.

A.2 Pair-wise Comparisons

A.2.1 Pair-wise Comparisons for Dataset A1

Pair-wise prediction score comparisons of feature sets generated based on data set A1 are presented. DMG, STC, and ALL stand for demographic feature set, spatio-temporal and choice behavior feature set, and combination of all the sets, respectively. For each comparison pair, t-score, degree of freedom (df), and significance levels (p value) are listed. $p*$, $p**$, and $p***$ represent values smaller than 0.05, 0.01, and 0.001, respectively.

Table A.2: Pair-wise comparison results for demographic and spatio-temporal and choice behavior features (Dataset A1).

Label	Label Set	DMG and STC		
		t-score	df	p value
inac-full	SB	33.15	7	$p*** < 0.001$
inac-3m	SB	28.82	7	$p*** < 0.001$
inac-l1m	SB	32.64	7	$p*** < 0.001$
inac-l2m	SB	28.24	7	$p*** < 0.001$
inac-l3m	SB	24.52	7	$p*** < 0.001$

Table A.3: Pair-wise comparison results for demographic features and combined model (Dataset A1).

Label	Label Set	DMG and ALL		
		t-score	df	p value
inac-full	SB	36.20	7	$p*** < 0.001$
inac-3m	SB	29.13	7	$p*** < 0.001$
inac-l1m	SB	36.61	7	$p*** < 0.001$
inac-l2m	SB	30.44	7	$p*** < 0.001$
inac-l3m	SB	25.14	7	$p*** < 0.001$

Table A.4: Pair-wise comparison results for spatio-temporal and choice behavior features and combined model (Dataset A1).

Label	Label Set	STC and ALL		
		t-score	df	p value
inac-full	SB	1.38	7	$p > 0.1$
inac-3m	SB	1.77	7	$p < 0.1$
inac-l1m	SB	2.83	7	$p* < 0.05$
inac-l2m	SB	2.15	7	$p* < 0.05$
inac-l3m	SB	1.68	7	$p < 0.1$

A.2.2 Pair-wise Comparisons for Dataset A2

Pair-wise prediction score comparisons of feature sets generated based on data set A2 are presented. DMG, STC, and ALL stand for demographic feature set, spatio-temporal and choice behavior feature set, and combination of all the sets, respectively. For each comparison pair, t-score, degree of freedom (df), and significance levels (p value) are listed. $p*$, $p**$, and $p***$ represent values smaller than 0.05, 0.01, and 0.001, respectively.

Table A.5: Pair-wise comparison results for demographic and spatio-temporal and choice behavior features (Dataset A2).

Label	Label Set	DMG and STC		
		t-score	df	p value
inac-full	SB	15.38	7	$p*** < 0.001$
inac-l1m	SB	16.23	7	$p*** < 0.001$
inac-l2m	SB	20.90	7	$p*** < 0.001$
inac-1m	SB	25.38	7	$p*** < 0.001$
inac-2m	SB	17.84	7	$p*** < 0.001$
cc-inacl3m	CC	49.92	7	$p*** < 0.001$
cc-inacl3m-ccowned	CC	31.16	7	$p*** < 0.001$
cc-ltal3m-ccowned	CC	30.46	7	$p*** < 0.001$
cc-zerostmt-l3m	CC	16.25	7	$p*** < 0.001$
cc-lowstmt-l3m	CC	16.75	7	$p*** < 0.001$
che-lowl3m	CA	26.89	7	$p*** < 0.001$

Table A.6: Pair-wise comparison results for demographic features and combined model (Dataset A2).

Label	Label Set	DMG and ALL		
		t-score	df	p value
inac-full	SB	18.36	7	$p*** < 0.001$
inac-l1m	SB	18.47	7	$p*** < 0.001$
inac-l2m	SB	22.23	7	$p*** < 0.001$
inac-1m	SB	31.49	7	$p*** < 0.001$
inac-2m	SB	17.70	7	$p*** < 0.001$
cc-inacl3m	CC	50.71	7	$p*** < 0.001$
cc-inacl3m-ccowned	CC	35.33	7	$p*** < 0.001$
cc-ltal3m-ccowned	CC	34.59	7	$p*** < 0.001$
cc-zerostmt-l3m	CC	18.84	7	$p*** < 0.001$
cc-lowstmt-l3m	CC	21.72	7	$p*** < 0.001$
che-lowl3m	CA	27.78	7	$p*** < 0.001$

Table A.7: Pair-wise comparison results for spatio-temporal and choice behavior features and combined model (Dataset A2).

Label	Label Set	STC and ALL		
		t-score	df	p value
inac-full	SB	1.05	7	$p > 0.1$
inac-l1m	SB	2.27	7	$p* < 0.05$
inac-l2m	SB	1.04	7	$p > 0.1$
inac-1m	SB	4.88	7	$p*** < 0.001$
inac-2m	SB	0.92	7	$p > 0.1$
cc-inacl3m	CC	1.46	7	$p < 0.1$
cc-inacl3m-ccowned	CC	0.79	7	$p > 0.1$
cc-ltal3m-ccowned	CC	1.95	7	$p* < 0.05$
cc-zerostmt-l3m	CC	2.54	7	$p* < 0.05$
cc-lowstmt-l3m	CC	2.99	7	$p* < 0.05$
che-lowl3m	CA	1.88	7	$p < 0.1$

A.2.3 Pair-wise Comparisons for Dataset B1

Pair-wise prediction score comparisons of feature sets generated based on data set B1 are presented. DMG, STC, and ALL stand for demographic feature set, spatio-temporal and choice behavior feature set, and combination of all the sets, respectively. For each comparison pair, t-score, degree of freedom (df), and significance levels (p value) are listed. $p*$, $p**$, and $p***$ represent values smaller than 0.05, 0.01, and 0.001, respectively.

Table A.8: Pair-wise comparison results for demographic and spatio-temporal and choice behavior features (Dataset B1).

Label	Label Set	DMG and STC		
		t-score	df	p value
inac-full	SB	21.03	7	$p*** < 0.001$
inac-3m	SB	18.79	7	$p*** < 0.001$
inac-l1m	SB	18.14	7	$p*** < 0.001$
inac-l2m	SB	20.40	7	$p*** < 0.001$
inac-l3m	SB	17.84	7	$p*** < 0.001$

A.2.4 Pair-wise Comparisons for Dataset B2

Pair-wise prediction score comparisons of feature sets generated based on data set B2 are presented. DMG, STC, and ALL stand for demographic feature set, spatio-temporal and choice behavior feature set, and combination of all the sets, respectively. For each comparison pair, t-score, degree of freedom (df), and signifi-

Table A.9: Pair-wise comparison results for demographic features and combined model (Dataset B1).

Label	Label Set	DMG and ALL		
		t-score	df	p value
inac-full	SB	23.82	7	$p*** < 0.001$
inac-3m	SB	19.84	7	$p*** < 0.001$
inac-l1m	SB	21.14	7	$p*** < 0.001$
inac-l2m	SB	25.04	7	$p*** < 0.001$
inac-l3m	SB	21.13	7	$p*** < 0.001$

Table A.10: Pair-wise comparison results for spatio-temporal and choice behavior features and combined model (Dataset A2).

Label	Label Set	STC and ALL		
		t-score	df	p value
inac-full	SB	1.24	7	$p > 0.1$
inac-3m	SB	1.26	7	$p > 0.1$
inac-l1m	SB	1.05	7	$p > 0.1$
inac-l2m	SB	1.01	7	$p > 0.1$
inac-l3m	SB	1.38	7	$p > 0.1$

cance levels (p value) are listed. $p*$, $p**$, and $p***$ represent values smaller than 0.05, 0.01, and 0.001, respectively.

Table A.11: Pair-wise comparison results for demographic and spatio-temporal and choice behavior features (Dataset B2).

Label	Label Set	DMG and STC		
		t-score	df	p value
inac-full	SB	18.66	7	$p*** < 0.001$
inac-l1m	SB	19.01	7	$p*** < 0.001$
inac-l2m	SB	18.43	7	$p*** < 0.001$
inac-1m	SB	21.17	7	$p*** < 0.001$
inac-2m	SB	19.26	7	$p*** < 0.001$
cc-inacl3m	CC	26.13	7	$p*** < 0.001$
cc-inacl3m-ccowned	CC	23.56	7	$p*** < 0.001$
cc-ltal3m-ccowned	CC	36.58	7	$p*** < 0.001$
cc-zerostmt-l3m	CC	14.27	7	$p*** < 0.001$
cc-lowstmt-l3m	CC	17.68	7	$p*** < 0.001$
che-lowl3m	CA	45.40	7	$p*** < 0.001$

Table A.12: Pair-wise comparison results for demographic features and combined model (Dataset B2).

Label	Label Set	DMG and ALL		
		t-score	df	p value
inac-full	SB	22.91	7	$p*** < 0.001$
inac-l1m	SB	22.30	7	$p*** < 0.001$
inac-l2m	SB	28.23	7	$p*** < 0.001$
inac-1m	SB	23.14	7	$p*** < 0.001$
inac-2m	SB	24.00	7	$p*** < 0.001$
cc-inacl3m	CC	26.57	7	$p*** < 0.001$
cc-inacl3m-ccowned	CC	23.06	7	$p*** < 0.001$
cc-ltal3m-ccowned	CC	35.34	7	$p*** < 0.001$
cc-zerostmt-l3m	CC	17.35	7	$p*** < 0.001$
cc-lowstmt-l3m	CC	18.14	7	$p*** < 0.001$
che-lowl3m	CA	42.25	7	$p*** < 0.001$

Table A.13: Pair-wise comparison results for spatio-temporal and choice behavior features and combined model (Dataset B2).

Label	Label Set	STC and ALL		
		t-score	df	p value
inac-full	SB	1.59	7	$p < 0.1$
inac-l1m	SB	1.21	7	$p > 0.1$
inac-l2m	SB	1.16	7	$p > 0.1$
inac-1m	SB	0.85	7	$p > 0.1$
inac-2m	SB	0.56	7	$p > 0.1$
cc-inacl3m	CC	0.43	7	$p > 0.1$
cc-inacl3m-ccowned	CC	0.12	7	$p > 0.1$
cc-ltal3m-ccowned	CC	1.52	7	$p < 0.1$
cc-zerostmt-l3m	CC	1.84	7	$p < 0.1$
cc-lowstmt-l3m	CC	3.50	7	$p** < 0.01$
che-lowl3m	CA	4.12	7	$p** < 0.01$

A.3 Churn Definitions

inac-full The customer is labeled as churner if he/she was evaluated to be in segment *inactive* by the bank for all of the months during the labeling window of the data collection period.

inac-3m The customer is labeled as churner if he/she was evaluated to be in segment *inactive* by the bank for any three of the months of the labeling window of the data collection period.

inac-l1m The customer is labeled as churner if he/she was evaluated to be in segment *inactive* by the bank for the last month of the labeling window of the data collection period.

inac-l2m The customer is labeled as churner if he/she was evaluated to be in segment *inactive* by the bank for the last two months of the labeling window of the data collection period.

inac-l3m The customer is labeled as churner if he/she was evaluated to be in segment *inactive* by the bank for the last three months of the labeling window of the data collection period.

inac-1m The customer is labeled as churner if he/she was evaluated to be in segment *inactive* by the bank for any of the months of the labeling window of the data collection period.

inac-2m The customer is labeled as churner if he/she was evaluated to be in segment *inactive* by the bank for any two of the months of the labeling window of the data collection period.

che-lowl3m The customer is labeled as churner if monthly checking account statements with low balance have been issued for him or her during the labeling window of the data collection period.

cc-zerostmtl3m The customer is labeled as churner if monthly credit card statements with zero balance have been issued for him or her during the labeling window of the data collection period.

cc-ltal3m-ccowned The customer is labeled as churner if he/she owned a credit card and made purchases with very low amounts during the labeling window of the data collection period.

cc-lowstmtl3m The customer is labeled as churner if monthly credit card statements with low balance have been issued for him or her during the labeling window of the data collection period.

cc-inacl3m The customer is labeled as churner if he/she never used credit card, or his/her credit card was inactive during the labeling window of the data collection period.

cc-inacl3m-ccowned The customer is labeled as churner if he/she never used credit card, and his/her credit card was active during the labeling window of the data collection period.

A.4 Parameter Values Applied to Prediction Model

A.4.1 Random Forest Parameters

The Random Forest classification model has been implemented by employing the SciKit-learn library. The parameters are listed (and abbreviated) as in the library's documentation in order to support straightforward reproduction of our results.

Number of Estimators (n_estimators) The number of trees in the forest. Applied Value: 500

The Quality of Split Criterion (criterion) The function to measure the quality of a split. Applied Criterion: Gini Impurity

Maximum Number of Features (max_features) The number of features to consider when looking for the best split. Applied Value: 2

Maximum Depth of the Tree (max_depth) The maximum depth of the tree. Applied Value: No specific limit has been put on the depth of the tree leading to the expansion of the nodes until either all leaves are pure or all leaves contain less than *min_samples_split* samples.

Minimum Number of Samples for Split in Internal Node (min_samples_split) The minimum number of samples required to split an internal node. Applied Value: 2

Minimum Number of Samples at Leaf Node (min_samples_leaf) The minimum number of samples required to be at a leaf node. Applied Value: 1

Limit on the Leaf Node Number (max_leaf_nodes) Grow trees with *max_leaf_nodes* in the best-first fashion. Best nodes are defined as relative reduction in impurity. Applied Value: No limit has been applied.

Minimum Impurity for Split (min_impurity_split) Threshold for early stopping in tree growth. A node will split if its impurity is above the threshold, otherwise it is a leaf. Applied Value: 10^{-7}

A.4.2 SVM-SMOTE Parameters

Synthetic minority over-sampling has been implemented with Imbalanced-learn library by Lemaître et al. We considered it useful to report over-sampling parameters with their abbreviations as listed in the library's documentation to support reproducibility of our study.

Minority and Majority Class Ratio (ratio) The number of samples in minority class over the number of samples in majority class. Applied Value: 0.25

Number of Nearest Neighbors (k_neighbors) Number of nearest neighbors that are used to construct synthetic samples. Applied Value: 5

Number of Nearest Neighbors for Danger Detection (m_neighbors) Number of nearest neighbors that are used to determine if a minority sample is in danger. Applied Value: 10

Extrapolation Step Size (out_step) The step size used in the extrapolation step. Applied Value: 0.5

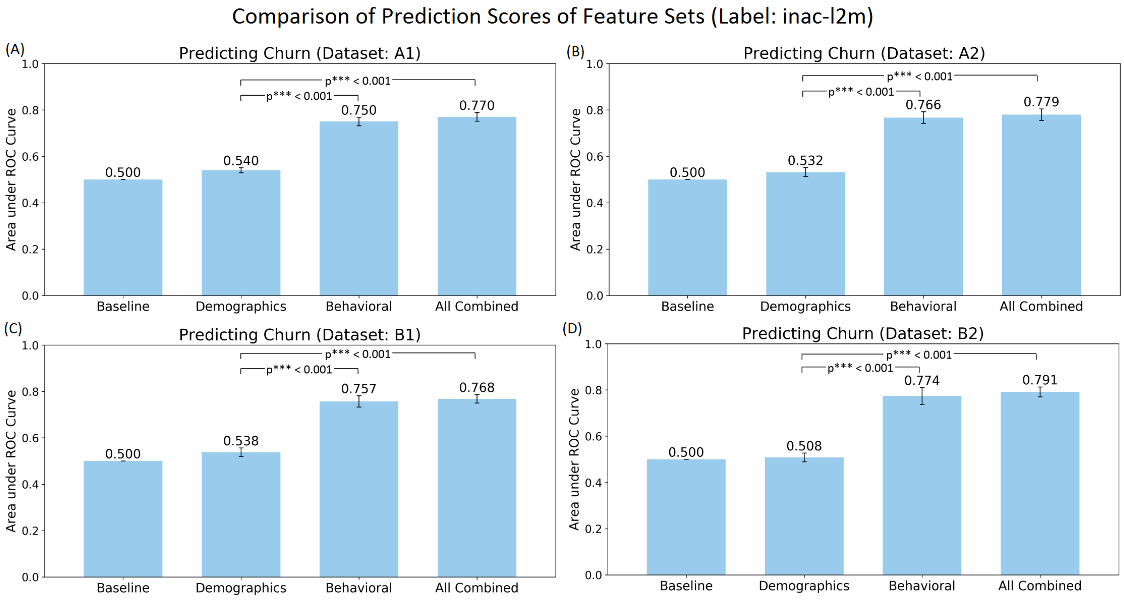

Figure A.1: Comparison of prediction scores of feature sets. Area under ROC curve metric comparison of *demographic* features, *spatio-temporal and choice behavior* features, and the combination of both feature sets for each of the data set versions. The length of the error bar corresponds to 1.0 standard deviation. The comparison is performed for the label *inac-l2m*. $p***$ denotes the significance level with p value smaller than 0.001.

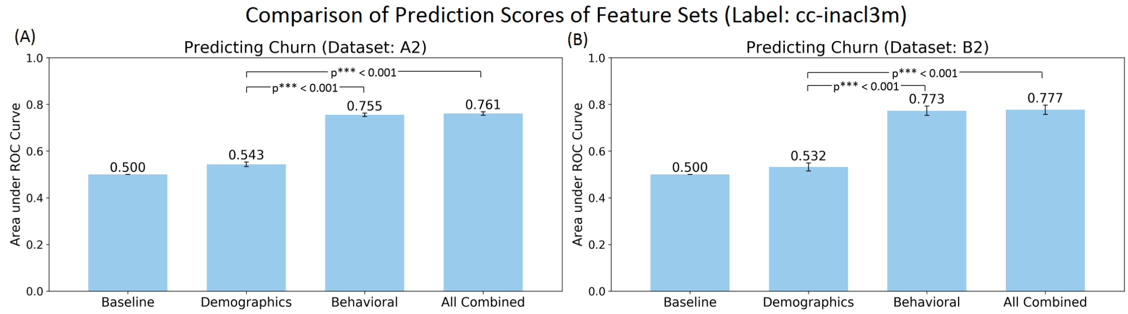

Figure A.2: Comparison of prediction scores of feature sets. Area under ROC curve metric comparison of *demographic* features, *spatio-temporal and choice behavior* features, and the combination of both feature sets for each of the data set versions. The length of the error bar corresponds to 1.0 standard deviation. The comparison is performed for the label *cc-inacl3m*. $p***$ denotes the significance level with p value smaller than 0.001.

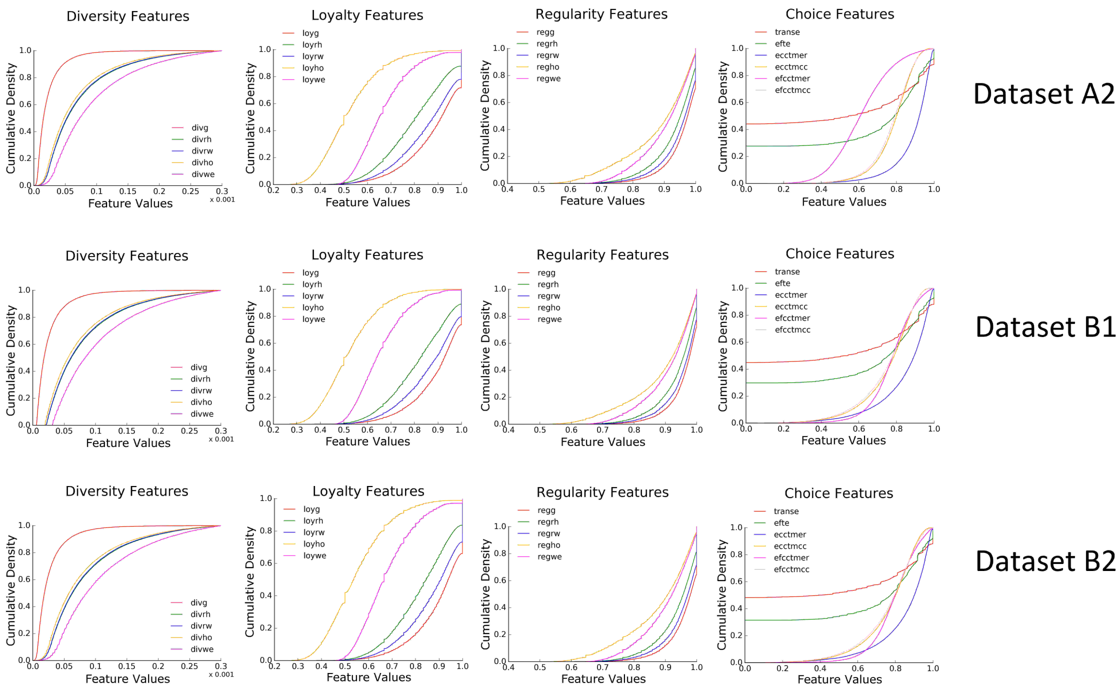

Figure A.3: Cumulative density functions for data sets A2, B1, and B2. In general, distributions of the features are similar across all the data sets.

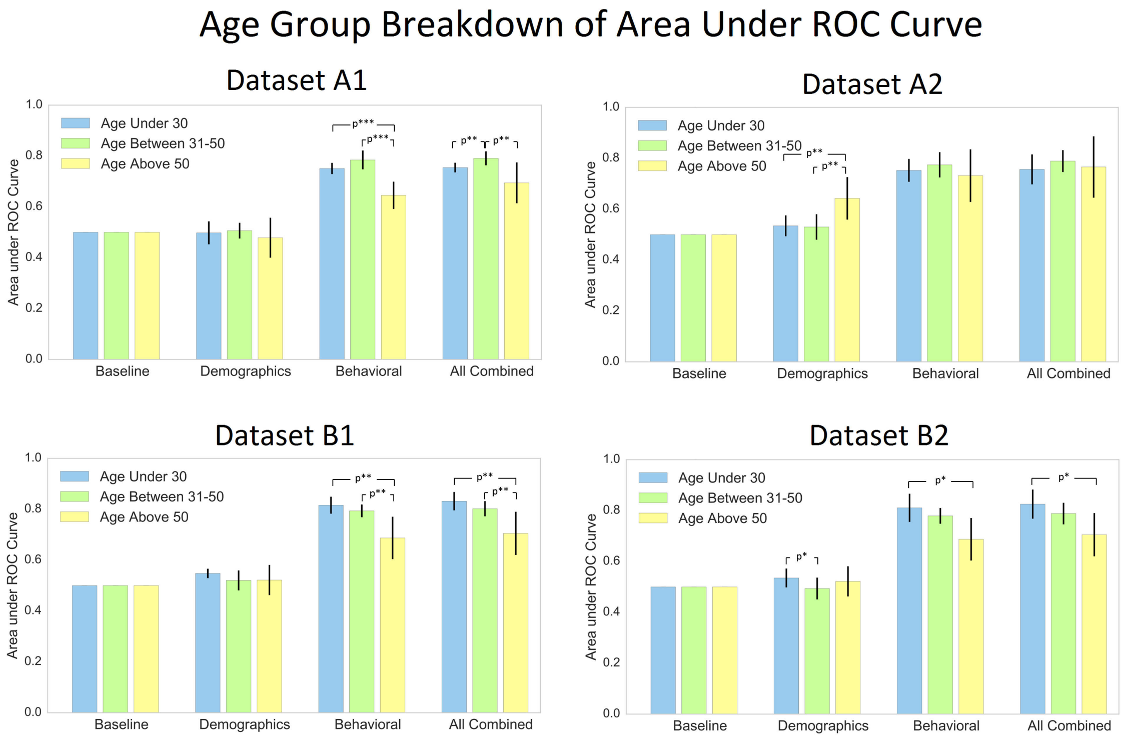

Figure A.4: Age group breakdown of feature set performance for data sets A2, B1, and B2. Area under ROC curve metric comparison of *demographic* features, *spatio-temporal and choice behavior* features, and the combination of both feature sets for the portions of data sets generated based on age groups under 30, between 30-50, and above 50. The length of the error bar corresponds to 1.0 standard deviation. The comparison is performed for the label *inac-full*. $p*$, $p**$, and $p***$ denote the significance levels with p values smaller than 0.05, 0.01, and 0.001, respectively.

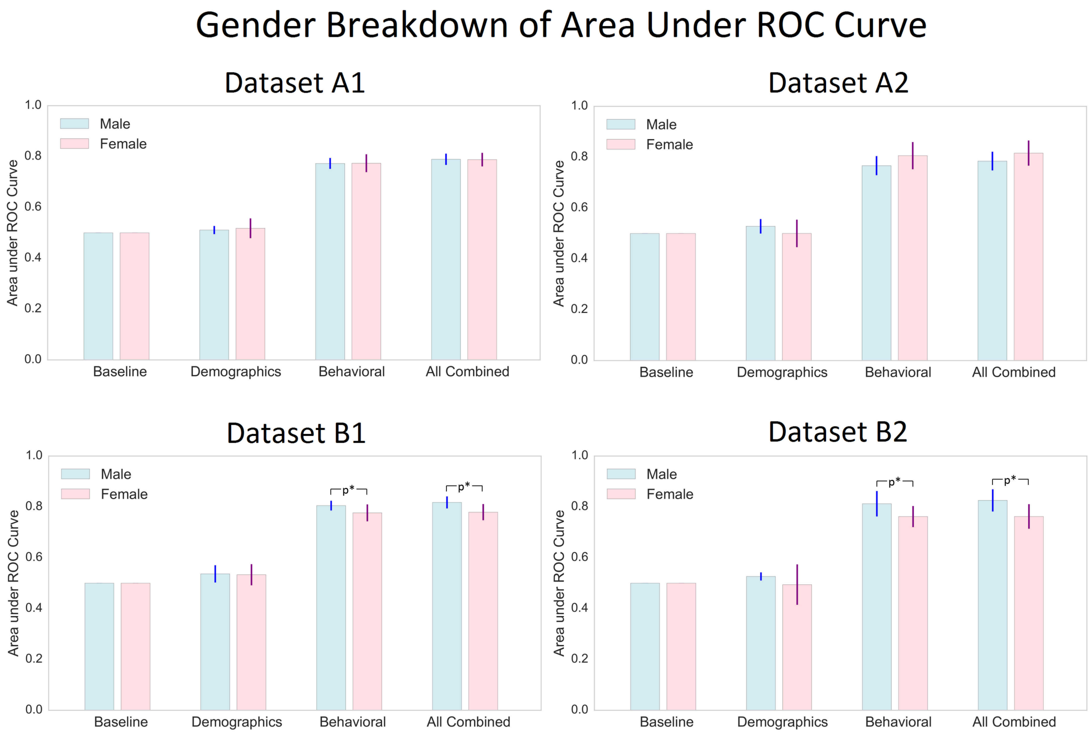

Figure A.5: Gender breakdown of feature set performance for data sets A2, B1, and B2. Area under ROC curve metric comparison of *demographic* features, *spatio-temporal and choice behavior* features, and the combination of both feature sets for the portions of data sets generated based on gender. The length of the error bar corresponds to 1.0 standard deviation. The comparison is performed for the label *inac-full*. $p*$ denotes the significance level with p value smaller than 0.05.

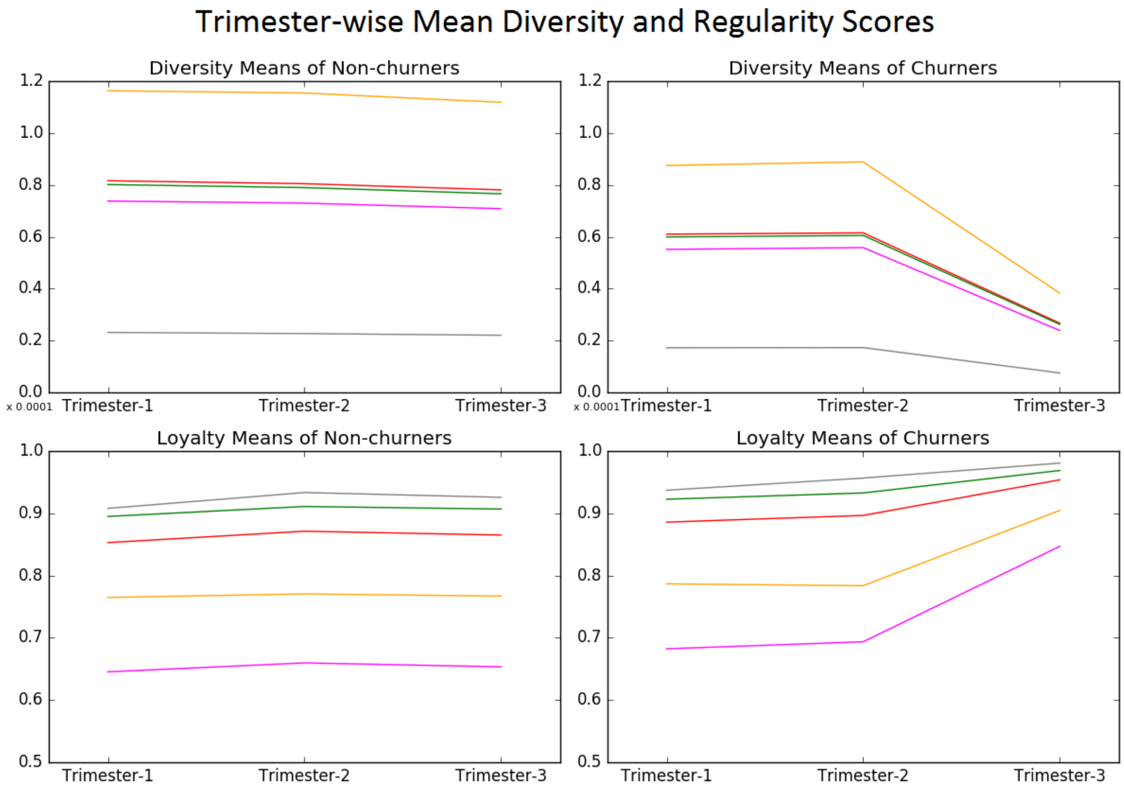

Figure A.6: Diversity and loyalty trends of churners and non-churners for data set A1. For the observation window, the diversity and loyalty feature values of churners have decreasing and increasing trend, respectively, whereas this is not the case for non-churners.

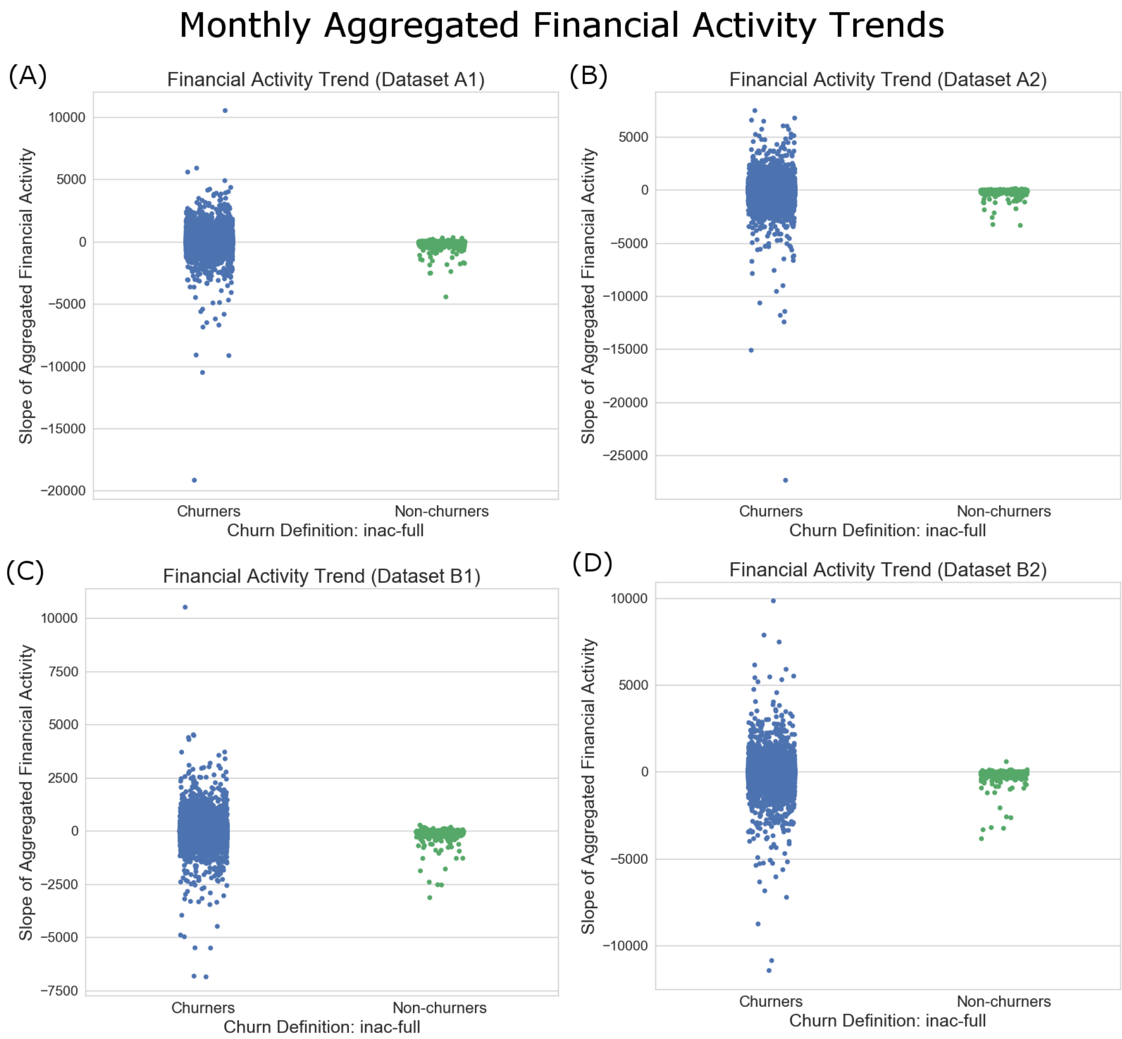

Figure A.7: Monthly financial activity trend plots of churners and non-churners. For all the data sets, churners and non-churners seem to have both positive and negative financial activity slopes with an exception for some of the churners with negative slopes.

APPENDIX B

FINANCIAL ANALYSIS CASE STUDY OBSERVATION DATA

B.1 Dataset

The data analyzed during the case study comprises credit card transaction features, customer demographics and financial metrics, and normalized features.

B.1.1 Credit Card Transaction Features

- **cust_id** Customer ID
- **tx_date** Transaction Date
- **tx_time** Transaction Time
- **tx_amount** Transaction Amount
- **merch_type** Merchant Type (MMC Code)
- **merch_id** Anonymized Merchant ID
- **online_tx** Online Transaction Flag (True or False)
- **exp_type** Expenditure Type
- **currency** Currency
- **pos_x** Longitude of the Transaction Location
- **pos_y** Latitude of the Transaction Location

B.1.2 Customer Demographic and Financial Metrics

- **marital_stat** Marital status of the customer making the transaction
- **edu_stat** Educational status of the customer making the transaction
- **job_type** The category of the job of the customer making the transaction
- **income** Monthly income of the customer making the transaction
- **age** Age of the customer making the transaction
- **bank_age** The number years spent as a customer of the bank of the customer making the transaction

- **cc_mean_risk** A score showing how risky the customer is from the bank's perspective
- **total_num_cc** Total number of the credit cards owned by the customer making the transaction
- **other_total_num_cc** Total number of credit cards issued by other banks and owned by the customer making the transaction
- **bank_cc_max_limit** The max credit card limit given to the customer by the bank during the data collection period
- **all_cc_max_limit** The max credit card limit given to the customer by all banks during the data collection period
- **total_transfer** The total within-bank wired transfer made by the customer making the transaction during the data collection period.
- **mean_transfer** The mean within-bank wired transfer amount made by the customer making the transaction during the data collection period
- **mean_eft** The total within-bank wired transfer made by the customer making the transaction during the data collection period
- **total_eft** The total between-banks EFT (Electronic Funds Transfer) amount made by the customer making the transaction during the data collection period
- **eft_entropy** A measure of the customer showing his/her diversity in issuing EFT to other banks. Higher value indicates that the customer distributed his/her EFT transfers evenly over the banks to which he/she performed an EFT transfer.
- **max_eft_bank** A categorical value showing the bank to which the customer performed EFT transaction most in terms of the money amount.
- **total_atm_withdrawal** The total amount of the ATM withdrawals made by the customer throughout the data collection period.
- **total_atm_deposit** The total amount of the ATM deposits made by the customer throughout the data collection period.
- **accept_percent** The rate of the total number of acceptance of the customers for various offerings made by the bank to the number of total offerings.
- **resp_score_mean** The average response score of the customers, calculated by the bank, indicating the responsiveness of the customers with respect to any interactions initiated by the bank.
- **resp_score_stddev** The standard deviation of the response scores of the customers calculated by the bank.

- **risk_score_mean**The average risk score of the customers, calculated by the bank.
- **mobile_total_transfer** The total within-bank wired transfer made by the customer over mobile application during the data collection period.
- **mobile_total_eft** The total between-banks EFT amount made by the customer over the mobile application during the data collection period.
- **mean_exp** The average credit card expenditures of the customer over the data collection period.
- **total_exp** The total credit card expenditures of the customer over the data collection period.

B.1.3 Normalized Features

- **tx_amount_n** The linear mapping of the tx_amount feature to the range [0,1].
- **income_n** The linear mapping of the income feature to the range [0,1].
- **age_n** The linear mapping of the age feature to the range [0,1].
- **bank_age_n** The linear mapping of the bank_age feature to the range [0,1].
- **total_num_cc_n** The linear mapping of the total_num_cc feature to the range [0,1].
- **bank_cc_max_limit_n** The linear mapping of the bank_cc_max_limit feature to the range [0,1].
- **all_cc_max_limit_n** The linear mapping of the all_cc_max_limit feature to the range [0,1].
- **total_transfer_n** The linear mapping of the total_transfer feature to the range [0,1].
- **mean_transfer_n** The linear mapping of the mean_transfer feature to the range [0,1].
- **mean_eft_n** The linear mapping of the mean_eft feature to the range [0,1].
- **total_eft_n** The linear mapping of the total_eft feature to the range [0,1].
- **total_atm_withdrawal_n** The linear mapping of the total_atm_withdrawal feature to the range [0,1].
- **total_atm_deposit_n** The linear mapping of the total_atm_deposit feature to the range [0,1].
- **mobile_total_transfer_n** The linear mapping of the mobile_total_transfer feature to the range [0,1].

- **mobile_total_eft_n** The linear mapping of the mobile_total_eft feature to the range [0,1].
- **mean_exp_n** The linear mapping of the mean_exp feature to the range [0,1].
- **total_exp_n** The linear mapping of the total_exp feature to the range [0,1].

B.1.4 Data Pruning

The data has been pruned in order to prevent various kinds of distortions in the visualizations such as stacking of data points to a relatively small area due to extremely high-valued points. The data has been pruned with respect to conditions listed below. For any condition, no more than 0.5% of the data has been pruned. As a result of pruning process approximately 4% of the data have been removed.
All the transaction tuples satisfying *all* of the following conditions formed the dataset used in the study.

- income <10000
- total_transfer <1000000
- mean_transfer <100000
- mean_eft <100000
- total_eft <1500000
- total_atm_withdrawal <80000
- total_atm_deposit <250000
- mobile_total_transfer <500000
- mobile_total_eft <800000
- mean_exp <5000

B.2 Tasks

The tasks and example workflow utilizing these tasks are discussed in detail in Section 4.2. Below is the mapping between our high-level tasks and the task abstractions of Yi et al. [9].

B.3 Selectable Features

Selectable features in the difference view of the DimXplorer are listed in Table B.2 and in "Normalized Features" subsection.

Task #	High-level Task	Abstract Actions
Task-1	Automatic Feature-based Subsegment Generation	selection, filtering, clustering, comparison
Task-2	User-defined Subsegment Definition	selection, filtering, clustering, comparison
Task-3	Segment Composition	retrieve, reconfigure
Task-4	Segment Fine-tuning	retrieve, filtering, comparison, selection
Task-5	Composed Segment Description	retrieve, elaborate

Table B.1: Mapping between high-level tasks and abstract actions based on Yi et al.'s [9] taxonomy.

B.3.1 Feature Set Selection

The set of features selected by the analysts during the analysis sessions were recorded and represented in Table B.3. Moment shows the time elapsed during the start of the corresponding sub-session until a feature set selection made by the analysts. Duration represents the time elapsed that the analysts spent on working on the selected feature set. The numbers in the feature set selection column are the index numbers of the features listed in Table B.2. The progression below, recorded as multiples of 25, shows the maximum percentage of the whole data contributed to the calculations as of the moment the analysts renounced the selected feature set. Insights-Questions-Hypotheses column shows the total number of insights, questions, and hypotheses arose during the time analysts worked with corresponding feature set.

Sub-session	Moment	Duration	FSS	PB	IQH
2-1	1:23	2:10	1,3,7	75	0-2-0
2-2	0:05	7:11	1,4,8	75	1-0-0
2-2	9:54	2:52	1,2,8,18	25	1-1-0
2-2	17:30	7:10	1,9	75	2-0-0
2-3	00:00	4:45	1,9	100	1-0-1
2-3	5:58	4:19	3,6	50	1-0-1
2-3	10:23	9:32	3,12,17,18	100	2-2-0
2-4	5:14	4:55	1,3	50	0-2-0
2-4	10:09	1:35	1,3,6,12	25	0-0-0
2-4	11:59	5:26	1,3,9,10,15	25	0-0-0
2-4	20:32	9:50	3,5,7	75	2-1-0
3-1	2:02	4:03	22,23,26	25	1-0-0

3-1	27:00	6:05	22,23,27,34,35	50	0-2-0
3-1	33:05	4:15	23,27,34,35	25	1-0-0
3-1	37:20	17:10	23,27,32,33	75	2-0-0
3-2	0:00	4:03	23,27,29,33	25	0-0-0
3-2	4:03	4:37	12,15,16,24,25	50	0-0-1
3-2	12:40	4:29	12,15,16,36	25	0-0-1
3-2	17:09	3:22	12,15,16,37	50	0-0-1
4-1	2:00	6:06	25,31,33	NP	0-0-1
4-1	10:42	8:53	12,15,26,36	NP	0-3-0
4-1	24:34	25:34	22,23,25	NP	0-0-0
4-1	25:34	11:35	22,23,25,26	NP	1-0-0
4-2	19:03	11:54	25,32,33,36,37	NP	1-1-0

Table B.3: Feature set selections performed by analysts during the course of the analysis sessions. NP stands for ”non-progressive,” FSS stands for “Feature Set Selection,” PB stands for “Progress Below,” and IQH stands for “Insights-Questions-Hypotheses.”

B.3.2 Insights-Questions-Hypotheses

Important take-away points were extracted from the analysis video records and are listed as follows.

Sub-session	Moment	Type of Inference	Content
Session 2-1	1:22	Question	(Q-1) Why do customers with high salary and credit card limits make transactions mainly around Istanbul? Why don’t they spend their money in other regions? Why do the customers seeming to be traveling more and having less salary travel more compared to the ones with higher salary?
Session 2-1	7:01	Quote	(Qu-1) Analyst-1: “Let’s try some other demographic features as this selection seems like not going to bring new patterns. We can generate so many new hypotheses in a very short time without waiting for the whole calculation to end.” Researcher: “Do you think that making different feature set selections could be distracting?” Analyst-1: “No. Instead, visualization is quite engaging as we don’t have to wait for even a moment to get some initial results.”
Session 2-2	2:46	Insight	(I-1) Customers with higher salary and not using credit card frequently seems to make most of transactions in Istanbul (might mean that they don’t travel a lot)
Session 2-2	5:05	Insight	(I-2) Why do customers with high salary and credit card limits make transactions mainly around Istanbul? Why don’t they spend their money in other regions? Why do the customers seeming to be traveling more and having less salary travel more compared to the ones with higher salary?
Session 2-2	5:26	Brief Interview	(Qu-2) Researcher: “How long would it take to reach to these insights if you were to use your own methods?” Analyst-2: “That would require me to form a database query selecting the transactions of the customers with high salary and compare the means of the transactions of those customers with mean of all transactions.” Analyst-1: “And the total process could take up 4-5 queries. However, I’m not sure whether we could come up with that question in a very short time. Not to mention the time that would take for queries to run over the transaction database. [Progressive] Visualization seems to help a lot.”

Session 2-2	13:35	Quote	(Qu-3) Analyst-1: "It seems like the clustering will not change. ... Almost all of the data has been calculated, let's switch to some other set."
Session 2-2	24:45	Insight	(I-3) Analyst-1: The customers working with other banks seem to be more profitable ones as their financial metrics draws a better picture (i.e. higher transfers, EFT, higher response score, etc.).
Session 2-2	26:37	Insight	(I-4) Customers with one credit card and high credit card limit seem to be have lower risk score.
Session 2-3	00:47	Testing	(Te-1) Hypothesis (insight) I-3 has been rejected. Customers with 2 or 3 credit cards and low credit card limits seem to represent low financial profile.
Session 2-3	03:41	Insight	(I-5) As the age of the customers increase the more likely their expenditures are grouped in the Istanbul area.
Session 2-3	14:12	Question	(Q-2) What is the relationship with the EFT entropy and the total credit number of a customer?
Session 2-3	14:33	Insight	(I-6) The EFT entropy seems to be independent from the total number of credit card.
Session 2-3	15:49	Insight	(I-7) Customers working with other banks seem to be managing their credit cards and their assets via those other banks, not via our bank. (Analyst-1)
Session 2-3	17:01	Question	(Q-3) Why do customers having more than one credit card have lower EFT entropy?
Session 2-4	5:17	Question	(Q-4) Can the transaction amount and the age of the customer be a good discriminator in terms of facilitating good segmentation? Can the transaction amount and the age of the customer be related to money deposit or withdrawal amount?
Session 2-4	11:24	Quote	(Qu-4) Analyst-4 accidentally deleted all the selections including the features, which immediately stopped all ongoing clustering and PCA calculations. Analyst-3: "Sigh, all the computations have gone." Analyst-1 "No worries. Make the selection again, please. They (clusters) should show up soon." However, after this dialog, they ended up starting with a slightly different feature set.
Session 2-4	12:06	Brief Interview	(Qu-5) Researcher: "To what extend do you think this tool (progressive visual analytic system) can effect your current analysis processes?" Analyst-1: "Recently we changed our policy which could be summarized as 'work on old but important hypotheses' to a stance encouraging our analyst teams for the production of new hypotheses. This approach, I believe, can take us one step forward as we will be trying out new alternatives. It seems like this tool is a good fit, at least conceptually, as our analysis is now more data-driven rather than goal-driven. ... (16:03) It is quite a new concept for my team to have almost-real-time response from a clustering calculation. ... (17:09) Typically, given the usual dataset size and analysis goals, we spare more than a day for clustering a model."
Session 2-4	21:18	Insight	(I-8) There seems to be a positive correlation with the age and income of the customer.
Session 2-4	25:47	Question	(Q-5) Do the transactions mostly done close to coastal line of the city form a cluster based on the feature set with age, mean risk and maximum credit card limit?
Session 2-4	29:25	Insight	(I-9) Online transactions seem to be clustered around the west half of the city.
Session 3-1	2:12	Quote	(Qu-6) Just after a new feature set has been selected, Analyst-3: "I think we can start as we have a good view of the clusters."
Session 3-1	3:10	Quote	(Qu-7) The team tried a new set of features and immediately observed a 'good' separation of data points. However, after only 15-20 seconds, the separation dramatically changed. After this, Analyst-2: "Well, I think waiting for a while might be a good thing."
Session 3-1	3:35	Insight	(I-10) As expected, there is a correlation between the credit card limit and income.
Session 3-1	27:29	Question	(Q-6) Are the features age and income together predictor of credit card limit or total transfer amount?
Session 3-1	30:02	Question	(Q-7) How do the clusters differ from each other in terms of total transfer amount?
Session 3-1	40:47	Insight	(I-11) The number of the credit card of the customer and the amount of expenditures made by him/her are not correlated.

Session 3-1	50:47	Observation	(Qu-8) Analyst-3 started to pose his idea and requested one of his fellow analyst to modify a particular visualization parameter. As soon as he noticed that visualization was changing on the screen, he said: “Well, let’s wait for a moment to let it settle down.” They waited for a couple of seconds so that DimXplorer could processed some more data.
Session 3-1	51:03	Design Informant	(Qu-9) During the course of the analysis, Analyst-3: “I’ve just seen a high response score for the selected cluster, but it has just gone away.” As the clustering algorithm continued to its calculations, the relevant data points have moved to other clusters changing the pattern Analyst-3 previously discovered: “Wouldn’t it be nice to have a button that simply pauses the progressive visualization?”
Session 3-2	08:11	Hypothesis	(H-1) The customers with low EFT entropy are also the ones with high response scores.
Session 3-2	08:40	Technical Problem	The analysis session has been paused. The session continued after approximately 4 minutes.
Session 3-2	14:59	Testing	(Te-2) H-1 has been verified. Customers with high response scores were also the ones with low EFT entropy. The reverse was also true, as it was clear on the comparison with PCA and cluster views.
Session 3-2	19:02	Hypothesis	(H-2) The customers with high response score are also the ones with high acceptance rate.
Session 3-2	20:31	Brief Interview	(Qu-10) Researcher: “I have been observing that you are making quick and brief inferences from the data.” They were able to interactively try out many different filters and features for the clustering in a short time. Previously, they reported that these kind of changes would actually require a work-day. “How do you think these brief findings can contribute to your actual analysis activities?” Analyst-3: “Well, we can further elaborate on the hypotheses that we derived from this tool. For example, we have seen that the customers with high EFT entropy tend to have low response scores. Well, we could verify this in more detail and, maybe, we can also investigate the outliers, the customers with high response scores and high EFT entropy, to find out why they discriminate from the majority.” Analyst-2: “For example, if they are not credit card customers, I would try to ‘get deep with them’ with credit card offers.” Analyst-3: “Moreover, if they make more EFT than they actually need, I would recommend them to apply for automatic payment services.” Analyst-4: “It could also help us to locate a customer group that we have never been aware of. In that case, we could devise new actions for that new particular group.”
Session 3-2	22:07	Brief Interview	(Qu-11) Researcher: “How do you usually go about inferring hypotheses during your daily analysis activities?” Analyst-4: “Experience and limited visualizations usually form a basis for our hypotheses extraction processes.” Analyst-3: “Well, first of all, we couldn’t get all those results that quickly. As we get some results from our data analytic models, we become aware of even more other options that we feel like we need to try. Well, we don’t have that much time.” Analyst-2: “And this situation may lead to indolence, particularly when your models take so much time to be calculated.” Analyst-3: “And [when working on this tool] we are able to look at the model altogether which is way better than one person working on his own.” Analyst-4: “Due to time limitations, we regularly eliminate the minor or insignificant cases, but we are able try them here in a short time.” Researcher: “Aren’t there situations that you work on the models as a team?” Analyst-2: “Of course there are. For example, one of us develops a model and we brainstorm on it. However, we usually get the feeling that the analyst working on that model must have spent much time on it and we cannot easily give up and try new models. However, during study I noticed that we can reject our own models as quickly as we generate them. It is easy to reject them on this tool.” Analyst-4: “Analysis, especially exploratory one, is usually like a lottery, the more ticket you have, the more like you win the prize. With this tool, we have many more tickets.”
Session 4-1	2:30	Injection	DimXplorer has been configured to work in non-progressive mode. For any calculation, if the calculations are completed in less than 60 seconds, the DimXplorer will respond in at least 60 seconds.
Session 4-1	2:53	Observation	During the delay (60 seconds) that was intentionally inserted to simulate non-progressive analytics, Analyst-4 started to talk about what they would be expecting at the end of the clustering calculation.
Session 4-1	2:58	Hypothesis	(H-3) There is a positive correlation between EFT entropy and total number of credit cards per customer.

Session 4-1	12:17	Observation	At the end of the NP clustering calculation they were satisfied with the result of the clustering; however, they noticed that they forgot to change the feature on the axes of cluster small multiples. Instead of having the result to be recalculated, they decided to continue to the analysis with the old view.
Session 4-1	17:17	Question	(Q-8) How does the campaign acceptance rate of the customers change with high EFT entropy and high number of credit cards?
Session 4-1	17:36	Question	(Q-9) There are customers with high EFT entropy but also falling into one of the subgroups with high and low acceptance rate. What causes this discrimination?
Session 4-1	19:01	Question	(Q-10) Is there a positive correlation between the credit card number and the credit card limit that was granted by our bank?
Session 4-1	22:29	Quote	(Qu-12) Analyst-3: "I think it might also be beneficial to be able to set the step size of the progression. For example, it can be distracting to look at an ever-changing visualization. Can we follow a different approach, say, process a certain fraction of the data at each step of calculation. This way we can have some time to talk about intermediate results."
Session 4-1	32:24	Insight	(I-12) Partial answer has been found for the Q-10. The credit card limit granted by the bank varies depending on the total number of credit cards the customers possess. The 'age' of the customers seems to be a discriminating factor; however, this needs to be verified in a different context.
Session 4-2	0:20	Quote	(Qu-13) Analyst-4: "This new [non-progressive] version of the calculations are better from my point of view. In the progressive version, I was having hard time to mention about the patterns and express my ideas as the clusters were changing so quickly." Analyst-3: "Rather than having the results continuously updated, it might be better to have them with lags, step by step, with a considerable amount of time between the steps. By the way, I suspect that the sampling mechanism gets the data rows randomly. If it did, I would expect a little more uniform distribution across the clusters. However, even if we worked on the same set for a long time, we could barely see such a result." Analyst-4: "Even getting updates at every five minutes is also a great improvement for us as we get a clustering model developed in ten minutes or so with small datasets in commercial [monolithic, non-progressive] tools." Analyst-3: "I believe it would be great to define the amount of the data to be calculated in each step of the calculation. If I know the size of my data, I would divide it, say, into 10 chunks and set progression step accordingly."
Session 4-2	11:34	Injection	DimXplorer has been configured in a way that it will respond to calculations in at least 20 seconds.
Session 4-2	19:18	Question	(Q-11) Is there a correlation between the number of credit cards and expenditure, and money withdrawal from ATMs?
Session 4-2	21:51	Insight	(I-13) The customers using their credit cards more often seem to withdraw money from ATMs less often.

Table B.4: Insights derived, questions formed, hypotheses generated by the analysts during the case study.

B.3.3 Self-reported Advantages of Progressive Approach on the Analysis Process

Subjective evaluations of progressive visualization and analytics were collected from the analysts verbally. Below is the summarized transcript relevant to the progressive aspect of the tool. *Brief Interviews* and notable *Quotes* listed on Table B.4 are also additional sources used as subjective measures that were not listed below.

Analyst-3: "The progressive visualization provides instant feedback on the filtering of the data. We usually filter our data with database queries and if we can think

Feature Code	Feature Name	Feature Code	Feature Name
1	tx_amount	21	tx_amount_n
2	income	22	income_n
3	age	23	age_n
4	bank_age	24	bank_age_n
5	cc_mean_risk	25	total_num_cc_n
6	bank_cc_max_limit	26	bank_cc_max_limit_n
7	all_cc_max_limit	27	all_cc_max_limit_n
8	total_transfer	28	total_transfer_n
9	mean_transfer	29	mean_transfer_n
10	mean_eft	30	mean_eft_n
11	total_eft	31	total_eft_n
12	eft_entropy	32	total_atm_withdrawal_n
13	total_atm_withdrawal	33	total_atm_deposit_n
14	total_atm_deposit	34	mobile_total_transfer_n
15	accept_percent	35	mobile_total_eft_n
16	resp_score_mean	36	mean_exp_n
17	resp_score_stddev	37	total_exp_n
18	mobile_total_transfer		
19	mean_exp		
20	total_exp		

Table B.2: Selectable features.

better queries during the data retrieval, we are not able to apply it immediately. When the data is big, this becomes overwhelming." (Qu-13)

Analyst-2: "To add more on that, even the effort in finding the proper filter to the data can sometimes be time consuming. With this tool, I felt like I had more ideas, and maybe due to the visual matter we saw, we worked hands-on together and carried more fruitful brainstorming sessions." (Qu-14)

Researcher: "Do you do 'hands-on team work' together as a part of your daily analytic tasks?"

Analyst-3: "Well, it is usually quite impractical as we frequently have to wait for a certain period of time even for simple queries." (Qu-15)

Analyst-1: "What we look at during the analyses is usually the numbers supported with traditional visualizations. At the end of the day, if we happen to doubt our model and say, ‘What if we tried this?’, well, this instantly become a story for another day. And start all over again!” (Qu-16)

Analyst-2: “Well, from this perspective, we had many chances to try out different ideas today. And the visualization also facilitated emergence of many ideas.” (Qu-17)

Analyst-4: “And, we try many clustering variants throughout an analysis period. The tool enabled us to try many options today. Well, I think I’ve become experience with this data more quickly than I used to do in my other analysis activities.” (Qu-18)

Analyst-3: “During the analysis, we feel obliged to look at to the portion of the data corresponding to the analysis goals and it is not quite often to run into different findings. With the visualization, we could notice also some interesting patterns that were not related to what we were supposed to analyze.” (Qu-19)

Researcher: “Did you ever feel interrupted, distracted or disengaged during your analysis sessions today?”

Analyst-1: “No, we were so engaged that even the computer couldn’t catch up with us [laughs].” The analyst implied the crashes of our visual analytics tool during the case study. Other analysts shortly answered that they did not feel interrupted. (Qu-20)

Bibliography

[1] K. A. Cook and J. J. Thomas, "Illuminating the path: The research and development agenda for visual analytics," 2005.

[2] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler, "Visual analytics: Scope and challenges," *Visual Data Mining*, pp. 76–90, 2008.

[3] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card, "The cost structure of sensemaking," in *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, ACM, 1993, pp. 269–276.

[4] B. Dervin, *An overview of sense-making research: Concepts, methods, and results to date*. The Author, 1983.

[5] D. Norman, *The design of everyday things: Revised and expanded edition*. Basic Books (AZ), 2013.

[6] *Proceedings of the 2017 chi conference on human factors in computing systems*. Available from: `http://dl.acm.org/citation.cfm?id=3025453&preflayout=flat`, Accessed: 2017-09-17.

[7] A. Dix, "Human-computer interaction," in *Encyclopedia of database systems*, Springer, 2009, pp. 1327–1331.

[8] Z. Liu and J. Stasko, "Mental models, visual reasoning and interaction in information visualization: A top-down perspective," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 999–1008, 2010.

[9] J. Yi, Y. ah Kang, J. Stasko, and J. Jacko, "Toward a deeper understanding of the role of interaction in information visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 6, pp. 1224–1231, 2007.

[10] T. Munzner, "A nested model for visualization design and validation," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 921–928, 2009.

[11] H. Haddadi, "Human-data interaction," *Encyclopedia of Human Computer Interaction*, 2016.

[12] R. Mortier, H. Haddadi, T. Henderson, D. McAuley, and J. Crowcroft, "Human-data interaction: The human face of the data-driven society," 2014.

[13] P. Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization," 2009.

[14] N. Andrienko and G. Andrienko, *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media, 2006.

[15] S. K. Card, G. G. Robertson, and J. D. Mackinlay, "The information visualizer, an information workspace," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '91, New Orleans, Louisiana, United States: ACM, 1991, pp. 181–186, ISBN: 0-89791-383-3.

[16] K. J. W. Craik, *The nature of explanation*. CUP Archive, 1967, vol. 445.

[17] *Interaction design foundation*. Available from: `https://www.interaction-design.org/`, Accessed: 2017-12-04.

[18] B. Gaver, T. Dunne, and E. Pacenti, "Design: Cultural probes," *interactions*, vol. 6, no. 1, pp. 21–29, 1999.

[19] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, IEEE, 2005, pp. 111–117.

[20] J.-D. Fekete, "Visual analytics infrastructures: From data management to exploration," *Computer*, vol. 46, no. 7, pp. 22–29, 2013.

[21] J.-D. Fekete and C. Silva, "Managing data for visual analytics: Opportunities and challenges.," *IEEE Data Eng. Bull.*, vol. 35, no. 3, pp. 27–36, 2012.

[22] C. Turkay, E. Kaya, S. Balcisoy, and H. Hauser, "Designing progressive and interactive analytics processes for high-dimensional data analysis," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 131–140, 2017.

[23] J.-D. Fekete, "Progressivis: A toolkit for steerable progressive analytics and visualization," in *1st Workshop on Data Systems for Interactive Analysis*, 2015, p. 5.

[24] J. E. Brophy and T. L. Good, "Teachers' communication of differential expectations for children's classroom performance: Some behavioral data.," *Journal of educational psychology*, vol. 61, no. 5, p. 365, 1970.

[25] V. K. Singh, B. Bozkaya, and A. Pentland, "Money walks: Implicit mobility behavior and financial well-being," *PloS one*, vol. 10, no. 8, e0136628, 2015.

[26] J. Mumm and B. Mutlu, "Human-robot proxemics: Physical and psychological distancing in human-robot interaction," in *Proceedings of the 6th international conference on Human-robot interaction*, ACM, 2011, pp. 331–338.

[27] R. L. Mack and J. Nielsen, "Usability inspection methods: Executive summary," in *Human-computer interaction*, Morgan Kaufmann Publishers Inc., 1995, pp. 170–181.

[28] M. Tory and T. Moller, "Evaluating visualizations: Do expert reviews work?" *IEEE computer graphics and applications*, vol. 25, no. 5, pp. 8–11, 2005.

[29] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE transactions on visualization and computer graphics*, vol. 18, no. 9, pp. 1520–1536, 2012.

[30] C. North, "Toward measuring visualization insight," *Computer Graphics and Applications, IEEE*, vol. 26, no. 3, pp. 6–9, 2006.

[31] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko, "Effectiveness of animation in trend visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 6, pp. 1325–1332, 2008.

[32] P. Saraiya, C. North, and K. Duca, "An insight-based methodology for evaluating bioinformatics visualizations," *IEEE transactions on visualization and computer graphics*, vol. 11, no. 4, pp. 443–456, 2005.

[33] A. Strauss and J. Corbin, "Discovery of grounded theory," 1967.

[34] M. D. Myers, *Qualitative research in business and management.* Sage, 2013.

[35] A. M. Thomson and J. L. Perry, "Collaboration processes: Inside the black box," *Public administration review*, vol. 66, no. s1, pp. 20–32, 2006.

[36] M. Li, S. Gao, and C. C. Wang, "Real-time collaborative design with heterogeneous cad systems based on neutral modeling commands," *Journal of Computing and Information Science in Engineering*, vol. 7, no. 2, pp. 113–125, 2007.

[37] J. Hou, C. Su, Y. Su, and W. Wang, "A methodology of knowledge management based on ontology in collaborative design," in *Intelligent Information Technology Application, 2008. IITA'08. Second International Symposium on*, IEEE, vol. 2, 2008, pp. 409–413.

[38] C. Bortolaso, M. Oskamp, G. Phillips, C. Gutwin, and T. Graham, "The effect of view techniques on collaboration and awareness in tabletop map-based tasks," in *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ACM, 2014, pp. 79–88.

[39] X. Chen, J. Fuh, Y. Wong, Y. Lu, W. Li, and Z. Qiu, "An adaptable model for distributed collaborative design," *Computer-Aided Design and Applications*, vol. 2, no. 1-4, pp. 47–55, 2005.

[40] X. Song, W. Dou, and J. Zhu, "Implementation of collaborative design system upon heterogeneous cad systems using a feature-based mapping set," in *Computer Supported Cooperative Work in Design (CSCWD), 2010 14th International Conference on*, IEEE, 2010, pp. 510–515.

[41] Y. Rahmawati, C. Utomo, N. Anwar, P. Setijanti, and C. B. Nurcahyo, "An empirical model for successful collaborative design towards sustainable project development," *Journal of Sustainable Development*, vol. 7, no. 2, p. 1, 2014.

[42] S. Son, S. Na, K. Kim, and S. Lee, "Collaborative design environment between ecad and mcad engineers in high-tech products development," *International Journal of Production Research*, vol. 52, no. 20, pp. 6161–6174, 2014.

[43] C. Anumba, O. Ugwu, L. Newnham, and A Thorpe, "Collaborative design of structures using intelligent agents," *Automation in construction*, vol. 11, no. 1, pp. 89–103, 2002.

[44] G. C. Gabriel and M. L. Maher, "Coding and modelling communication in architectural collaborative design," *Automation in construction*, vol. 11, no. 2, pp. 199–211, 2002.

[45] J. J. Shah, N. Vargas-Hernandez, J. D. Summers, and S. Kulkarni, "Collaborative sketching (c-sketch)—an idea generation technique for engineering design," *The Journal of Creative Behavior*, vol. 35, no. 3, pp. 168–198, 2001.

[46] J. S. Linsey, E. Clauss, T Kurtoglu, J. Murphy, K. Wood, and A. Markman, "An experimental study of group idea generation techniques: Understanding the roles of idea representation and viewing methods," *Journal of Mechanical Design*, vol. 133, no. 3, p. 031 008, 2011.

[47] *Sketchup*, `https://www.sketchup.com/`, Accessed: 2017-06-13, 2017.

[48] *Autocad*, `https://www.autodesk.com/products/autocad/overview`, Accessed: 2017-06-13, 2017.

[49] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer, "World-stabilized annotations and virtual scene navigation for remote collaboration," in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, ACM, 2014, pp. 449–459.

[50] M. Fan, A. N. Antle, C. Neustaedter, and A. F. Wise, “Exploring how a co-dependent tangible tool design supports collaboration in a tabletop activity,” in *Proceedings of the 18th International Conference on Supporting Group Work*, ACM, 2014, pp. 81–90.

[51] A. Lee, H. Chigira, S. K. Tang, K. Acquah, and H. Ishii, “Annoscape: Remote collaborative review using live video overlay in shared 3d virtual workspace,” in *Proceedings of the 2nd ACM symposium on Spatial user interaction*, ACM, 2014, pp. 26–29.

[52] D. Leithinger, S. Follmer, A. Olwal, and H. Ishii, “Physical telepresence: Shape capture and display for embodied, computer-mediated remote collaboration,” in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, ACM, 2014, pp. 461–470.

[53] Y Rahmawati, C Utomo, N Anwar, C. Nurcahyo, and N. Negoro, “Theoretical framework of collaborative design issues,” *Jurnal Teknologi*, vol. 70, no. 7, pp. 47–53, 2014.

[54] S. D. Scott, K. D. Grant, and R. L. Mandryk, “System guidelines for co-located, collaborative work on a tabletop display,” in *ECSCW 2003*, Springer, 2003, pp. 159–178.

[55] C. Shen, K. Ryall, C. Forlines, A. Esenther, F. D. Vernier, K. Everitt, M. Wu, D. Wigdor, M. R. Morris, M. Hancock, *et al.*, “Informing the design of direct-touch tabletops,” *IEEE computer graphics and applications*, vol. 26, no. 5, pp. 36–46, 2006.

[56] C. Shen, K. Everitt, and K. Ryall, “Ubitable: Impromptu face-to-face collaboration on horizontal interactive surfaces,” in *International Conference on Ubiquitous Computing*, Springer, 2003, pp. 281–288.

[57] P. Girard and V. Robin, “Analysis of collaboration for project design management,” *Computers in industry*, vol. 57, no. 8, pp. 817–826, 2006.

[58] A. S. Vivacqua, A. C. B. Garcia, and A. Gomes, “Boo: Behavior-oriented ontology to describe participant dynamics in collocated design meetings,” *Expert Systems with Applications*, vol. 38, no. 2, pp. 1139–1147, 2011.

[59] H. Seichter, “Augmented reality and tangible interfaces in collaborative urban design,” *Computer-Aided Architectural Design Futures (CAADFutures) 2007*, pp. 3–16, 2007.

[60] X. Wang, “Augmented reality in architecture and design: Potentials and challenges for application,” *International Journal of Architectural Computing*, vol. 7, no. 2, pp. 309–326, 2009.

[61] J. Underkoffler and H. Ishii, "Urp: A luminous-tangible workbench for urban planning and design," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM, 1999, pp. 386–393.

[62] D. Leithinger and H. Ishii, "Relief: A scalable actuated shape display," in *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, ACM, 2010, pp. 221–222.

[63] I. Hurkxkens and G. Munkel, "Speculative precision: Combining haptic terrain modelling with real-time digital analysis for landscape design," *Digital Landscape Architecture*, 2014.

[64] T. Mezher, M. A. Abdul-Malak, I. Ghosn, and M. Ajam, "Knowledge management in mechanical and industrial engineering consulting: A case study," *Journal of management in engineering*, vol. 21, no. 3, pp. 138–147, 2005.

[65] A. Keyvanfar, M. Z. A. Majid, A. Shafaghat, H. Lamit, A. Talaiekhozan, M. W. Hussin, C. T. Lee, R. B. M. Zin, and M. A. Fulazzaky, "Application of a grounded group decision-making (ggdm) model: A case of micro-organism optimal inoculation method in biological self-healing concrete," *Desalination and Water Treatment*, vol. 52, no. 19-21, pp. 3594–3599, 2014.

[66] M. Majid, W. Zakaria, H. Lamit, A Keyvanfar, A Shafaghat, and E. Bakti, "Construction information systems for executive management in monitoring work progress," *Advanced Science Letters*, vol. 15, no. 1, pp. 169–171, 2012.

[67] M.-L. Chiu, "An organizational view of design communication in design collaboration," *Design studies*, vol. 23, no. 2, pp. 187–210, 2002.

[68] F. Sudweeks and S. Rafaeli, "How do you get a hundred strangers to agree?" *Computer Networking and Scholarly Communication in the Twenty-first Century, New York: State University of New York*, pp. 115–136, 1996.

[69] *Kinect for xbox one*, `https://en.wikipedia.org/wiki/Kinect`, Accessed: 2017-04-14, 2017.

[70] C. Reas and B. Fry, "Processing: Programming for the media arts," *AI & SOCIETY*, vol. 20, no. 4, pp. 526–538, 2006.

[71] *Fingertracker librray for processing*, `http://makematics.com/code/FingerTracker/`, Accessed: 2017-04-23, 2017.

[72] *Pixelflow: A processing library for high performance gpu-computing (glsl)*, `http://thomasdiewald.com/processing/libraries/pixelflow/`, Accessed: 2017-04-23, 2017.

[73] J. Blomberg, J. Giacomi, A. Mosher, and P. Swenton-Wall, "Ethnographic field methods and their relation to design," *Participatory design: Principles and practices*, pp. 123–155, 1993.

[74] *Aster gdem v2.* Available from: `asterweb.jpl.nasa.gov/gdem.asp`, Accessed: 2015-01-04.

[75] *Openstreetmap.* Available from: `www.osm.org`, Accessed: 2015-01-04.

[76] *Kinect.* Available from: `https://www.microsoft.com/en-us/kinectforwindows/`, Accessed: 2015-01-04.

[77] R. Rosenthal and R. L. Rosnow, *Essentials of behavioral research: Methods and data analysis.* McGraw-Hill Humanities Social, 1991.

[78] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction.* John Wiley & Sons, 2010.

[79] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen, "Collaborative visualization: Definition, challenges, and research agenda," *Information Visualization*, vol. 10, no. 4, pp. 310–326, 2011.

[80] R. M. Baecker, *Readings in groupware and computer-supported cooperative work: Assisting human-human collaboration.* Elsevier, 1993.

[81] M. C. Chuah and S. F. Roth, "Visualizing common ground," in *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, IEEE, 2003, pp. 365–372.

[82] F. B. Viegas, M. Wattenberg, F. Van Ham, J. Kriss, and M. McKeon, "Manyeyes: A site for visualization at internet scale," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 6, pp. 1121–1128, 2007.

[83] *Tableau public*, `https://public.tableau.com/s/`, Accessed: 2015-09-28.

[84] R. Lissermann, J. Huber, M. Schmitz, J. Steimle, and M. Mühlhäuser, "Permulin: Mixed-focus collaboration on multi-view tabletops," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2014, pp. 3191–3200.

[85] C. Bortolaso, M. Oskamp, T. Graham, and D. Brown, "Ormis: A tabletop interface for simulation-based training," in *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, ACM, 2013, pp. 145–154.

[86] Y. Jansen, P. Dragicevic, and J.-D. Fekete, "Evaluating the efficiency of physical visualizations," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2013, pp. 2593–2602.

[87] *Tangible city scape*, http://tangible.media.mit.edu/project/tangible-cityscape/, Accessed: 2015-09-28.

[88] U. von Zadow, W. Büschel, R. Langner, *et al.*, "Sleed: Using a sleeve display to interact with touch-sensitive display walls," in *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, ACM, 2014, pp. 129–138.

[89] A. Pentland, *Social physics: How good ideas spread-the lessons from a new science*. Penguin, 2014.

[90] N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright, "Effects of four computer-mediated communications channels on trust development," in *Proceedings of the SIGCHI conference on human factors in computing systems*, ACM, 2002, pp. 135–140.

[91] T. Munzner, "Process and pitfalls in writing information visualization research papers," in *Information visualization*, Springer, 2008, pp. 134–153.

[92] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman, and H. Pfister, "Evaluation of artery visualizations for heart disease diagnosis," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2479–2488, 2011, ISSN: 1077-2626. DOI: 10.1109/TVCG.2011.192.

[93] M. Hicks, C. O'Malley, S. Nichols, and B. Anderson, "Comparison of 2d and 3d representations for visualising telecommunication usage," *Behaviour & Information Technology*, vol. 22, no. 3, pp. 185–201, 2003.

[94] A. Cockburn and B. McKenzie, "Evaluating the effectiveness of spatial memory in 2d and 3d physical and virtual environments," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 2002, pp. 203–210.

[95] M. D. Plumlee and C. Ware, "Zooming versus multiple window interfaces: Cognitive costs of visual comparisons," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 13, no. 2, pp. 179–209, 2006.

[96] A. Kjellin, L. W. Pettersson, S. Seipel, and M. Lind, "Evaluating 2d and 3d visualizations of spatiotemporal information," *ACM Transactions on Applied Perception (TAP)*, vol. 7, no. 3, p. 19, 2010.

[97] A. M. MacEachren, *How maps work: representation, visualization, and design*. The Guilford Press, 2004.

[98] A. M. MacEachren and M.-J. Kraak, "Research challenges in geovisualization," *Cartography and Geographic Information Science*, vol. 28, no. 1, 2001.

[99] M.-J. Kraak, "The space-time cube revisited from a geovisualization perspective," in *Proc. 21st International Cartographic Conference*, 2003, pp. 1988–1996.

[100] T. Slocum, *Thematic cartography and geovisualization, 3rd*, 2009.

[101] L. Wilkinson and M. Friendly, "The history of the cluster heat map," *The American Statistician*, vol. 63, no. 2, 2009.

[102] P. F. Fisher, *Developments in spatial data handling.* Springer, 2005.

[103] A. Mehler, Y. Bao, X. Li, Y. Wang, and S. Skiena, "Spatial analysis of news sources," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 12, no. 5, pp. 765–772, 2006.

[104] N. Willems, H. Van De Wetering, and J. J. Van Wijk, "Visualization of vessel movements," in *Computer Graphics Forum*, Wiley Online Library, vol. 28, 2009, pp. 959–966.

[105] M. Yuan, "Temporal gis and spatio-temporal modeling," in *Proceedings of Third International Conference Workshop on Integrating GIS and Environment Modeling, Santa Fe, NM*, 1996.

[106] J. F. Roddick, K. Hornsby, and M. Spiliopoulou, "An updated bibliography of temporal, spatial, and spatio-temporal data mining research," in *Temporal, Spatial, and Spatio-Temporal Data Mining*, Springer, 2001, pp. 147–163.

[107] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: An analytical review," *Journal of Visual Languages & Computing*, vol. 14, no. 6, pp. 503–541, 2003.

[108] T. Abraham and J. F. Roddick, "Survey of spatio-temporal databases," *GeoInformatica*, vol. 3, no. 1, pp. 61–99, 1999.

[109] T.-M. Rhyne, A. MacEachren, and J. Dykes, "Guest editors' introduction: Exploring geovisualization," *IEEE Computer Graphics and Applications*, pp. 20–21, 2006.

[110] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M.-J. Kraak, A. MacEachren, and S. Wrobel, "Geovisual analytics for spatial decision support: Setting the research agenda," *International Journal of Geographical Information Science*, vol. 21, no. 8, pp. 839–857, 2007.

[111] U. Demšar and K. Virrantaus, "Space–time density of trajectories: Exploring spatio-temporal patterns in movement data," *International Journal of Geographical Information Science*, vol. 24, no. 10, pp. 1527–1542, 2010.

[112] R. Scheepens, N. Willems, H. van de Wetering, G. Andrienko, N. Andrienko, and J. J. van Wijk, "Composite density maps for multivariate trajectories," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2518–2527, 2011.

[113] R. Scheepens, N. Willems, H. van de Wetering, and J. J. van Wijk, "Interactive visualization of multivariate trajectory data with density maps," in *Pacific Visualization Symposium (PacificVis), 2011 IEEE*, IEEE, 2011, pp. 147–154.

[114] U. D. Turdukulov, M.-J. Kraak, and C. A. Blok, "Designing a visual environment for exploration of time series of remote sensing data: In search for convective clouds," *Computers & Graphics*, vol. 31, no. 3, pp. 370–379, 2007.

[115] P. Gatalsky, N. Andrienko, and G. Andrienko, "Interactive analysis of event data using space-time cube," in *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, IEEE, 2004, pp. 145–152.

[116] S. Hadlak, C. Tominski, H.-J. Schulz, and H Schumann, "Visualization of attributed hierarchical structures in a spatiotemporal context," *International Journal of Geographical Information Science*, vol. 24, no. 10, pp. 1497–1513, 2010.

[117] P. Shanbhag, P. Rheingans, and M. desJardins, "Temporal visualization of planning polygons for efficient partitioning of geo-spatial data," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, IEEE, 2005, pp. 211–218.

[118] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne, "Flowstrates: An approach for visual exploration of temporal origin-destination data," in *Computer Graphics Forum*, Wiley Online Library, vol. 30, 2011, pp. 971–980.

[119] C. A. Waldspurger and W. E. Weihl, "Lottery scheduling: Flexible proportional-share resource management," in *Proceedings of the 1st USENIX conference on Operating Systems Design and Implementation*, USENIX Association, 1994, p. 1.

[120] J. Kruger and R. Westermann, "Acceleration techniques for gpu-based volume rendering," in *Visualization, 2003. VIS 2003. IEEE*, 2003, pp. 287–292. DOI: 10.1109/VISUAL.2003.1250384.

[121] S. Ghani, N. Elmqvist, and J. S. Yi, "Perception of animated node-link diagrams for dynamic graphs," in *Computer Graphics Forum*, Wiley Online Library, vol. 31, 2012, pp. 1205–1214.

[122] L. H. Hardy, G. Rand, and M. C. Rittler, "Tests for the detection and analysis of color-blindness," *JOSA*, vol. 35, no. 4, pp. 268–271, 1945.

[123] B. Tversky, J. B. Morrison, and M. Betrancourt, "Animation: Can it facilitate?" *International journal of human-computer studies*, vol. 57, no. 4, pp. 247–262, 2002.

[124] P. Baudisch, D. Tan, M. Collomb, D. Robbins, K. Hinckley, M. Agrawala, S. Zhao, and G. Ramos, "Phosphor: Explaining transitions in the user interface using afterglow effects," in *Proceedings of the 19th annual ACM symposium on User interface software and technology*, ACM, 2006, pp. 169–178.

[125] M. Tory, A. E. Kirkpatrick, M. S. Atkins, and T. Moller, "Visualization task performance with 2d, 3d, and combination displays," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 12, no. 1, pp. 2–13, 2006.

[126] C. Ware, *Information visualization: perception for design*. Elsevier, 2012.

[127] C. Turkay, P. Filzmoser, and H. Hauser, "Brushing dimensions – a dual visual analysis model for high-dimensional data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2591 –2599, 2011.

[128] C. Turkay, A. Lundervold, A. Lundervold, and H. Hauser, "Representative factor generation for the interactive visual analysis of high-dimensional data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2621–2630, 2012.

[129] J. Choo and H. Park, "Customizing computational methods for visual analytics with big data," *Computer Graphics and Applications, IEEE*, vol. 33, no. 4, pp. 22–28, 2013.

[130] Z. Liu, B. Jiang, and J. Heer, "Immens: Real-time visual querying of big data," in *Computer Graphics Forum*, Wiley Online Library, vol. 32, 2013, pp. 421–430.

[131] C. D. Stolper, A. Perer, and D. Gotz, "Progressive visual analytics: User-driven visual exploration of in-progress analytics," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 12, pp. 1653–1662, 2014.

[132] H. J. Schulz, M. Angelini, G. Santucci, and H. Schumann, "An enhanced visualization process model for incremental visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 7, pp. 1830–1842, 2016, ISSN: 1077-2626. DOI: 10.1109/TVCG.2015.2462356.

[133] S. Agarwal, B. Mozafari, A. Panda, H. Milner, S. Madden, and I. Stoica, "Blinkdb: Queries with bounded errors and bounded response times on very large data," in *Proceedings of the 8th ACM European Conference on Computer Systems*, ACM, 2013, pp. 29–42.

[134] D. Fisher, I. Popov, S. Drucker, *et al.*, "Trust me, i'm partially right: Incremental visualization lets analysts explore large datasets faster," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, ACM, 2012, pp. 1673–1682.

[135] C. Turkay, F. Jeanquartier, A. Holzinger, and H. Hauser, "On computationally-enhanced visual analysis of heterogeneous data and its application in biomedical informatics," in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, Springer, 2014, pp. 117–140.

[136] M. F. Cohen, S. E. Chen, J. R. Wallace, and D. P. Greenberg, "A progressive refinement approach to fast radiosity image generation," *SIGGRAPH Comput. Graph.*, vol. 22, no. 4, pp. 75–84, 1988.

[137] J. Heer and G. Robertson, "Animated transitions in statistical data graphics," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 13, no. 6, pp. 1240–1247, 2007.

[138] R. Rensink, J. O'Regan, and J. Clark, "To see or not to see: The need for attention to perceive changes in scenes," *Psychological Science*, vol. 8, no. 5, pp. 368–373, 1997.

[139] J. Foley, A. Van Dam, S. Feiner, J. Hughes, and R. Phillips, *Introduction to computer graphics*. Addison-Wesley, 1994, vol. 55.

[140] W. Barfield and C. Hendrix, "The effect of update rate on the sense of presence within virtual environments," *Virtual Reality*, vol. 1, no. 1, pp. 3–15, 1995.

[141] S. Albers, "Online algorithms: A survey," *Mathematical Programming*, vol. 97, no. 1, pp. 3–26, 2003.

[142] A. Blum, "On-line algorithms in machine learning," in *In Proceedings of the Workshop on On-Line Algorithms, Dagstuhl*, Springer, 1996, pp. 306–325.

[143] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008.

[144] D. Sculley, "Web-scale k-means clustering," in *Proceedings of the 19th international conference on World wide web*, ACM, 2010, pp. 1177–1178.

[145] M. Harrower and C. Brewer, "Colorbrewer. org: An online tool for selecting colour schemes for maps," *Cartographic Journal, The*, vol. 40, no. 1, pp. 27–37, 2003.

[146] C. Turkay, A. Slingsby, H. Hauser, J. Wood, and J. Dykes, "Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 20, no. 12, pp. 2033–2042, 2014.

[147] S. Chan, L. Xiao, J. Gerth, and P. Hanrahan, "Maintaining interactivity while exploring massive time series," in *Visual Analytics Science and Technology. IEEE Symposium on*, IEEE, 2008, pp. 59–66.

[148] S. Radoš, R. Splechtna, K. Matkovic, M. Duras, E. Gröller, and H. Hauser, "Towards Quantitative Visual Analytics with Structured Brushing and Linked Statistics," *Computer Graphics Forum*, 2016, ISSN: 1467-8659. DOI: `10.1111/cgf.12901`.

[149] E. Catmull, "The problems of computer-assisted animation," *SIGGRAPH Comp. Graph.*, vol. 12, no. 3, pp. 348–353, 1978, ISSN: 0097-8930.

[150] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko, "Effectiveness of animation in trend visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 6, pp. 1325 –1332, 2008, ISSN: 1077-2626.

[151] N. Elmqvist, P. Dragicevic, and J. Fekete, "Rolling the dice: Multidimensional visual exploration using scatterplot matrix navigation," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 6, pp. 1539–1148, 2008.

[152] C. Bentley and M. Ward, "Animating multidimensional scaling to visualize n-dimensional data sets," in *Information Visualization, Proceedings IEEE Symposium on*, 1996, pp. 72–73, 126.

[153] D. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang, "Ipca: An interactive system for pca-based visual analytics," *Computer Graphics Forum*, vol. 28, no. 3, pp. 767–774, 2009.

[154] D. Archambault, H. Purchase, and B. Pinaud, "Animation, small multiples, and the effect of mental map preservation in dynamic graphs," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 4, pp. 539–552, 2011.

[155] Y. Frishman and A. Tal, "Online dynamic graph drawing," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 4, pp. 727–740, 2008.

[156] R. Johnson and D. Wichern, *Applied multivariate statistical analysis*. Prentice Hall, 2007, vol. 6.

[157] P. Tan and K. V. Steinbach M., *Introduction to data mining*. Addison-Wesley, 2005.

[158] M. Wijffelaars, R. Vliegen, J. Van Wijk, and E. Van Der Linden, "Generating color palettes using intuitive parameters," *Computer Graphics Forum*, vol. 27, no. 3, pp. 743–750, 2008.

[159] E. Bertini, A. Tatu, and D. Keim, "Quality metrics in high-dimensional data visualization: An overview and systematization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 12, pp. 2203–2212, 2011.

[160] L. Van der Maaten and G. Hinton, "Visualizing non-metric similarities in multiple maps," *Machine learning*, vol. 87, no. 1, pp. 33–55, 2012.

[161] H. Guo, S. R. Gomez, C. Ziemkiewicz, and D. H. Laidlaw, "A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 22, no. 1, pp. 51–60, 2016.

[162] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 12, pp. 2431–2440, 2012.

[163] M. M. Gaber, A. Zaslavsky, and S. Krishnaswamy, "Mining data streams: A review," *ACM Sigmod Record*, vol. 34, no. 2, pp. 18–26, 2005.

[164] Y. Tillé, *Sampling algorithms*. Springer, 2011.

[165] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, *et al.*, "Computational social science," *Science*, vol. 323, no. 5915, p. 721, 2009.

[166] A. Pentland and T. Heibeck, *Honest signals: how they shape our world*. MIT press, 2010.

[167] F. F. Reinchheld, "The loyalty effect: The hidden force behind growth, profits, and lasting value," *Long Range Planning*, vol. 6, no. 29, p. 909, 1996.

[168] *The standout customer loyalty stats of 2017*. Available from: `http://www.socialannex.com/blog/2017/01/10/standout-customer-loyalty-stats-2017/`, Accessed: 2017-05-30.

[169] *Customer acquisition vs. retention costs - statistics and trends.* Available from: `https://www.invespcro.com/blog/customer-acquisition-retention/`, Accessed: 2017-05-30.

[170] *Accenture global consumer pulse survey.* Available from: `https://www.accenture.com/us-en/insight-digital-disconnect-customer-engagement`, Accessed: 2017-05-30.

[171] S. V Nath and R. S. Behara, "Customer churn analysis in the wireless industry: A data mining approach," *Proceedings-Annual Meeting of the Decision Sciences Institute*, 2003.

[172] N. V. Patil and A. Dixit, "Survey on profit maximizing metric for measuring classification performance of customer churn prediction models," *International Journal*, vol. 4, no. 12, 2014.

[173] B. Huang, M. T. Kechadi, and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414–1425, 2012.

[174] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1659–1665, 2014.

[175] Y. Huang, F. Zhu, M. Yuan, K. Deng, Y. Li, B. Ni, W. Dai, Q. Yang, and J. Zeng, "Telco churn prediction with big data," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, ACM, 2015, pp. 607–618.

[176] Y. Liu and Y. Zhuang, "Research model of churn prediction based on customer segmentation and misclassification cost in the context of big data," *Journal of Computer and Communications*, vol. 3, pp. 87–93, 2015.

[177] W. Bi, M. Cai, M. Liu, and G. Li, "A big data clustering algorithm for mitigating the risk of customer churn," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1270–1281, 2016.

[178] A. Wangperawong, C. Brun, O. Laudy, and R. Pavasuthipaisit, "Churn analysis using deep convolutional neural networks and autoencoders," *arXiv preprint arXiv:1604.05377*, 2016.

[179] Y. Xie, X. Li, E. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, 2009.

[180] B. Larivière and D. Van den Poel, "Predicting customer retention and profitability by using random forests and regression forests techniques," *Expert Systems with Applications*, vol. 29, no. 2, pp. 472–484, 2005.

[181] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert systems with applications*, vol. 34, no. 1, pp. 313–327, 2008.

[182] E. G. Castro and M. S. Tsuzuki, "Churn prediction in online games using players' login records: A frequency analysis approach," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 3, pp. 255–265, 2015.

[183] W. Buckinx and D. Van den Poel, "Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual fmcg retail setting," *European Journal of Operational Research*, vol. 164, no. 1, pp. 252–268, 2005.

[184] T. Verbraken, W. Verbeke, and B. Baesens, "A novel profit maximizing metric for measuring classification performance of customer churn prediction models," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 5, pp. 961–973, 2013.

[185] W.-H. Au, K. C. Chan, and X. Yao, "A novel evolutionary data mining algorithm with applications to churn prediction," *IEEE transactions on evolutionary computation*, vol. 7, no. 6, pp. 532–545, 2003.

[186] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, "Applying data mining to telecom churn management," *Expert Systems with Applications*, vol. 31, no. 3, pp. 515–524, 2006.

[187] C. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[188] N. Eagle, M. Macy, and R. Claxton, "Network diversity and economic development," *Science*, vol. 328, no. 5981, pp. 1029–1031, 2010.

[189] W. Pan, N. Aharony, and A. Pentland, "Fortune monitor or fortune teller: Understanding the connection between interaction patterns and financial status," in *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, IEEE, 2011, pp. 200–207.

[190] H. Selye, "The general adaptation syndrome and the diseases of adaptation," *The Journal of clinical endocrinology & metabolism*, vol. 6, no. 2, pp. 117–230, 1946.

[191] U. D. Prasad and S Madhavi, "Prediction of churn behavior of bank customers using data mining tools," *Business Intelligence Journal*, vol. 5, no. 1, pp. 96–101, 2012.

[192] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[193] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, “Svms modeling for highly imbalanced classification,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 281–288, 2009.

[194] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[195] G. Lemaître, F. Nogueira, and C. K. Aridas, “Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning,” *CoRR*, vol. abs/1609.06570, 2016. [Online]. Available: `http://arxiv.org/abs/1609.06570`.

[196] J. Burez and D. Van den Poel, “Handling class imbalance in customer churn prediction,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 4626–4636, 2009.

[197] N. V. Chawla, “Data mining for imbalanced datasets: An overview,” in *Data mining and knowledge discovery handbook*, Springer, 2005, pp. 853–867.

[198] *Ic2s2: International conference on computational social science.* Available from: `https://ic2s2.org/`, Accessed: 2017-05-30.

[199] *Netmob: The conference on the scientific analysis of mobile phone datasets.* Available from: `http://netmob.org/`, Accessed: 2017-05-30.

[200] K. Krippendorff, “Reliability in content analysis,” *Human communication research*, vol. 30, no. 3, pp. 411–433, 2004.

[201] S. R. Langton, R. J. Watt, and V. Bruce, “Do the eyes have it? cues to the direction of social attention,” *Trends in cognitive sciences*, vol. 4, no. 2, pp. 50–59, 2000.

[202] J. P. Otteson and C. R. Otteson, “Effect of teacher’s gaze on children’s story recall,” *Perceptual and Motor Skills*, vol. 50, no. 1, pp. 35–42, 1980.

[203] R. V. Exline and L. C. Winters, “Affective relations and mutual glances in dyads,” *Affect, cognition, and personality. New York: Springer*, vol. 1, no. 5, 1965.

[204] M. F. Mason, E. P. Tatkow, and C. N. Macrae, “The look of love: Gaze shifts and person perception,” *Psychological Science*, vol. 16, no. 3, pp. 236–239, 2005.

[205] S. Kiesler, J. Siegel, and T. W. McGuire, "Social psychological aspects of computer-mediated communication.," *American psychologist*, vol. 39, no. 10, p. 1123, 1984.

[206] M. Lea and R. Spears, "Computer-mediated communication, de-individuation and group decision-making," *International journal of man-machine studies*, vol. 34, no. 2, pp. 283–301, 1991.

[207] L. Wu, B. N. Waber, S. Aral, E. Brynjolfsson, and A. Pentland, "Mining face-to-face interaction networks using sociometric badges: Predicting productivity in an it configuration task," 2008.

[208] B. N. Waber, D. Olguin Olguin, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Organizational engineering using sociometric badges," 2007.

[209] S. Smith, R. D. Bergeron, and G. G. Grinstein, "Stereophonic and surface sound generation for exploratory data analysis," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1990, pp. 125–132.

[210] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbæk, "Opportunities and challenges for data physicalization," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 2015, pp. 3227–3236.

[211] D. Bilgili and S. Balcisoy, "Physvis: A data physicalization pipeline enhanced with augmented reality.," *Poster presented at IEEE VIS 2017 Conference*,

[212] *Gnu project (2015). gnu pspp for gnu/linux (version 1.0.1) [computer software]*, `https://www.gnu.org/software/pspp/`.