Olga Donkovtceva

KNOWLEDGE DISCOVERY AND VISUAL ANALYTICS IN A DATA-DRIVEN BUSINESS ENVIRONMENT

Master's Thesis in Information Systems
Master's Program in Governance of Digitalization
Supervisor: Associate Prof. Jozsef Mezei
Faculty of Social Sciences, Business and Economics
Åbo Akademi University

Åbo 2021

ABSTRACT

Subject: Information Systems	
Writer: Olga Donkovtceva	
Title: Knowledge discovery and visual analytics in a data-driven business environment	
Supervisor: Associate Prof. Jozsef Mezei	

Abstract:

Nowadays, becoming data centric and insights driven is a critical strategy for nearly all businesses. Data can be regarded today as oxygen, which is needed in order to thrive in the data-driven business environment, as they conceal ample opportunities for business organizations for discovering new insights and knowledge. Visual analytics has become an indispensable part of business, as it enables analysts to explore data via visualizations and discover knowledge in innovative ways. Visual representations of information reduce complex cognitive efforts required to perform certain tasks. However, converting data into meaningful visualizations is not a trivial task and it cannot necessarily be automatically improved by continuously increasing computational capacity.

The purpose of this study is to examine why visualization is critical in today's business organizations for innovative knowledge discovery and what drives knowledge discovery in the context of visual culture in organizations. The study attempts to explain what facilitates the process of knowledge discovery from the holistic perspective including the role of an analyst (human discourse), the tool-based viewpoint in terms of features and the organizational perspective in terms of the visual culture concept. It also encompasses the notion of visual discovery by design and endeavors to contemplate its implications in the knowledge discovery process.

In order to fulfil the purpose of the study, an appropriate theoretical background was built initially. Further, mixed research methods were applied to the study. The qualitative data were collected in the form of semi-structured interviews. Afterwards, a survey was conducted to collect quantitative data. The target group for both qualitative and quantitative data collections comprised business analysts. Finally, the thesis provides evaluations of the findings while comparing the literature review results in a discussion form and giving concluding points.

Keywords: Visual Analytics, Visualization, Knowledge discovery, Interactive visualization

Date: dd.mm.yyyy	**Number of pages:** xxx

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

1 INTRODUCTION

In today's data-driven economy, in almost every domain there is an enormous demand for visual analytics to discover knowledge in innovative ways, since data and analytics have become imperative drivers for business strategy. Yet, it is complicated and challenging to fully leverage data. Complex, heterogenous and multidimensional data continuously accrue concealing ample opportunities for business organizations to discover new insights and make data-driven, timely and swift decisions by means of visual analytics approaches. Numerous advanced and top-notch analytics and Business Intelligence (BI) tools are at business's disposal nowadays, which offer business users solutions for augmented analytics enabling them to focus on data exploration and knowledge discovery rather than on complicated analytics processes and computational algorithms (Dewani 2020; Dar 2020; Gartner PeerInsights 2021; Sosulski 2019). That is, with the help of such tools, analysts can fully embrace their perceptual and cognitive capabilities in the analytical work while exploiting advanced computational capabilities enhancing the discovery process.

In the age of information overload or so called "data deluge", visual analytics has become an indispensable part of business. The aim of visual analytics is to convert data overload into meaningful information and eventually into knowledge by utilizing three primary areas: data mining, information retrieval and human interaction. In visual analytics, human perception and cognition coupled with interaction and augmented capabilities of a visualization tool facilitate data exploration, data comprehension and eventually knowledge discovery. Visual representations of information reduce complex cognitive efforts required to perform certain tasks. In addition, visualizations encourage creative thinking, which is usually not associated with data analysis, but can lead to unexpected insights. However, converting data into meaningful visualizations is not a trivial task and this process can not necessarily be automatically improved by continuously increasing computational capacity. In order to achieve effective visual data analysis, advanced concepts of representation, interaction, perception and decision-making must be considered and applied.

Numerous studies (Stahl et al. 2013; Endert et al. 2017; Bertini & Lalanne 2010) have investigated how visualization and data mining can be best integrated to achieve the synergism of the best of human and computer capabilities for effective knowledge discovery. Some pieces of research (Bertini & Lalanne 2010; Elmqvist et al. 2011; Keim et al. 2010; Chang et al. 2014) have been conducted from the machine-to-human perspective while analysing various ways of augmentation and enhancement of visualization and data mining techniques and methods to facilitate knowledge discovery. Some studies (Green et al. 2009; Pohl et al. 2012) have analysed human capabilities to perceive, relate and conclude and how these capabilities can be exploited and leveraged in visual analytics. Ultimately, certain human cognition frameworks for information visualization have been suggested, and principles and guidelines for visualization design have been developed.

This research attempted to investigate why visualization is critical in today's business organizations for innovative knowledge discovery and what drives knowledge discovery in the context of visual culture in organizations by distilling the most important findings in different standpoints of visual analytics.

1.1 Research objectives and questions

Data-driven economy forces organizations of different sizes to embrace and exploit data strategically. Various BI and visual analytics tools boost companies' abilities to handle complex and massive data in order to gain insights and extract knowledge. This research endeavored to explain drivers of the knowledge discovery process from the perspectives of the organizational culture, tool-based features and human information discourse.

The research objectives and questions placed for the thesis are the following:

Research question (RQ) 1: What drives insight and knowledge discovery in visual analytics?

- Understand the role of visual data discovery in the context of data-driven environment.

- Identify the drivers of knowledge discovery by means of visualization tools.

Research question (RQ) 2: Why and how does visualization simplify cognitive tasks?

- Understand why visualizations support human's high-level cognitive functioning.

- Identify how visual representations facilitate human cognitive capabilities.

Research question (RQ) 3: What are the implications of the visual discovery by design for knowledge generation?

- Understand what discovery by design is and what it entails for knowledge generation.

1.2 Scope of the thesis

The scope of the thesis covers consideration of data and information visualization within visual analytics in terms of knowledge discovery in the context of data-driven business world. The study attempts to describe and explain what facilitates the process of knowledge discovery from the holistic perspective including the role of an analyst (human discourse), the tool-based viewpoint in terms of features and the organizational perspective in terms of the visual culture concept, and why visualization plays a crucial role in visual analytics. It also encompasses the notion visual discovery by design and endeavors to contemplate their implications in the knowledge discovery process.

The purpose of the research is to understand the very drivers of knowledge discovery in visual analytics rather in a broader sense by considering the role of a user in the interaction with a VA tool and the role of visual culture in the process of knowledge discovery.

1.3 Thesis outline

The structure of the thesis is illustrated in Figure 1.

Introduction depicts the background and motivation of the research; it also illustrates research questions and objectives. Chapter 2 provides a theoretical part by overviewing theories, concepts and frameworks most important and relevant to the research scope. It scrutinizes first literature regarding data-driven organizations and the concept of visual data discovery and its role for knowledge discovery in the context of data-driven business environment. Then it provides a literature review on such aspects as visualization and knowledge discovery, visualization and human cognition, visualization and interaction. Finally, the theoretical part includes an overview of the concept of visual discovery by design. Chapter 3 provides description of methodology and data collection for the research. Further, the findings are reported in Chapter 4. Discussion and conclusion are included in Chapter 5 and 6.

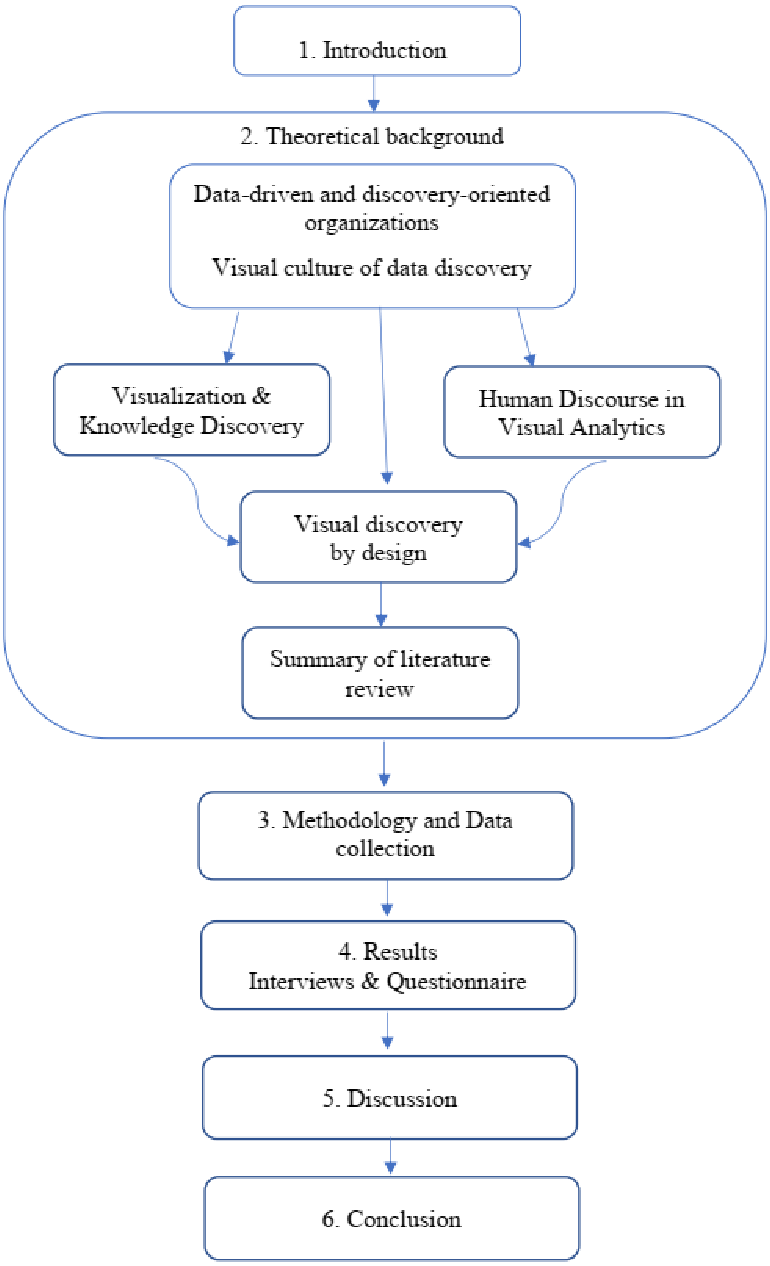


Figure 1 The structure of the thesis

2 THEORETICAL BACKGROUND

This chapter introduces the main theories, frameworks and concepts essential for the research. It discusses the importance of discovery-oriented and visual culture in organizations and its implications in the context of data-driven environment and knowledge discovery. This section also elaborates on data and information visualization and its role in knowledge discovery as well as human information discourse and interactive visualization. Finally, a concept of discovery by design is introduced and explained.

2.1 Data-driven and discovery-oriented organizations

The Table 1 lists and explains definitions of the key concepts in the framework of the data-driven business environment.

Table 1 Key terms and definitions

Key terms	Explanation
Data	refer to symbols representing properties of objects, events and their environments and being of no value until they are processed into a useable form (Ackoff 1989)
Information	is inferred from the data into a meaningful form answering the questions at hand (what, who, where, when, how many) (Ackoff 1989); refers to the data resulting from the computational process for the assignment of meanings to the data or the transcripts of the meanings assigned by the analyst (Chen 2009)
Knowledge	refers to the application of relevant information to answer the "why" questions (Ackoff 1989; Chen 2009); refers to the capability to determine required information relating to the problem at hand and interpret it (Ylijoki 2019)
Business Intelligence (BI)	combines business analytics, data mining, data visualization, that is, applications, tools and infrastructure along with best practices which enable access to and analysis of data and

	information to improve and optimize decisions and performance. (Tableau 2020; TIBCO 2020)
Analytics	"is used to describe statistical and mathematical data analysis that clusters, segments, scores and predicts what scenarios are most likely to happen." (Gartner IT Glossary 2020)
Advanced analytics	"is the autonomous or semi-autonomous examination of data or content using sophisticated techniques and tools, typically beyond those of traditional business intelligence (BI), to discover deeper insights, make predictions, or generate recommendations. Advanced analytic techniques include those such as data/text mining, machine learning, pattern matching, forecasting, visualization, semantic analysis, sentiment analysis, network and cluster analysis, multivariate statistics, graph analysis, simulation, complex event processing, neural networks." (Gartner IT Glossary 2020)
Data-driven decision-making	refers to the extent to which an organization embraces new ideas which challenge current practices based on data-driven insight while collecting and having data available for making decisions and depending on data-based insights. (Cao et al. 2015)
Big data	are characterized by volume, velocity, variety, veracity, validity, value, variability, venue, vocabulary and vagueness (Tsai et al. 2015) are "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation". (Gartner IT Glossary 2020)

The era of big data and innovative technologies has created both numerous challenges and tremendous opportunities for businesses to explore and apply appropriate technologies for an effective data governance and creation of agile and more streamlined data analytics. Today's business setting demands from companies to build a cohesive environment, which not only enables integration of data management and data analyses, but also provides agile analytics on an ad hoc basis and delivers real-time reports. However, not every organization is adept in making sense of heterogeneous data it amasses in high volume and at high velocity (Davenport 2018). Today, more than ever before, there is a critical need in the ability to make agile, data-driven decisions based on real-time analytics. Data continuously accrue and drive rapid disruption, reinvention, and digital transformation, which, in turn, forces organizations to transform to visual, data-driven and discovery-oriented organizations. (Ryan 2016)

In literature, different authors use different terms for defining organizations, which embrace and leverage data, data visualization (DataViz) tools and analytics in an effective way and ultimately gain insights and discover knowledge. Simon (2014) defines such organizations as Visual Organizations. Visual organizations are those with a data-driven mindset and which routinely take advantage of visualizing data from different sources and of different types by utilizing contemporary, interactive and flexible DataViz tools and applications or creating new ones, if necessary. The core principles of visual organizations are first, the ease of data access, data discovery and data processing for everyone; second, the ability to visualize data regardless of its size, type and source; third, swiftness and agility in finding data. (Simon 2014)

Data-driven or data-centric are another, more common, terms used to describe organizations, which embrace data and analytics. Insights-driven organizations appear in literature when investigating data-driven business (Saulles 2019). As the significance and implications in exploiting data and analytics for organizations change, the definitions of data-driven organizations are continuously broadening and gaining new meanings. Kirin (2017) distinguishes three levels of data-driven organizations. On the first level, there are organizations with well-governed data and commitment from both parties - leadership and employees. When organizations start handling data as a core asset and as an imperative element in building strategies and transforming relationships with stakeholders, then they move to the second level. Finally, with mature capabilities in data and analytics, organizations become adept in establishing new, innovative business models and operations, which might bring about fundamental alterations within an organization. (Kirin 2017) The significance of creating a data-driven culture in an organization, specifically from top-down, is obvious (Saulles 2019). In order to thrive in the data-driven world, where volume, velocity, variety and veracity of data are rising, organizations have to build a well-established and fine-tuned ecosystem to be able to effectively generate, collect, assimilate, analyze and discover insights and knowledge.

There is also the term big data phenomenon (Ylijoki 2019), which refers to "the paradigm shift towards data-driven businesses and ecosystems" and implies big and ubiquitous cultural or organizational change towards a new data-oriented mindset. The author emphasizes the need of organizations to move towards the data-driven culture, in which analytics and insights should be embedded into processes and decision-making. (Ylijoki 2019)

Ryan (2016) extends the definition of data-driven culture and refers to discovery-oriented culture. According to Ryan (2016, p.8), discovery-oriented culture implies agile and iterative process from information to insight in a more proactive, actionable and real-time way. Data-centric companies embrace their analytical culture to generate knowledge in four principal areas: customer insights, product insights, optimization and innovation (Ryan 2016). Ryan (2016) argues that data-centric is on par with competitive advantage. In order to gain this competitive advantage as a data-centric organization, companies must not only accumulate data from different sources, but also embrace ongoing analytics, think long term and simultaneously act on analytics real time. Analytics must be swift and prompt enough to make a difference. In the establishment of a successful data-driven culture, one of the most significant aspects is the cultivation of principles of data discovery and data visualization. Obviously, this requires a cultural change within an organization. (Ryan 2016)

Increasing availability of heterogenous data stimulates the data discovery process which, in turn, demands the ability to explore, dig deeper into data and reveal hidden information to gain insights and create knowledge. As this is an iterative process, a multitude of iterations might be required before a meaningful insight or knowledge could be discovered. Therefore, the IT department's role is transforming to enablement technology, that is, it not only enables the analytical process but fosters it, while allowing analysts to focus on the primary tasks without cessations (Ryan 2016). In this sense, Nguyen (2016) defines the analytical data life cycle as consisting of several stages which require a specific role in an organization. This cycle comprises data exploration, data preparation, model development and model deployment. Nguyen (2016) also refers to the term of enabling technology implying in-database processing and in-memory analytics, which enable organizations to amass, process, and analyse data in an effective, efficient and agile way without any possible interruptions.

2.2 Visual culture of data discovery

Today, data must be converted into information quickly to be analysed further. For this purpose, organizations exploit fast analytics through visualization. Fast analytics is used synonymously with visual analytics and refers to the ability to quickly obtain and visualize data (Larson & Chang 2016). Visual or fast analytics is more about discovery, which can result in knowledge and which, in turn, creates a refinement of the visual product. (Larson & Chang 2016)

Visual representation of data is an influential mechanism in the visual data discovery process. Fast analytics and visualization are conducted as parts of exploratory data analysis (Larson & Chang 2016). The competency of data visualization must be cultivated nowadays in order to fully embrace the ethos of visual culture of data discovery. According to Ryan (2016), creating a visual culture of data discovery within an organization implies balancing design, information and organizational culture with the emphasis on gaining insights, discoveries and ability to visually communicate. Nowadays, the significance of data visualization is growing due to the fact that organizations are demanded to make faster data-driven decisions and swiftly deliver data-driven solutions, wherein abilities in visualizing huge volumes of multidimensional data stimulates insight and knowledge discovery. There are numerous impactful, flexible and user-friendly data visualization and visual analytics tools at companies' disposal today, which enable companies to visualize and analyze their data in an innovative and effective way for knowledge discovery. Nowadays more than ever raw data can be interpreted on the fly for ad hoc analyses. Coupled with enablement technology, these DataViz tools allow organizations to gain insights and discover knowledge with the further swift decision-making. (Simon 2014)

2.3 Visualization and knowledge discovery

This section investigates the core questions of this research by attempting to find answers to why visualize and how it helps in knowledge discovery from the tool-based perspective and from the perspective of a business analyst.

2.3.1 Visualization in knowledge discovery process

Knowledge discovery has become strategically significant for businesses. Yet, effective generation of knowledge from massive datasets remains challenging for many organizations. (Begoli & Horey 2012)

The term datafication is used to depict the pervasive nature of data and increasing significance of data. Lycett (2013) defines datafication as "an information technology driven sense-making process". According to Ackoff (1989), data does not have value as such, until it is processed into usable form. Ackoff presented a model explaining how data generates value, many other studies have been based on this model later. The model describes a data-information-knowledge-wisdom hierarchy and helps in the understanding of the value creation process by conceptualizing this process in an abstract model (Ylijoki 2019). Three types of data exist in every entity (Murray & Chabot 2016, p. 5): first, "known data" contained in daily, weekly and monthly reports providing the visibility of operations and the basic context for discussions; second, "data you know you need to know" providing answers to questions which can arise from the known data; third, "data you don't know you need to know" which can be captured through interaction leading to uncovering patterns and outliers, which are invisible in previous types of data.

Visualization is one part of visual analytics, incorporating also "algorithmic data analysis and analytical reasoning, which takes advantage of visualization and interactions as suitable tools to integrate human judgement into KDD process to visually discover explainable patterns (knowledge) and to gain insight into large and complex data sets" (Cui 2019, p. 81559).

Visualization is not a new concept and is nowadays a crucial aspect in fast analytics. Visualization enables users to swiftly comprehend complex data sets produced from statistical analysis or analytical models. As today data are determined not only by volume, velocity and variety, but also by value, validity, veracity and visibility, it bears immense significance of visualization. (Larson & Chang 2016)

According to Colin Ware (Ware, 2004), visualization is considered to be a visual representation of data with numerous benefits: it enables to understand vast amounts of data, its structure and quality, it allows to identify trends and patterns in data, it supports in gaining insights and decision-making. The purpose of visualization, wherein interactivity is a fundamental aspect, consists in supporting users to intuitively explore data within a visual context by leveraging the human visual system to provide insight and understanding about data, to assist in reasoning and problem solving. (Kim 2016, Patterson et al. 2014)

Ryan (2016, p. 179) defines data visualization as “a visual display of information that is transformed by the influence of purposeful design decisions with the intent of encoding and conveying information that would otherwise be either difficult to understand or unlikely (or impossible) to connect with in a meaningful way”. The author describes data visualization as a process of drawing values to visuals, as a tool assisting in exploration of trends, patterns, correlations and in revealing results and insights, and as a mechanism for data communication, exploration and explanation. Azzam (2013) also explains data visualization as a threefold concept: it is *a process* based on quantitative or qualitative data, it is *results* in visual representations of raw data, which is interpretable by *users* and supports user’s exploration, investigation and communication of data.

Knowledge discovery is a data-intensive process enabling users to identify trends, models, patterns hidden in vast amounts of heterogeneous data. Numerous methods assisting in detecting meaningful information and discovering useful knowledge have been introduced, for instance, decision trees, evidence theory, cluster analysis., etc.

KDD or knowledge discovery in databases (or knowledge discovery and data mining) is the process of making sense out of data by exploiting data mining techniques with the purpose of discovering or extracting trends and patterns, which can be transformed into knowledge (Sacha et al. 2014). KDD enables to extract implicit, previously unknown, and potentially meaningful information. This process includes 4 stages (Figure 2): first, preparing data relevant for analysis which includes selecting, cleaning and converting data; second, utilizing suitable data mining technique; third, assessing discovered patterns; and forth, integrating discovered knowledge into decision-making system. (Ltifi 2016) Begoli & Horey (2012) argue that effective KDD requires from enterprises to implement effective organizational and technological practices, which comprise the following elements:

> • Data collection, storage and organization practices
> • Understanding and effective application of the modern data analytic methods (including tools)
> • Understanding of the problem domain and the nature, structure and meaning of the underlying data (Begoli & Horey 2012)

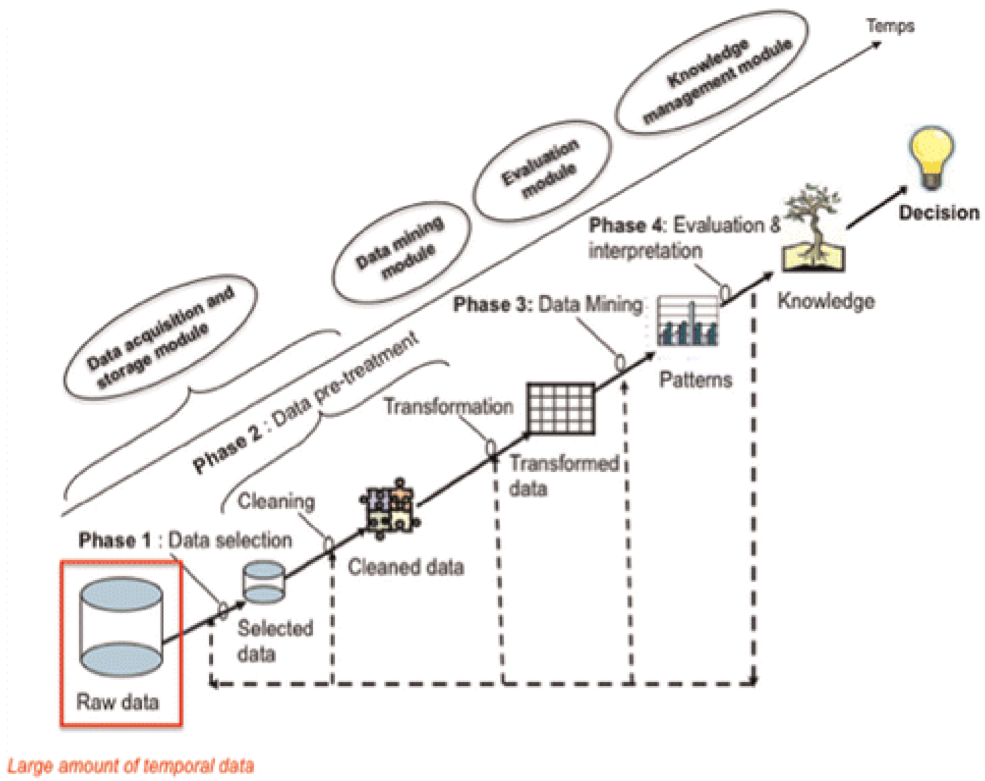


Figure 2 KDD process (Ltifi 2016, p.33)

Visualization often accompanies the KDD process and has become an essential mechanism in knowledge discovery regardless of the data type. Moreover, Leece (1999) emphasises that "visualization of multidimensional data is one of the key tools for data mining as part of the overall process of knowledge discovery in databases (KDD)".

If viewing data visualization and data mining separately in terms of knowledge discovery, the former presents a human-centered approach to knowledge discovery, whereas the latter is commonly machine-driven approach exploiting tools for automatic extraction of patterns and models from data, for devising information and eventually knowledge

(Bertini & Lalanne 2010). Keim et al. (2008) determine gaining insight or knowledge discovery as the final step in the process of visual analytics. Fekete et al. (2008) explain the value of visualization so that "demonstrations should start using a simple question, show that a good representation answers the question at once and then argue about additional benefits, i.e., questions the users did not know they had". For users it implies that a good visualization will support their prior knowledge, provide a prompt answer to the question asked, give them insights resulting in the so called "a-ha" moment, wherein users feel they comprehend the dataset.

Users play a central role in constructing knowledge from the snippets of information retrieved from visual data analysis. Recent studies in visual analytics theories emphasise the need in the shift from "human in the loop" to "human is the loop" concept in order to merge human work processes with analytics (Sacha et al. 2014). Sacha et al. (2014) developed a knowledge generation model for visual analytics (Figure 3) illustrating integration of a visual analytics system on the one side and knowledge generation process of a human on the other side. From the human cognitive processes side, it is a reasoning process which consists of exploration, verification and knowledge generation loops. (Sacha et al. 2014) Visualization exploits data or models created from data and allows the user to identify patterns and relationships. Visualization is generally the primary interface between the user and visual analytics system, whereas the comprehension of models requires cognitive efforts. (Sacha et al. 2014)

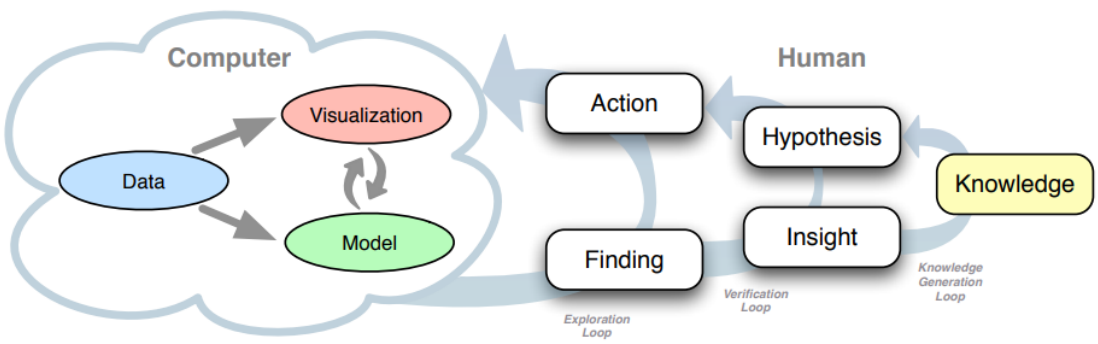


Figure 3 Knowledge generation model for visual analytics (Sacha et al. 2014, p.14)

In the knowledge generation model the exploration loop presents user interactions with the system. Findings or a user's goal specifies actions taken in the exploration loop. A user's specific goal might be missing, or findings are not necessarily related to this goal and eventually they can bring about new insights or new analytical directions (Sacha et al. 2014). Sacha et al. then break down the process model into action and cognition paths: data preparation including data collection or data selection, model building related to KDD process and its configuration, model usage such as statistical calculation or data clustering, then visual mapping for creating data visualizations, model-vis mapping outlining models into visualizations. Then manipulating with visualization alters the output or stresses striking data without the need to change the model-vis mapping. (Sacha et al. 2014)

The exploration loop is steered to the verification loop to confirm hypotheses or formulate new ones. Visualization itself facilitates hypothesis formation (Ware 2004). By being able to translate findings from the exploration loop in the context of the problem realm, the user can obtain new insights, which might be not related to the initial hypothesis or which in turn can result in new hypotheses or further investigations. Moreover, the user obtains additional knowledge when collecting more reliable insights. (Sacha et al. 2014)

In the knowledge generation loop the user's prior knowledge about the data, the problem domain, visual analytics tools and methodology determine the analysis strategy and procedure. During the visual analytics process the user attempts to identify evidence for hypotheses or discover new knowledge about the problem realm, which can be defined as "justified belief" (Sacha et al. 2014). However, the characteristics of the evidence can affect the trustworthiness of the discovered knowledge, and the findings of the statistical testing may be perceived more reliable than patterns in visualization. Hence, it is of the user's decision on whether the collected evidence is sufficient to trust a new insight or discovered knowledge or whether further investigation is required. Sacha et al. (2014) state that the evaluation of new knowledge trustworthiness should start from data collection and is critical in the course of the entire analysis process. (Sacha et al. 2014)

Figure 4 illustrates how the knowledge generation model (Sacha et al. 2014) is related to other models and theories. That is, along with the KDD process represented by data, model and their mapping, there is also InfoVis Pipeline corresponding to data, visualization and their mappings. When data is converted to visualization, the user can generate knowledge based on perception through interactive exploration. Green et al. (2009) argue that perception plays a crucial role in interactive exploration, which is associated with reasoning and which oftentimes results in knowledge acquisition.

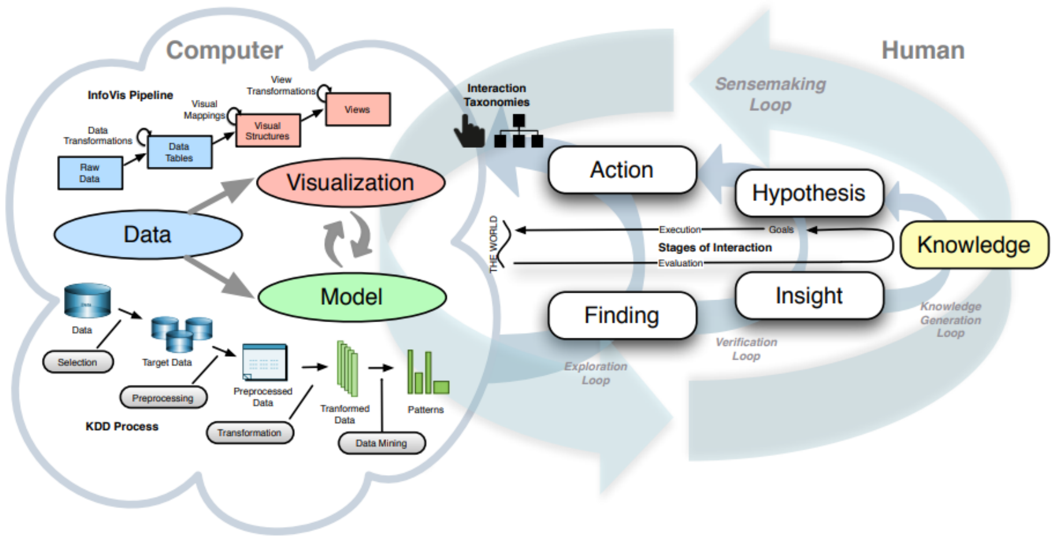


Figure 4 Relation of the knowledge generation model with other models and theories (Sacha et al. 2014, p.20)

Figure 5 illustrates a typical visualization pipeline in terms of two contexts of visualizing – *viewing* and *seeing*, which correspond to different parts in the visualization pipeline. The process of defining important or noteworthy information, generating suitable visual representations and communicating visual representations to the target audience is referred to as *viewing*. In the literature, it is clarified in terms of "making visible to one's eyes". *Seeing* relates to a viewer's cognitive processes and experiences of interpreting visuals by translating them into mental representations in order to comprehend what message they convey. In the literature, it is explained in terms of "making visible to one's mind". Viewing implies concentration on a visualization process, which is supported by a tool or a system and includes filtering, mapping, rendering. Seeing implies optimizing the effectiveness and usefulness of a visualization process and includes creating visual metaphors, designing visuals, assessing visualization outcome and user experience. (Chen & Florodi 2013) In addition, the authors broadly group visualizations into three main categories depending on the user's tasks: information retrieval, information analysis and information dissemination. In terms of knowledge discovery, information analysis serves as the most essential objective of a visualization for gaining insights from data and comprises various analytical tasks, such as filtering, clustering, sorting, identifying patterns and anomalies, combining and portioning data, comparing and finding correlations, evaluating hypotheses. (Chen & Florodi 2013)

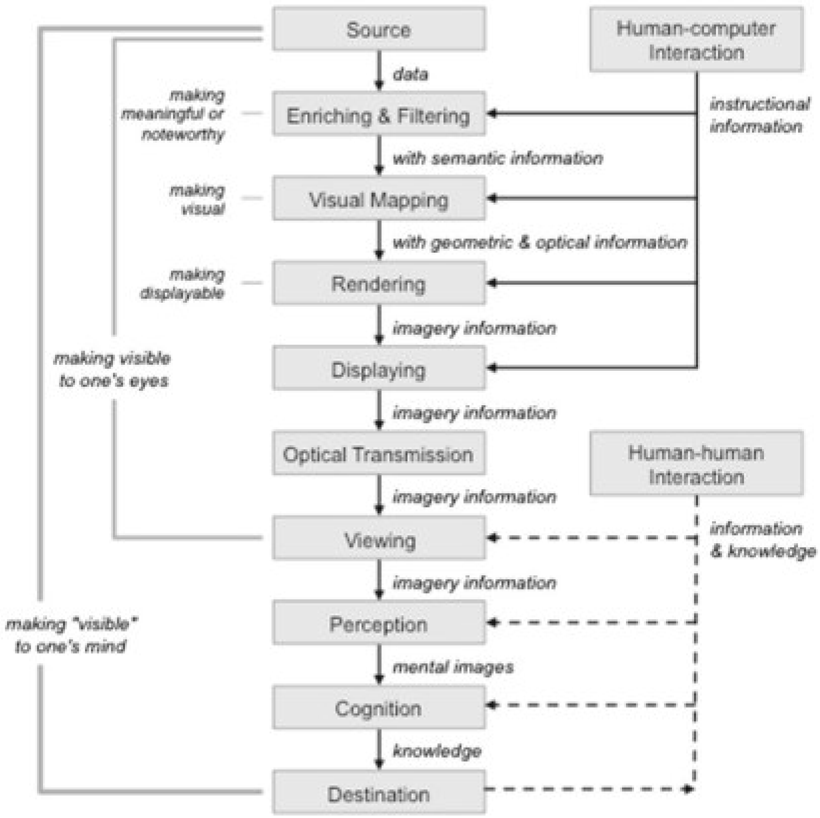


Figure 5 A typical visualization pipeline (Chen & Florodi, 2013, p.3422)

The authors (Chen & Florodi 2013) also define requirements in order for data to become information and for information to become new knowledge. In a broad sense, in order to become information, data need to be, first, well-formed, implying correct data collection by strictly following certain rules (syntax), second, meaningful, implying compliance with the meanings of the chosen code (semantics), and, third, truthful, implying possibility to recognize mis- or disinformation. Then new knowledge is extracted based on available information and already existing knowledge through different cognitive activities including, reasoning, learning, association. (Chen & Florodi 2013)

Cui (2019) demonstrates visual analytics process as a sense-making loop (Figure 6), which consists of and adjusted from various models, including visualization models, knowledge conversion processes and knowledge generation models for visual analytics as well as analytical processes in visual analytics.

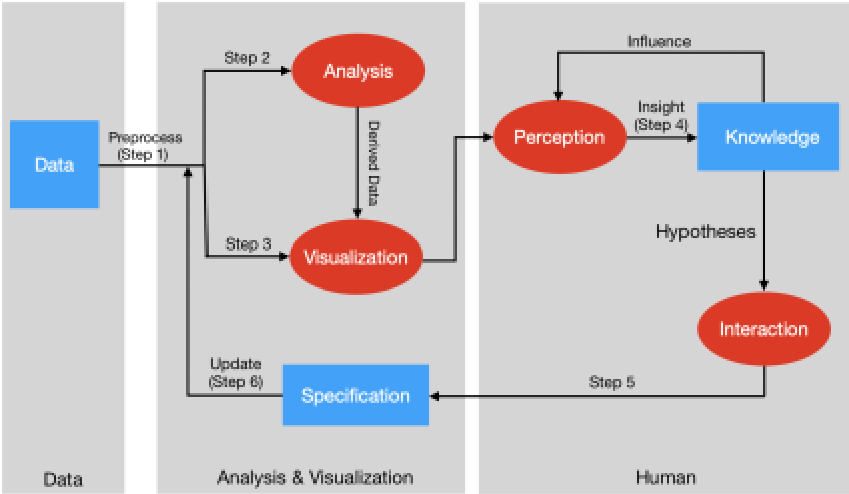


Figure 6 Visual Analytics as a sense-making loop (Cui 2019, p.81561)

Li et al. (2016) state that it is crucial to have a comprehensive framework outlining the entire analytic process. The authors argue that traditional knowledge discovery and data mining process (KDDM) models do not suffice in the modern business context and are not always able to address challenges imposed by the data-driven economy and have to be revised. They offer an updated knowledge discovery via data analytics (KDDA) process model, named Snail Shell KDDA process model (Figure 7). This model has eight phases, which are iterative and consist of different tasks. It is regarded as a project life cycle, however, sequences between phases are not strictly determined. Importantly, the project more often starts with Problem Formulation (PF). Li et al. (2016) declare that the quality of a business problem formulation can potentially impact the findings of the following stages in the KDDA process. According to the authors, business users must have reasonable expectations regarding possible achievements of analytic models and the PF phase should be introduced to direct the systematic formulation of achievable analytic problem statements based on specific business goals. KDDA model is able to deal with more aspects of the analytic model life cycle in comparison to the traditional KDDM models, as it comprises a model maintenance phase, which includes all model management activities. (Li et al 2016)

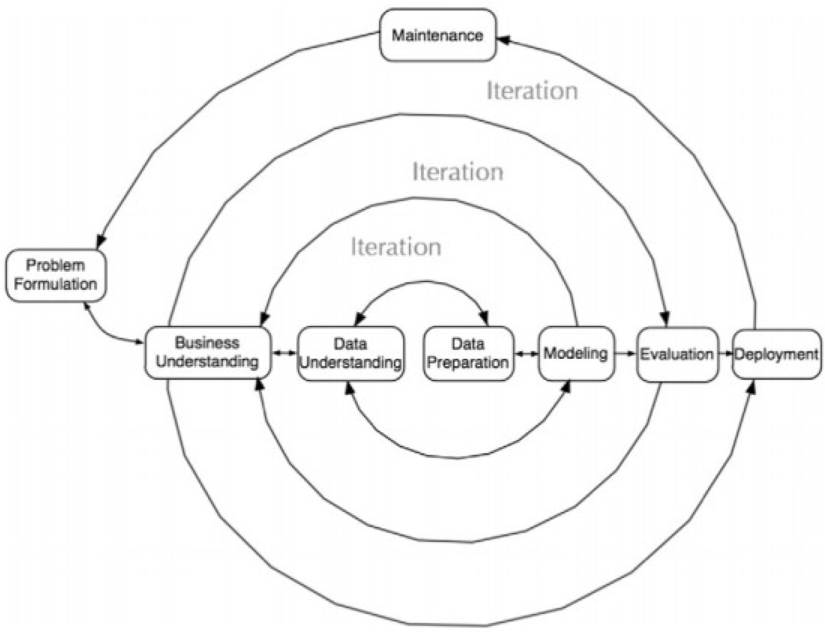


Figure 7 Snail Shell KDDA process model (Li et al 2016, p.3)

2.3.2 Tool-based drivers of knowledge discovery

Bertini & Lalanne (2010) reflect on the strengths and weaknesses of the current state of visual analytics, specifically on the most effective ways of integrating data analysis and visualization in order to facilitate knowledge discovery. Investigation is conducted from the perspective of human-machine interaction. Various data mining techniques and visualization methods are investigated while classifying them into separate categories with regard to the knowledge discovery matter. The following major categories are identified: first, *computationally enhanced visualization*, second, *visually enhanced mining*, and third, *integrated visualization and mining*. (Bertini & Lalanne 2010)

The computationally enhanced visualization category can be considered as a somewhat pure visualization technique implying that it is a primary data analysis means with supplementary automatic computation features. In terms of benefits which could be derived from automatic computation, this visualization technique can be enhanced by such automatic methods as projection (e.g., multidimensional scaling), intelligent data reduction (e.g., dimension filtering) and pattern disclosure (graph clustering). These methods could not only reduce data complexity and controlled information loss, but also enhance visual configuration by making patterns more salient. (Bertini & Lalanne 2010)

Visually enhanced mining consists in data mining as a primary data analysis providing an interactive interface for displaying the output. This approach can benefit from visualization by exploiting model presentation (e.g., nomograms) and pattern exploration and filtering (e.g., sequential patterns visualization). These methods facilitate interaction with data for its deeper understanding and gaining insights. (Bertini & Lalanne 2010)

Integrated visualization and mining imply a synergic combination of both approaches. However, strategies used for their integration can be different. Two kinds of integration are identified, first, white-box integration, which is characterized by the highest degree of collaboration between the user and the computer, second, black-box integration or in other words feedback loop, which is more common, and which provides the user with a set of solutions and guides the user by offering alternatives. (Bertini & Lalanne 2010)

Given the observed particularities of each category, the authors suggest possible contributions of data analysis and visualization in visual analytics systems. Thus, the first category, computationally enhanced visualization, which has limitations in providing more than simple pattern recognition and extending the problem for higher level reasoning, would benefit from incorporating such functions as visual model building, verification and refinement, and prediction. For the visually enhanced mining approach, the following functions could be beneficial in terms of enabling the user to involve in the knowledge discovery process at early stages: visualizing parameter space and alternatives, and model-data linking. These functions would equip systems with the direct representation of connections between parameters and results, between models and raw data as well as with visualization of structures to better compare alternatives. The third, integrated visualization and mining, category should be aimed at achieving a full mixed-initiative KDD process, since this approach already implies the complementary work of humans and computers. This category of systems can leverage collaboration of humans and machines by incorporating both visualization and data mining techniques, wherein models are built by mining techniques or derived by human perception and cognition. (Bertini & Lalanne 2010)

Ultimately, suggested contributions lead to the conclusion that both visualization and data mining equally contribute to the knowledge discovery process. However, both techniques necessitate further investigation. Particularly, the authors suggest directing research to the following issues: from the perspective of data mining, research should focus on exploring how the mining process can be augmented with flexibility to simplify exploration of alternatives, and how robustness and verification capabilities can be added to visualization. From the visualization perspective, research should explore how effective visualization can be provided so to interpret, understand and verify models, how model visualization differs from data visualization in terms of interaction techniques and design principles, and finally, how automatic data analysis can contribute to simple extraction of relevant and accurate patterns. (Bertini & Lalanne 2010)

2.4 Human discourse in visual analytics

This section studies the visualization process in terms of human cognitive abilities and interactivity.

2.4.1 Visualization and human cognition

The visual analytics process is described by the interplay between data analysis, visualization and human cognition in order to discover knowledge, which can be exploited for optimization of processes and operations, improvement of business strategies or innovations. Figure 8 depicts how human perception, cognition and analytical reasoning are exploited in visual analytics. Its essence lies in the effective integration of human analytical reasoning abilities and knowledge into the data analysis process to gain insights or discover knowledge, which would be difficult with solely visualization or analysis techniques. (Cui 2019)

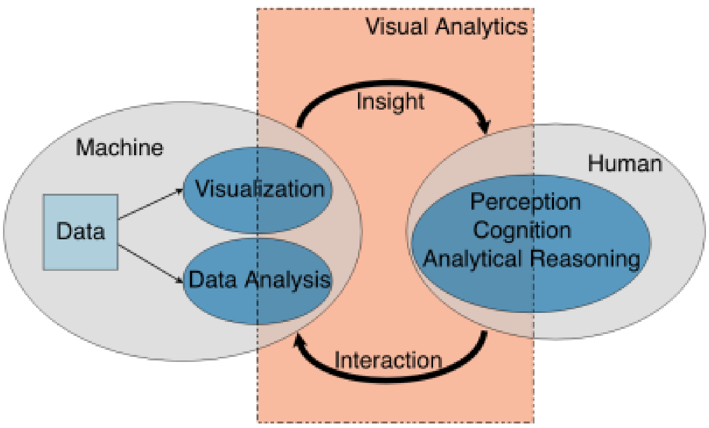


Figure 8 Visual analytics as the interaction between data analysis, visualization and human cognition (Cui 2019, p.81560)

Human perception and human abilities in pattern recognition as well as human readiness to creative discoveries and serendipities are at the core of the effectiveness in visualizations. Human cognition is almost always based on mental models, and visualization as a cognition amplifier facilitates building those mental models. (Valle 2013) Cognitive science suggests taking into detailed consideration the relationships between mental models as a special form of human's internal representation of visualizations and external representations (Bačić & Fadlalla 2016). In the context of information visualization, a mental model is defined as following:

> A mental model is a functional analogue representation to an external interactive visualization system with the following characteristics:
> • The structural and behavioral properties of external systems are preserved in mental models.
> • A mental model can preserve schematic, semantic or item-specific information about the underlying data.
> • Given a problem, a mental model of an interactive visualization can be constructed and simulated in working memory for reasoning. (Liu & Stasko 2010)

Patterson et al. (2014) view data visualization as "a human cognition-augmentation issue" and suggest that well-designed visualization leverages human visual perception capabilities and promotes reasoning and comprehension processes by involving high-level cognitive activities such as retrieval from long-term memory. From the design perspective, there have been developed design principles, which can be used to leverage human cognition in the visual analytics process. These design principles are based on numerous studies and research including Gestalt principles, visual imagery, cognitive fit and preattentive attributes (Bačić & Fadlalla 2016). Gestalt laws (e.g. proximity, similarity, continuity, closure) and related preattentive attributes (e.g. high-speed level visual perception occurring unconsciously) have been leveraged for visual representations of data. Cognitive Fit Theory implies that handling data representations involves the profound cognitive effort in user effectiveness and efficiency, and that if the problem representation matches the task type, the "cognitive fit" can be reached and this will improve the decision-making performance (Baker et al. 2009). That is, there must be a "cognitive fit" between information presentation, problem-solving task, and the user's decision processes, which can reduce the user's cognitive effort by exploiting visual displays. (Bačić & Fadlalla 2016)

Creating an information visualization for effective analysis corresponds to creating visual representations of data for "information extraction" rather than "data availability", since in "information extraction" visual representation of data facilitates human cognition by benefiting from graphical elementary cognitive codes such as length, area, angle, etc. (Patterson et al. 2014). Information visualization is also extended to high-level cognitive processes. Patterson et al. (2014) suggest a human cognition framework promoting those processes. Since human cognition entails the reciprocity between two processes: bottom-up (visual information drives pattern building) and top-down (demand in attention reinforcing relevant information) (Ware 2008), the human cognition framework for information visualization focuses on top-down processing as it "guides the way in which the bottom-up information is processed in order to activate organized knowledge structures represented in long-term memory" (Patterson et al. 2014). The human cognition framework consists of such components as exogenous attentional capture, encoding, working memory, long-term memory, pattern recognition and decision. In their cognition framework, Patterson et al. (2014) conclude that well-designed visual representations should encourage high-level cognitive functioning, such as gaining insight, making sense and understanding. For this purpose, the authors define so called "leverage points" as an access mechanism for the visualization design process, which eventually improves knowledge extraction (Patterson et al. 2014). Table 1 briefly explains the leverage points.

Table 2 Visualization leverage points defined by Patterson et al (2014)

Leverage point	Explanation
1. Capture exogenous attention	Using salient cues as signals to changes in or essential attributes of a visualization. This minimizes so called "inattentional blindness" of a user. Examples: colour, texture cues, motions, changes over time, space
2. Guide endogenous attention	Providing proper organization of information or interaction options to engage endogenous attention and reduce distractions. Important for encoding and maintaining the information in working memory.

	Examples: using clear labels for a structure of a visualization, arrow pointing to relevant information, zooming in on particular details
3. Facilitate chunking	Providing visualization parameters with strong grouping cues. Important for reducing capacity limitations of working memory and promoting relevant knowledge structures in long-term memory. Examples: use of Gestalt principles
4. Aid reasoning with mental models	Arranging material based on mental models to give strong retrieval cues for knowledge structures in long-term memory to assist in reasoning.
5. Aid analogical reasoning	Structuring information to provide strong retrieval cues for knowledge structures to help in analogical reasoning. Analogous patterns and situations are more explicit for users.
6. Encourage implicit learning	Training regimes for implicit learning about statistical regularities within a visualization.

2.4.2 Interactive visualization

Visualization is a medium for human interaction with data, involving human cognitive reasoning processes.

Tasks in visual analytics are strategic in nature and require converting into operational questions in the process of analysis (Endert, 2014). The significance of visual analysis tools is high for the knowledge discovery process by enabling the user to visually and intuitively explore datasets. As visualizations considerably rely on human cognitive abilities, they have to be highly interactive.

The theory of distributed cognition seeks to understand and explain the interaction between users and computers. It explains that knowledge is distributed among users and artefacts and that user knowledge is embodied in artefacts, and the interaction is a mediator between users and artefacts. In addition, the theory of distributed cognition highlights that cognitive processes can be distributed among other users in a group and require coordination, which is also regarded as a crucial element of visual analytics by other researchers. (Hollan, 2000; Isenberg, 2012)

Interaction is critical to data visualization, as it amplifies human cognition in active data exploration, in which the user controls information space, and these interactive manipulations enable the user to construct, test, refine and share knowledge (Dimara & Perin 2020). “Further, interaction leverages humans’ natural abilities through new visualization shapes, modalities, and input technologies, helping to make visualizations accessible to broader audiences” (Dimara & Perin 2020).

The first information visualization mantra was introduced by Shneiderman (Sun et al. 2013):

Overview first, zoom/filter, details on demand. (Sun et al. 2013)

Then Keim et al. (Sun et al. 2013) extended the mantra to visual analysis for gaining profound insights, which combines algorithmic data analysis and interactive visual interface:

Analyse first, show the important, zoom/filter, analyse further, details on demands. (Sun et al. 2013)

Dimara and Perin (2020) summarize concepts characterising good interaction in HCI. The first concept consists in *dialogue*, which views interaction as a series of communication acts between a computer through input and output and a user through action and perception. The author then defines good interaction based on Norman’s action theory, describing the route from goal formulation by the user to actions needed to move towards that goal and user’s perception and evaluation of the system’s output related to that goal. Thus, good interaction amplifies simplicity, directness and “naturalness” of the dialogue, it also gives the user a strong feeling of understanding and control (Dimara & Perin 2020). The next concept considers interaction as a *transmission* of information between the computer and the user (Dimara & Perin 2020). Good interaction is measured in transmission rate, which is “the amount of error-free information per time unit that is transferred over a communication channel in the presence of noise” based on information theory (Dimara & Perin 2020). The third concept, *control*, is based on control theory, where interaction is continuous and performed via the system comprising goals, input, output, feedback, feedforward and states, and wherein the user carries out “goal-driven actions, with feedback on the system state and feedforward on its prospect state” (Dimara

& Perin 2020). Thus, good interaction implies minimization of errors and gaps to the user's goal as well as provision of swift and steady convergence to the aimed state. *Tool use* is the concept based on activity theory and considers interaction as a set of tools allowing the user to interact with the system. Thus, good interaction is defined by boosting utility or usefulness of these tools and simultaneously increasing the user's capabilities. (Dimara & Perin 2020) The rationality paradigm and adaptation theories such as information foraging reveal that "humans tend to approximate optimal adaptation of their motor, perceptual and cognitive processes" (Dimara & Perin 2020). *Optimal behaviour* constitutes the next concept of good interaction implying maximization of utility to the best of the user's capabilities corresponding to such constraints as task, interface and environment. It has to be mentioned here that utility is not equal to efficiency in this sense. (Dimara & Perin 2020) The concept of *embodiment* is based on ecological psychology and views interaction as the act of being and participating in the world, wherein the user perceives technology as an extension of them, and where the focus lies in the user's lived experiences combining the user's intention, actions and context. Thus, good interaction implies leveraging artifacts which do not interfere with the user's pursuit and help him or her "in the wild" (Dimara & Perin 2020). The last concept, *experience*, considers interaction as an aspect constructed by the user's expectations, reactions and recall, and is based on user experience research focusing on aesthetics, emotions, stimulations, fulfilment and others, that is, pursuing a non-utilitarian objective. And thus, good interaction implies maximized satisfaction of the user's psychological needs. (Dimara & Perin 2020)

Considering these all concepts the authors further elaborate on good interaction within the scope of HCI. As user intent in HCI is not a requirement, good design of tools should explain how to use the tool, not to impose what to do with the tool. Thus, successful interaction design implies tools supporting unforeseen user intents. In visualization, user intents are more explicit, and it is often stated that insight or knowledge discovery is the ultimate goal of visualization (Dimara & Perin 2020). The authors then reveal that interaction with visualization can have several goals. The authors further distinguish HCI and visualization in terms of entities, emphasis, intent and flexibility of interaction. HCI includes two entities involved in the interplay – human and computer, and HCI concept is intended for tools simplifying the use and minimizing efforts while simultaneously increasing performance. Whereas visualization includes also the third entity – data entity, and thus computer is a mediator between human and data, where visualization has its objective to facilitate knowledge generation and sense-making. As opposed to minimizing efforts in HCI concept, visualization intends to make the user think and reflect on data being explored. Therefore, good interaction can imply in this sense effort and may ostensibly moderate interaction in order to gain deeper understanding of data. Dimara and Perin (2020) highlight that the user benefits by being provided with various means of interacting with visualization. (Dimara & Perin 2020)

According to Ware (2004), interactive visualization consists of four interlinking feedback loops: *data manipulation loop* at the lowest level, which involves basic eye-hand coordination skills to select and place objects; then, *exploration and navigation loop* at the middle level, where analysts orient themselves in the visual data space; and finally, *problem-solving loop* at the highest level, where hypotheses are formulated, and visualization is refined. This third process may consist of several visualization cycles when adding new data or reformulating a problem or finding a possible solution. Visualization can externalize the problem at hand, as a result this can enhance the cognitive process.

Knowledge discovery as any form of discovery has a serendipitous nature and might ensue provided that right factors, which are difficult to control, are present and aligned. However, there are also certain factors, which are controllable, and, thus, maximizing those factors in design can facilitate the process of knowledge discovery (Begoli & Horey 2012). The former factors, which are difficult to control, are intuition, acuteness, probability of observation, the controllable factors include availability of proper tools, comprehensibility, organization and layout of data as well as domain expertise (Begoli & Horey 2012). There are certain principles based on these controllable factors, maximization of which enables the user to explore, analyse and interact with data in a more effective and easier way (Begoli & Horey 2012). Ware (2004) defines several principles and concepts, whose incorporation into the interface make the process of interaction with visualization more efficient. For instance, the Hick–Hyman law provides an understanding of choice reaction time in relation to the number of choices. According to this law, a person provided with more information has a longer decision-making process than a person provided with less information. Another principle is Fitt's law, which explains a relationship between task difficulty and movement time. Task difficulty is calculated as the index of difficulty specified by the distance to the targets and their size. That is, a simple selection task can take more time, if the target gets smaller. (Ware 2004, p. 318-320) Thus, incorporating these rules into interfaces can improve the interaction process in that the optimal amount of information is presented for quick decision-making and task completion time is optimised.

Another important concept applied to interfaces is Gibson's concept of affordances, which is extended from the properties of the physical environment to the user interface design and representation of the data. (Ware 2004, p.18-20, p.327) Affordances are properties of objects displaying what actions the user can take. Different kinds of interface icons, pictograms, fields and buttons present a diverse group of visual affordances. These affordances should be easily perceived simply by looking without any instruction.

By and large, interaction is a fundamental aspect in visual analytics. To integrate human perception and reasoning into data analysis in visual analytics, users apply their judgement to the data analysis process through interactions with visualization, such as setting and changing the parameters of data models, filtering data, zooming and hover queries, distortion and elision techniques, and other techniques, which play a crucial role in solving the focus-context problem.

2.5 Visual discovery by design

Visualization encourages creative thinking, which is not always associated with data analysis. Yet, it can result in seeing interesting patterns and gaining valuable insights. Visual discovery by design is not accidental and can be regarded as a creative process. It requires diligence, perseverance and design thinking (Ryan 2016). Visual discovery by design is defined as "a method for the practical and creative resolution of problems that is custom-built to meet an intended result" (Ryan 2016, p. 244). It is compared to excellence, which is not an act, rather it is intended by deliberate planning and designing. As organizations become more data-centric, they strive to find more ways to leverage data. Ryan (2016) highlights a necessary shift for such companies from self-service to self-sufficiency, which implies enabling as many users as possible within the organization to participate in the discovery process. This means that both technical and non-technical users have similar tools, access and setting in order to contribute to the discovery process. (Ryan 2016)

Discovery by design is facilitated by visualization, flexibility and collaboration. Current vendors of Analytics and BI tools focus more on providing tools which facilitate the visual discovery by design process and on delivering technology solutions, which enable both technical and non-technical users to perform more independently and contribute to the discovery process (Ryan 2016). Visualization in discovery by design is a critical amplifier and facilitator of human cognitive capabilities, it enriches the discovery process by allowing technical as well as non-technical users to access enormous volumes of heterogeneous data, collaborate and transfer knowledge. “Visualization within discovery is the vehicle to move from shared understanding to shared insight” (Ryan 2016, p. 251). Figure (Ryan 2016) introduces the four angles of visual discovery by design, which explains how interactive and collaborative sharing environment, self-sufficiency along with visualization and flexibility empower data-centric organizations to gain valuable insights and discover knowledge in order to make prompt and informed decisions.

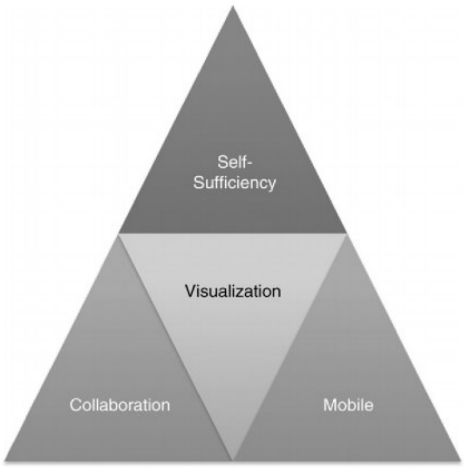


Figure 9 The four angles for visual discovery by design (Ryan 2016, p.252)

Several research projects (Heer & Agrawala 2008; Isenberg et al. 2012;) illustrate how beneficial collaborative analysis is and offer tools which promote such collaboration. The projects discuss various design considerations fostering collaboration in both synchronous and asynchronous data analysis while advocating the use of visualization during more stages in the analysis life cycle.

Daradkeh (2018) also introduces a concept of “a single seamless analytical environment” and points out that in order for a company to increase the utilization and the business value of visual analytics, it necessitates to ensure the infrastructure supporting the visual analytics tool, it needs to engage all possible users in the company by customizing training, it also requires to incorporate this tool into current business processes as well as it needs managerial support and consideration of security and privacy requirements. Although the visual analytics tool is exploited by different users in the company having mostly different access rights, they "must successfully complement one another in a single seamless analytical environment" (Daradkeh 2018).

Similar concept is pursued by Segid et al. (2012) who conceptualize the visual analytics structure by dividing it into 5 spaces (Figure 10). The components of the spaces must be designed and function in a harmonious manner to propel effective visual analytics activities.

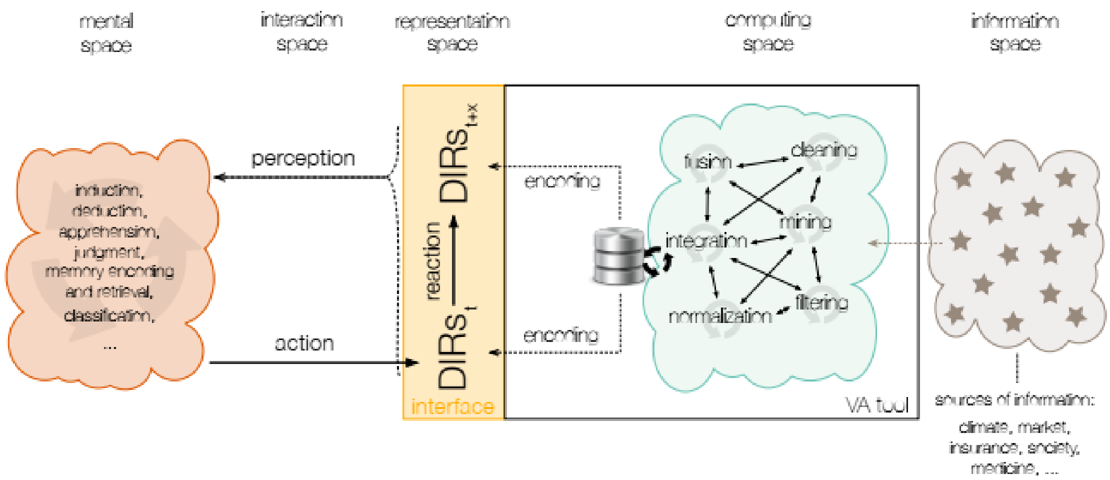


Figure 10 Spaces of the VA structure (Sedig et al. 2012, p.2)

2.6 Summary of the literature review

This section summarizes the main concepts and points relevant to the topics of the research (Table 3).

Table 3 Summary of the literature review

Findings	Reference

Visual culture of data discovery implies the balance between design, information and organizational culture with the emphasis on gaining insights, discoveries and ability to visually communicate. It embraces the concept of Enablement Technology, which in conjunction with DataViz tools allows organizations to gain insights and discover knowledge in an agile and innovative way.	Ryan (2016) Simon (2014)
Three types of data: 1) "known data" contained in daily, weekly and monthly reports providing the visibility of operations and the basic context for discussions; 2) "data you know you need to know" providing answers to questions, which can arise from the known data; 3) "data you don't know you need to know", which can be captured through interaction leading to uncovering patterns and outliers, which are invisible in previous types of data.	Murray & Chabot (2016)
Data visualization is a visual display of information influenced by the purposeful design for encoding and conveying information that would otherwise be either difficult to understand or unlikely (or impossible) to connect with in a meaningful way, and, thus, is aimed at supporting users to intuitively explore data within a visual context by leveraging the human visual system to provide insight and understanding about data, to assist in reasoning and problem solving.	Ryan (2016) Kim (2016) Patterson et al. (2014)
Visualization in visual analytics incorporates algorithmic data analysis and analytical reasoning while taking advantage of visualization and interaction to integrate human judgement into the KDD process to visually discover explainable patterns (knowledge) and to gain insight into large and complex data sets.	Cui (2019)
KDD (or KDDM) is the process of making sense out of data by exploiting data mining techniques with the purpose of discovering or extracting trends and patterns, which can be	Sacha (2014) Leece (1999)

transformed into knowledge. Visualization serves as an essential mechanism of the overall process of knowledge discovery in databases. However, merely KDD does not suffice in the modern context. KDDA (Snail Shell model) is required, which often starts with problem formulation and has a nature of a project life cycle.	Li et al. (2016)
The concept: shift from "human in the loop" to "human is the loop", which implies integrating the visual analytics systems and knowledge generation process of the user, wherein visualization is a primary interface.	Sacha et al. (2014)
The user's prior knowledge about the data, the problem domain, visual analytics tools and methodology determine the analysis strategy and procedure. The user evaluates the trustworthiness of new insight or knowledge, which must start from data collection and is critical in the course of the entire analysis process.	Sacha et al. (2014)
A typical visualization pipeline has two parts: 1) viewing - "making visible to one's eyes" – the process of defining important information, creating suitable visual representations and communicating them to the target audience; 2) seeing - "making visible to one's mind" - a viewer's cognitive processes and experiences of interpreting visuals by translating them into mental representations in order to comprehend what message they convey.	Chen & Florodi (2013)
The visual analytics process is the interplay between data analysis, visualization and human cognition in order to discover knowledge. Well-designed visualization leverages human visual perception capabilities and promotes reasoning and comprehension processes by involving high-level cognitive activities such as retrieval from long-term memory.	Cui (2019) Patterson et al. (2014)

Interaction is critical to data visualization, as it amplifies human cognition in active data exploration, in which the user controls information space, and these interactive manipulations enable the user to construct, test, refine and share knowledge. Successful interaction design implies that tools support unforeseen user intents.	Dimara & Perin (2020)
The theory of distributed cognition explains that knowledge is distributed among users and artefacts and that user knowledge is embodied in artefacts, and the interaction is a mediator between users and artefacts as well as cognitive processes can be distributed among users.	Hollan (2000) Isenberg (2012)
Knowledge discovery as any form of discovery has a serendipitous nature and might ensue provided that right factors, which are difficult to control (intuition, acuteness, probability of observation), are present and aligned. Whereas controllable factors (design) must be maximized, which can facilitate the process of knowledge discovery.	Begoli & Horey (2012)
Visual discovery by design entails the shift from self-service to self-sufficiency, which implies deliberate planning and designing of the process and enabling as many users as possible within the organization to participate in the discovery process by embracing visualization, mobility and collaboration.	Ryan (2016)

3 METHODOLOGY AND DATA COLLECTION

This chapter introduces the research design and methods exploited in the study. Further data collection and data analysis methods are described.

3.1 Research design

Building an appropriate theoretical background has been a considerable part of the research process. Thus, the research project included a vast part of literature review. The literature review consisted in the examination of various secondary data, such as books, articles, handbooks, which enabled to familiarize with key definitions, frameworks and concepts on the topic. The literature review process helped to clarify research ideas and questions. The process initially included the definition of the parameters to research questions and objectives, followed by generation and refinement of search terms. After the retrieval of scientific as well as non-scientific articles and books, evaluation and recording of relevant information were conducted. As the ideas and thoughts developed during the literature review process, the parameters for further search were redefined.

The review of principal terms, concepts and frameworks were taken from the scientific research articles and books. A summary of the most relevant and essential findings from the review is provided at the end of the theoretical background of the thesis.

The nature of the research is descriptive and explanatory, that is, it attempts to seek the answers to what-questions and investigate the relationships between two or more variables by explaining why and how certain phenomenon occurs. (Saunders et al. 2012)

This research was conducted in a combination of different approaches (Table 1). Namely, it included a deductive approach, as the research started with studying literature, an inductive approach was applied when conducting interviews to define preliminary hypotheses and new parameters for literature review. Based on interviews and literature review, a questionnaire was developed for further data collection.

Table 4 A combination of research approaches

Approach	Process	Result

Deduction	Literature review	Conceptual framework Questions for interview
Induction	Interview	Preliminary hypotheses and new parameters for further literature review
Abduction	Reflections on interviews, refining hypotheses	Questionnaire for data collection

Thus, the research design included mixed methods combining both qualitative and quantitative research. Johnson et al. (2007) define mixed methods research as following:

> Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration. (Johnson et al. 2007)

In hindsight, the research design can be defined as a sequential mixed methods research, wherein qualitative methods are followed and supported by quantitative ones (Saunders et al. 2012; Venkatesh et al. 2013). A mixed methods design for a research has numerous benefits and reasons. The definition of the mixed methods research provided by Johnson et al. (2007) illustrates general benefits for research. Venkatesh et al. (2013, p.36) suggest that mixed methods approach should be employed if the intention of a research is "to provide a holistic understanding of a phenomenon for which extant research is fragmented, inconclusive, and equivocal". Mixed methods research is able to provide more solid inferences than single method design (Venkatesh et al. 2013). One of the advantages in the mixed methods research is the opportunity to combine deductive and inductive approaches to finding more accurate answers to research questions (Denzin & Lincoln 2005). The reason and the advantage of the chosen mixed methods design for this research consists, particularly, in the initiation. That is, this kind of methodology allows to determine the nature and the scope of sequential research by providing contextual background and better understanding of a research problem. In addition, it can help in formulating or refining research questions, interview questions and questionnaire items. Namely, the purpose of the mixed methods design for the study is developmental, which implies that a qualitative study is used to develop constructs and hypotheses and a quantitative study is conducted to test the hypotheses (Venkatesh et al. 2013). Other reasons for the choice of the mixed methods research are complementarity and expansion, which enables findings and meanings to be confirmed, illustrated and linked as well as the understanding from one study to be explained and expanded (Saunders et al. 2012; Venkatesh et al. 2013). Besides, the combination of the research methods provides triangulation by ascertaining whether the findings from one method corroborate the findings from another method. That is, the validity of the qualitative methods can be strengthened by the application of the quantitative method. (Saunders et al. 2012)

3.2 Data collection

Through the qualitative methods, data were collected from different scientific articles, papers and books. Non-scientific publications, such as whitepapers and reports, were also analysed. As a result of literature review, the conceptual framework was developed, which served as a direction for compiling a set of questions for interviews. The qualitative methods also included conducting semi-structured interviews, leading to developing and refining questions for the questionnaire, which composed a quantitative research method.

The target persons and group for interviews and the questionnaire are business and data analysts from different industries and with various experiences. The reason for the choice of semi-structured interviews is that they provide more freedom to interviewees to describe and discuss ideas, processes and relationships without losing focus on the specific topics. The main factor for the choice of an interviewee was his or her experience with a visual analytics tool and the use of this tool for the explorative data analysis. The interviewees were found through the author’s personal contacts and LinkedIn. A preliminary short conversation was conducted to ensure that a person’s experience suited to the research objective. Open-ended questions were sent to the interview participants in advance, which gave them an opportunity to contemplate on the topic and recall experiences.

The qualitative methods were followed by the quantitative data collection, as supposed in the mixed methods research. A study can benefit from the use of both qualitative and quantitative methods by leveraging the complementary strengths of both approaches. A qualitative data collection approach, particularly in the form of interviews, can deepen the study by proving insights from abundant narratives, whereas surveys, a quantitative data collection approach, can provide a breadth to a study by enabling to collect data about different aspects of a phenomenon from numerous respondents. (Venkatesh et al. 2013)

Quantitative research is applied to examine also relationships between variables. It is primarily associated with survey research strategies conducted in the form of questionnaires, structured interviews or structured observations (Saunders et al. 2012). For this research, a survey in the form of an online questionnaire was conducted to collect more data. The survey was based on the results from the interviews and literature review. As questionnaires are specifically useful for descriptive and explanatory research and enable examination and explanation of relationships between variables, this method of data collection supported the objective of the study (Saunders, 2012).

3.2.1 Interviews

In total, four interviews were held. The interviews were conducted remotely via Zoom and WhatsApp and lasted approximately one hour. One interview was undertaken via email due to the time zone difference, as a person is located in the United States. All the interviews were held in English, except one interview, which was in Russian, and was translated later into English for the transcript. The interviews adhered to an interview guide comprising a set of predefined questions. However, the procedure was flexible and enabled the participants to provide more details and, as the conversation developed, follow-up and clarifying questions were asked in addition to the predefined questions. The conversations were audio-recorded and subsequentially transcribed. This method of data collection was aimed at gathering more opinionated and valuable insights from the experts in business and data analytics. Although the number of interviewees was small, the domain experts were from different companies and industries with various experience in business analytics ranging from 2 to 9 years and with experience in working with diverse visual analytics and BI tools, which provided insightful points about the topics. The interviewed business and data analysts perform their analyses in the following realms: markets and competitors, finance, business control and operational efficiency, sales and sales operations, and insurance. The analysts synthetize data from different sources and discover insights and knowledge through fused VA tools.

Due to flexibility and the flow of a semi-structured interview, the questions and the topics were not discussed in the same order for all participants.

Validity and reliability of the interviews

The promotion of validity and reliability of semi-structured and in-depth interviews can be achieved by the supply of the relevant information to interviewees before the interview (Saunders et al. 2012). The prior provision of research topics and themes to interview participants gives them an opportunity to prepare for the interview. In addition, as validity also refers to the extent to which a researcher is aware of an interviewee's knowledge and experience, and is able to derive meanings from a participant's intentions, a brief conversation took place before the interview and its preparation in order to understand the analysts' tasks and day-to-day work practices and the context they perform their analyses (Saunders et al. 2012).

The interview included the set of predefined questions (Appendix 1) comprised ten questions, nine were open-ended, the last question was for ranking the features of VA tools in their degree of importance for the effective process of knowledge discovery. The open-ended questions were provided to the interviewees beforehand. In addition, various follow-up or clarifying questions arose during the conversations.

The interview participants

The Table 5 breaks down the information about the interviewees.

Table 5 Information on interview participants

	Role	**Experience (years)**	**Area**	**Company size**	**Tools**
Interviewee 1	Data Analyst	5	Business control, Finance, Operational efficiency	250 and more	Microsoft Power BI, Tableau, Alteryx
Interviewee 2	Business Intelligence Analyst	3	Markets and competitors	250 and more	Microsoft Power BI, Excel
Interviewee 3	Business Analyst	9	Sales and Sales operations	250 and more	Tableau, Salesforce TableauCRM

Interviewee 4	Business Analyst	2	Insurance	250 and more	Tableau, Excel

The first interviewee is an experienced data analyst in a large Finnish company. They use Microsoft Power BI and Tableau for their analyses. Several years ago, they had a sizable project for implementing the integration of the advanced VA tools into existing systems and operations in order to leverage the data. The primary area, in which this data analyst conducts analyses, is business control, finance and operational efficiency. The data analyst has 5-year experience in business analytics. The second interviewee, a business intelligence analyst, works also in a large Finnish company, however, they are at the beginning of their journey to the full integration of the advanced VA tool Microsoft Power BI. They have partially implemented the project so that their analysts have started to exploit the tool. Yet, the analyses are conducted primarily through Excel, where their principal deliverable is PowerPoint presentation connected with Excel via think-cell add-on. This business analyst performs data analyses on markets and competitors and has 3-year experience in business analytics. The third interview participant works as a business analyst in a large company in the United States. The company utilizes Tableau within the Salesforce environment. The interviewed business analyst has considerable experience with VA tools (9 years in different companies) and currently makes data analyses in the realm of sales and sales operations. The fourth interviewed business analyst conducts data analyses in the realm of insurance by exploiting various tools including Tableau and Excel. The analyst's experience in the role comprises 2 years.

3.2.2 Online questionnaire

A survey in the form of an online questionnaire took place after the interviews. The online questionnaire was conducted in Google Forms and consisted of rating and ranking questions (Appendix 2). The questionnaire comprised three sections: first, the background of the respondents and the context of their analytics work; second, the questions related to the first research question regarding drivers of knowledge discovery from the analyst and VA tool perspectives; third, the questions related to the third research question regarding visual discovery by design.

For the rating and ranking questions, the five-point Likert scale was adapted to measure the extent of the agreement or importance. The five-level sentiment levels chosen for the questionnaire were considered as being optimal for this research, as it provides a reasonable array of opinions which, in turn, can help in better understanding of respondents' attitudes towards a phenomenon at hand.

Respondents

The table below briefly illustrates information on the respondents including their background and the context of their work.

Table 6 Information on questionnaire respondents

Total number	Experience (years)	Area	Company size	Tools
13	2 - 13	Financial, Controlling, Marketing, Sales, Insurance Local Government	50 – 249 250 and more	Microsoft Power BI (70%), Tableau (40%) Excel (85%) Other tools

Validity and Reliability

In order to ensure the achievement of the target audience and to prevent irrelevant responses, the preliminary description of the goal of the survey was provided. As the number of responses is low, there is a threat that the results from the same questionnaire for a larger audience might vary. However, certain measures to reduce risks were adopted. The statements were formulated as detailed and clear as possible and were based on the results from the qualitative data collection including literature review. The prior revision and feedback were performed involving an interview participant.

As the study was designed as a mixed methods research, meta-inferences have to be developed. Drawing meta-inferences is critical and essential in the mixed methods research, which are defined as “theoretical statements, narratives, or a story inferred from an integration of findings from quantitative and qualitative strands of mixed methods research” (Venkatesh et al. 2013, p.38). There are two approaches to develop meta-inferences: bracketing and bridging. Bridging is more suitable for this research, since it was performed as a sequential mixed methods study. (Venkatesh et al. 2013) Venkatesh et al. (2013, p. 39) define bridging as a process of developing a consensus between qualitative and quantitative findings, which help to understand transitions and other boundary conditions associated with research context. Therefore, the discussion part attempts to discuss the findings from the quantitative data collection in relation to results from the qualitative data collection.

4 RESULTS

This chapter presents the results of the collected qualitative and quantitative data and provides the findings of the research.

4.1 Interview findings

The section summarizes the major findings distilled form the analysis of the qualitative data.

The qualitative content analysis of the interviews included such methods as coding, categorization and meaning condensation. Coding was performed by reading through the transcripts line by line and by tagging describing words. Condensation and categorization were applied to distil central themes and allocate them to the categories which were predefined from the literature review.

Below are the most relevant and important extracts from the interviewees' responses presented in the form of meaning condensation answering the interview questions related to the RQ1 and RQ3.

Question 4: What do you think facilitates knowledge discovery in visual analytics most?

Natural unit	Central theme
"*There are 2 sides of a knowledge product. In visual analytics, as in Power BI and Tableau, there are two sides: one side is data visualization, the other side is interactivity, which allows users to fine-tune filters, to drill and explore data. That is, it enables to conduct explorative data analysis in an interactive way and gain insights from visualized data. These both sides contribute to that how well you can get understanding of your data. Visualization gives you a possibility to instantly recognize patterns and trends in data. When you have data presented in tables, they are kind of not interpreted. That is, they have columns and rows titles, numbers have meanings, but to get something from them, you have to put some cognitive effort to understand how these numbers relate. That is, you make*	**Tool features:** Visualization and interactivity; **Analyst:** Understanding business context and the meaning of data

additional cognitive efforts to, for example, understand that 200 is more than 100. … An important aspect here is to understand the business context and the meaning of data.”	
“*I would answer simply just a person who does visual analytics. You should understand why you do certain steps. And not just browsing through data and looking at the table, but being smart enough and conscious enough about the industry, the structure of the data, and then only if you know this, you will be able to result in knowledge discovery, if not, then not.*”	**Analyst:** knowledge about industry and understanding of data
“*The visualizations provided by these tools are the most powerful aspect of the tool we use. Data by itself inherently does not aid in "telling the story" that managers and leadership would like to see. In addition to the visualizations, the flexibility is also important in that we can embed filters to isolate important variables that managers want to see.”*	**Tool features:** Visualization and interactivity
"…I think if they're clear and easy, and offer first of all simple plots and trend plots that are maybe the best way to go. And those are like the good starting points, if you want to do analysis or create a report for some management people that aren't so good at analytics. And I think the best thing is that you have many options to dive deeper and maybe with some filters you can get deeper knowledge from the data."	**Tool features:** both simple and advanced visualizations Interactivity

Central topics derived from the interviewees' answers to the question indicate that tool features including visualization capabilities and interactivity are emphasized as significant aspects in driving the process of knowledge discovery most. Two of the interview participants highlighted the principal role of an analyst in the process of knowledge discovery, wherein an analyst's domain knowledge and data understanding are crucial.

Question 5: What features must a visualization tool have to drive the process of knowledge discovery?

Natural unit	Central theme

"…an important feature of visual analytics is its interactivity…. such employees as business and data analysts, financial analysts and business controllers do need a broad functionality in a tool to be able to deeply explore data. By, for example, filtering, different interaction techniques. They need unlimited access to different, more detailed, levels of data. If a business controller had same access rights to data as, say, a department manager, he/she would not be able to gain all the insights hidden in data. That is, a summarized figure doesn't tell him/her much, as he/she is responsible for the validity of this figure and what factors influence it."	Broad functionality incl. diverse interactive techniques => **Interactivity**
"The first thing about features, ... it should be simple and intuitive. It is used not only by business analysts, but by the CEO of the company and other employees as well…. The second thing is diversity. Tools should have various functionality in terms of what kind of graphs they can provide. ... the third thing is that it should be easily integrated into your current operations."	Simple and intuitive; => **Simplicity** Diversity in visualization capabilities => **Visual competence** Integration with other systems => **Connectivity**
"The ability to filter and drill down on data whether it be temporal or a very specific sales variable (such as region, product segment, or vertical). The tool must also be very interactive and intuitive. ... In order to work with Big Data you must have a powerful tool to filter through the noise and identify the drivers for performance. By leveraging analytics tools in addition to AI we are able to do just that. We can create stories that take our specific data sets and derive insights via analytic functions such as correlation. These tools facilitate knowledge discovery in that they make the invisible visible by producing meaningful outputs for our sellers and the teams."	Diverse interactive techniques Interactive and intuitive => **Interactivity** Processing of Big Data => **Scalability** VA + AI Analytic functions => **Integrated visualization and data mining techniques**
"I think it needs to be easy to use for both the analyst and for the end user. But it should offer options, for example, for the analyst to use harder solutions for complex analysis. But for the end user, it needs to be easy to use for simple analysis."	Easy to use => **Simplicity** (for all types of users)

… maybe if I have more advanced analytics tools within the visualization tool. So for example, in Tableau, there is some project that it could execute Python script, like a customized tool, maybe that would enhance the analytic cycle."	Solutions for complex analysis => **Enhanced data mining**

Central themes gleaned from the responses to the question illustrate that the most significant tool features facilitating the process of knowledge discovery comprise interactivity and simplicity followed by visual competence, connectivity, scalability and more advanced data mining techniques. Notably, simplicity was highlighted by the interviewees who has less experience in the role of an analyst.

Question 6: In what way does the analyst impact the process of knowledge discovery?

Natural unit	Central theme
"Business analysts' knowledge about and understanding of business in general. … When these data are visualized, a user, who did visualization, can make own interpretation of data. … It is a responsibility of a person who does this visualization to illustrate data from the right side so that the target users who sees theses visualizations do not draw misleading conclusions. …Visual literacy is essential. … I would say it is an indispensable part of visual analytics. It is a perspective on art, creativity. One should understand which type of a diagram is best for which purpose, where to put values. All this is included into a necessary set of skills of a successful business analyst. … If we are talking about a specific tool, then an analyst's experience with this tool directly affects the analytic process. If a tool is new to an analyst, then it bears additional cognitive load to first understand how it works and functions. I would compare it to playing a musical instrument. If you cannot play a musical instrument, at the beginning it would be a kind of suffering and an extra load. But after achieving a certain level, you already don't think about it, you just play and enjoy it. It concerns any kind	business knowledge; visual literacy; experience with the tool soft skills

of tools. You should first learn how it functions to get the most from it. The more difficult a tool is, the more effort you should put to learn using it. First, it is important to learn how you can use it to leverage all its capabilities…. … In addition to that, I would also mention that people or soft skills are important for an analyst as well. Ability to influence and work with different people. Because when working in such area, you deal with technology and data, but at the end of the day you communicate findings to people. And you have to understand people in your organization, who will work with your reports and findings. Not everybody has willingness and commitment to learning new tools. Somebody perhaps can have difficulties with explaining what reports they need and they would rather conduct them on their own in Excel. And a business analyst can bring also a value to the organization in this way, with his/her ability to work with different people."	
"An analyst impacts this process in 100 percent way. Because if you have a good analyst who knows the tools, who knows the data, who knows the industry, then you will get amazing insights. … Here comes a question about bias, because the analysts are eager to show what they want to show or what their mangers want to see.."	experience and understanding of the tool, data, business
"The analyst places a central role in the requirements gathering process. They work in conjunction with business stakeholders to ensure an effective transfer of knowledge (both of the business and the analysis application). The analyst should have a thorough understanding of the tool as they will play a key role in the solutioning process. They need to understand how the features and functionalities of the tool will satisfy the requirements of the business. The analyst does not conduct all steps in a vacuum. It is important that the analyst be joined by solution architects and developers when making key decisions on how to provide the best solutions for the business."	cooperation with business stakeholders for effective transfer of knowledge; understanding of the tool's functionalities satisfying business requirements; collaborative work with tech people

"...So, maybe the analyst helps in this process, by creating easy reports for the end users. And in some cases, I think the analyst needs to help the end user to come to the right discovery." "...by picking the right chart types to deliver the right message. And I think the colors are also important in visualization. If you highlight some aspects of the data with the color, it would enhance visualization… it's also design work…"	visualization design appropriate for the end user => visual literacy and soft skills
"...It's very important, especially for the business analyst to know that business you're working in." "...I think working together with the senior analysts [could help]. ... I think, if you have a big team of analysts, then it's easier to share the knowledge and work together."	domain knowledge collaboration with senior analysts
"I think I spent a lot of time to try to get something out from Tableau that I wanted. Now I know what I'm doing most of the cases, so it helps a lot. And I have some different kinds of graphs that I use most. Now I know what [visualization] gives the best thresholds to end users. I think you can learn the tableau in a day or two, but you need a lot of time to get used to that and know where things are in Tableau that you need."	experience with the tool

The extracted central topics depict that an analyst significantly contributes to the process of knowledge discovery by providing business understanding and domain knowledge along with visual and data literacy. Experience with VA tools is also indicated among essential contributors followed by cooperation and collaboration with different business stakeholders as well as with senior colleagues and tech people. Communication and other soft skills of an analyst are defined by three interviewees as of an utmost importance for making the process of knowledge discovery more effective.

Question 7: Does the organizational culture impact the process of knowledge discovery? In what way?

Natural unit	Central theme
"Employees have different roles and functions in the organization. They need to get own insights and conclusions from massive datasets for different purposes. ... when building infrastructure, it is important to take into consideration end users' needs and responsibilities in terms of the organization's data. .... One should understand that any analytic process in an organization should lead to some action, decision or broader understanding.... ... when designing diagrams and graphs, one has to understand who will make what decision based on this data and what actions can be taken. That is, for what purposes one visualizes this data. When you know all this, it is much easier then for you to comprehend where you should focus. ...And to understand what requires attention, one has to have understanding of the organization's requirements and its users' needs as well as of a business domain."	Employees' needs and responsibilities understanding organization's needs and requirements and business domain
"This can affect the results. ... Because how the company positions itself, in this way you as a business analyst feel that your insights also have to show good results."	Results determined by the organization's goal
"Absolutely, not just from an adoption perspective but also how the sellers and managers interact with the data sources themselves. In order to make accurate predictions you must have accurate data. With the SalesforceCRM being the system of record the culture as it relates to the tool is important. As you have to have executive buy-in and also ensure the data input into the system is quality data."	Adoption perspective; Interaction with data
"I think it impacts in some ways. For example, in our company we have that rule for a long time, that data is a very important part in process of decision making. It used to be a bit different when I first came as an analyst to the department. But nowadays, management wants really specific data-based calculations." "....I think it comes from the upper management that wants to get better knowledge from the data. ... And that's a route from the up to the bottom." ... "I think	encouraging data-driven decision making executive advocacy

it's very important that the upper management understands and realizes the importance and transmits this knowledge and this awareness."	

The responses to the question indicate the importance of the organizational culture including the role of the executive advocacy. Consideration of end users' needs and responsibilities and their interaction with data are important aspects in building infrastructure which encourages data-driven decision making.

Question 8: Could you recall the procedure of data analysis when you gained an insight or discovered knowledge (your strategy)? Which steps did you take? What led you to an insight? What was the role of visualization in this process?

All the interview participants provided nearly similar answers regarding the steps in the analytic process prior to what led to an insight or knowledge discovery. The process starts with the understanding of the data by overviewing the structure followed by preparing, transforming and cleansing the data. Understanding of data is performed through visualization. Visualization was emphasized as an important factor in data exploration. That is, one interview participant declared that intuitive application of different simple charts and graphs to the data helped to gain insights: "*So after transformation you have the data that reflects your goals. So then, to get an insight from the data, I try to apply different generic graphs, just simple ones, to see whether it's growing or not. If needed, I go deeper into data and apply more graphs. I cannot say which one, they just come to my mind during the process."*

Some of interviewees were able to recall or state what helped or could help in knowledge discovery. For instance, one of the interviewees mentioned that knowledge gained from the previous reports enhanced the process leading to "a-ha" moment and then to knowledge discovery. Particularly, reading and learning from the reports created by other colleagues aided the own analytic process. In addition, the collaboration with senior colleagues helped to discover a new insight.

Another interview participant stated: "*As the tool does most of the work for you really the procedure is just interpreting the outputs and determining if results correlations are accurate based on subject matter knowledge. ... Using the visualization one can derive as to whether the model is accurate and is coaching is necessary…. Overall, it is imperative in my experience that the analyst possess the relevant subject matter knowledge both on the application and to some extent the business."* That is, the combination of diverse competences including subject matter knowledge, visual literacy and experience with the tool enhances the process, as the lack of those competences *"... would elongate the process and lead to procedural inefficiencies as knowledge transfer won't be as effective nor will be the requirements gathering process."*

Question 9: What do you think about visual discovery by design? Does your company implement this concept?

Natural unit	Central theme
".... the entire infrastructure incorporating all the aspects of visual analytics and taking into account the organization's and its users' needs is important, regardless of a tool. ... However, not every user has the access rights to the data to explore and drill data. In our organisation, we have different access rights to the data. That is, for example, our managers, they have limited access rights to the data, as they don't need to deeply explore the data. They just need to view reports and they have possibility to do it by months or years. And that's it. They don't have any other possibilities to explore data, not to confuse them and not to damage visuals they see. And then we have people whose job is to drill into data and that is why they have more tools to deeply dig into data. And thanks to these drilldown tools, they can find causes. "	Taking into account organization's and its users needs; Various users with different rights for data access
"... this is what we are trying to get now and it is very beneficial for the company… It is implemented at some levels at my company, there are groups of business people who have access to certain data and	Certain people with limited access to certain data;

can visualize them. ... And what is important to say here is access rights. Not everyone in the company can see the same data, of course…. It's challenging because the company is huge. The information is very diverse…"	Issue with data access
"The amount of data generated on a daily basis [in our company] is higher than most companies produce in a year. The intent to deploy Salesforce and subsequently TableauCRM was an intentional effort to help make sense out of this large amount of data that is produced. Our sales teams are extremely mobile as the company does not just operate brick and mortar stores. Our company has certainly focused on being mobile first especially since the start of the Covid-19 pandemic… It is important that every functional group have a stake in the development if they indeed have access and will be using the tool. It is not efficient to engage every possible employee in the process but rather a subset of the group known as super-users or SMEs. From a statistical perspective you need a sufficient amount of people to provide input that is relative to whole. But not too many people."	subset of the group engaged in the process is more effective
"So, it means that data are available for everyone and everyone can visualize them…. Well, yeah, it's really good. And I think we have it in some cases in Tableau. It's a really good idea, but it's hard to manage. Because there must be some kind of data rights management, that wrong people don't get data access." ... "it would speed up the process. And in some cases, you don't need even the analyst. For example, I get a lot of ad-hoc questions that are quite easy for analysis. And I think if they [business people] could have those data and knew how to use Tableau properly, they could do it by themselves. So it would free up a lot of time for me. And I could do more advanced analytics. But for this people need to get better with Tableau, need more training to do simple analytics by themselves."	issues with access rights more effective and efficient process data skills at all people levels

The responses to the questions illustrate that visual discovery by design is associated with data access issues. One of the interview participants argued that the process of knowledge discovery would be more effective if a subset of the group is engaged into the process. Another interviewee conversely stated that involving business people into the analytic process would make it more effective and efficient, provided that they have developed necessary data skills, as this could eliminate the workload of an analyst dealing with simple analysis. Instead, they could conduct more complex data analysis. In terms of different access rights to data, the interviewee mentioned that it is crucial to be mindful of end users' needs.

Question 10: Please rank the following features of a visual analytics tool in terms of the effectiveness for knowledge discovery.

The responses for the question vary in terms of the tool features. The interview participants tended to combine the features. For instance, the interviewees grouped simplicity and interactivity, or simplicity and visual competence, or scale and connectivity. One of the interviewees stated: "*I think the scale is quite important nowadays. If the tool can't handle big data, it's useless. In many cases we have millions of rows in the datasets. If the tool can't handle it, it's useless.*" Similarly, other interviewee argued: "*Scale is something that I didn't mention but it's super important because if the company is huge like ours, there are huge data and it could take couple of days to process them.*"

4.2 Questionnaire findings

The findings from the online survey relate to RQ1 and RQ3. The total number of the respondents in the questionnaire was 13. The first section of the questionnaire included information on the respondents. The description of the respondents is provided in the charts (Figure 11, Figure 12, Figure 13 and Figure 14) to illustrate their background and the context of work.

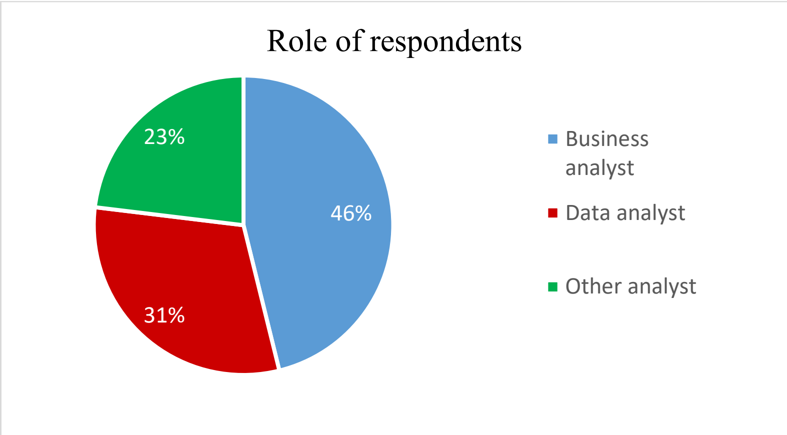


Figure 11 The role of the respondents in the questionnaire

The areas of business, in which the respondents work and conduct data analysis, include financial, controlling, sales, marketing, insurance, local government. The majority performs business analytics in the financial realm.

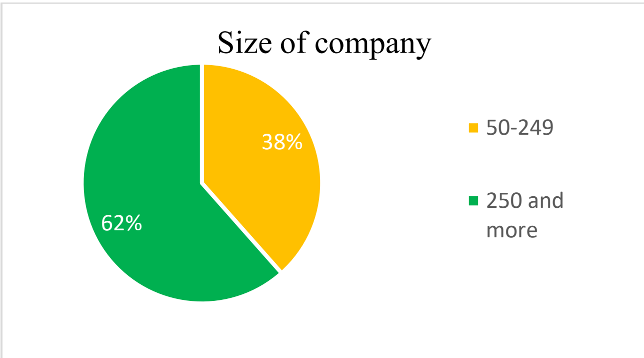


Figure 12 The size of the company

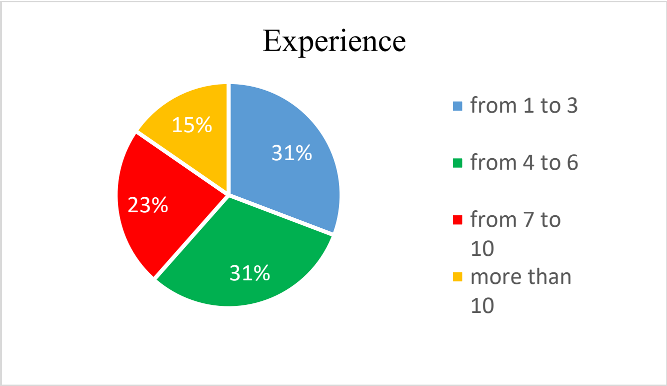


Figure 13 Experience (in years)

The respondents work mostly in large companies, their experiences in business and data analytics range from 2 to 13 years.

In terms of tools, the respondents utilize chiefly a combination of several tools. Figure 14 depicts that the most prevailing tools for data analysis are Microsoft Power BI, Tableau, Excel, QlikView and SAP Analytics.

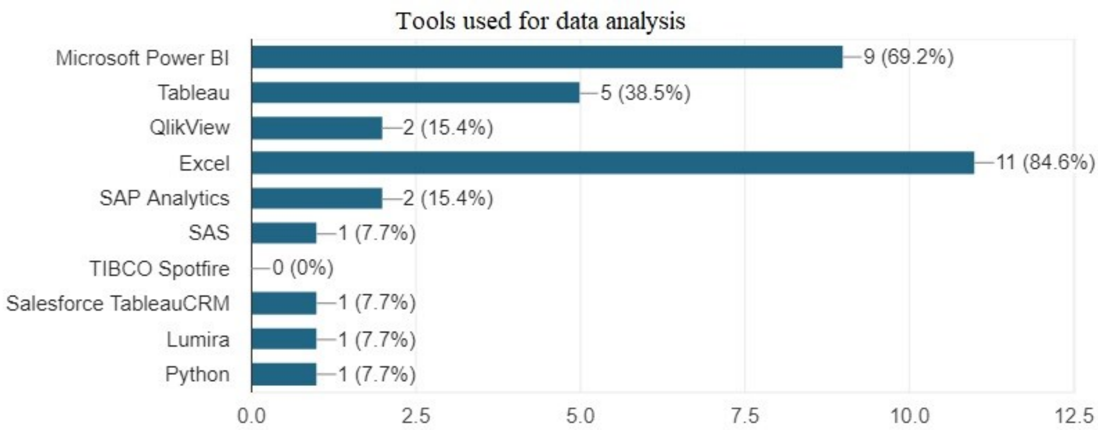


Figure 14 Tools used for data analysis

The next section of the survey was composed of four statements regarding facilitating aspects in the process of knowledge discovery. Table 7 presents descriptive statistics of the findings on the statements 6, 7, 8 and 9 by converting the items into numerical values, where 5 represents “Strongly agree”, 4 represents “Agree”, 3 performs as “Neither agree or disagree (Neutral)”, 2 performs as “Disagree” and 1 stands for “Strongly disagree”. The frequency of each item choice is visualized in Figure 15, which demonstrates the distribution of responses on the statements in terms of agreement.

Table 7 Descriptive statistics of results on Statements 6, 7, 8 and 9

	N	**Sentiment score (Median)**
S6 - Only the analyst can significantly facilitate the process of effective knowledge discovery regardless of VA tools.	13	3
S7 - Advanced VA tools with diverse features can considerably drive the effective knowledge discovery process, even if the analyst is new to the industry and to the company.	13	4
S8 - For the effective knowledge discovery process, it is important for the analyst to have a solid experience with VA tools.	13	4
S9 - VA tools can significantly facilitate the process of knowledge discovery regardless of the analyst's knowledge about and experience in the business domain.	13	4

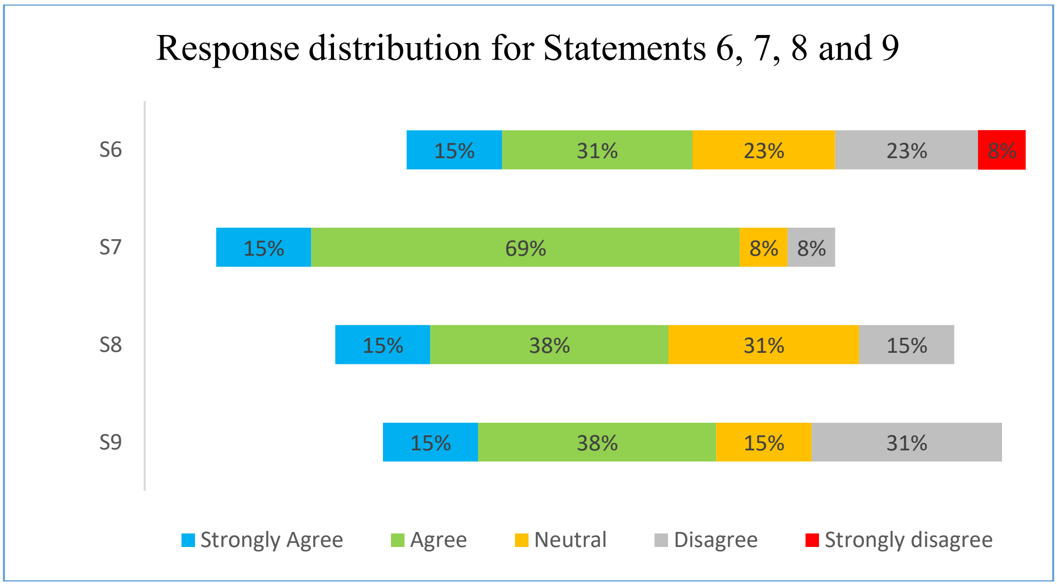


Figure 15 Response distribution in Statements 6, 7, 8 and 9

The distribution of the responses for Statement 6, which investigates whether the analyst is the primary and foremost driver in the knowledge discovery process, tends to more neutral score. Whereas the frequency of the item choice for Statement 7 clearly indicates that advanced VA tools offering diverse features can significantly drive the effective process of knowledge discovery in the opinion of the majority of the respondents. The results of Statement 8, which analyses how important solid experience with a VA tool is for driving knowledge discovery, demonstrate 53% of the respondents agree with the statement, whereas the rest tend to think neutral or disagree. Similarly, the response distribution for Statement 9 (VA tools can significantly facilitate the process of knowledge discovery regardless of the analyst's knowledge about and experience in the business domain) shows the agreement of slightly more than a half of the respondents, while the rest of participants disagrees or stays neutral in their opinions regarding this item.

The next question explored the importance of certain competencies of an analyst for the effective process of knowledge discovery. Table 8 provides descriptive statistics on Question 10 by converting the items into numerical values, where 5 represents “Extremely important”, 4 represents “Moderately important”, 3 performs as “Important”, 2 performs as “Slightly important” and 1 stands for “Not important”. Figure 16 illustrates the frequency of each item choice by providing the distribution of the responses. The item “Not important” is excluded from the legend in the visualization, as it was not chosen by the respondents.

Q10 - How important are the following competencies of an analyst in terms of facilitation of the effective knowledge discovery process?

Table 8 Descriptive statistics for Question 10

	N	**Sentiment score (Median)**
10-a - Knowledge and expertise in business and processes	13	5
10-b - Experience with the tool	13	3
10-c - Solid technical skills (diverse programming languages)	13	3
10-d - Soft skills (communication, influencing skills)	13	5

10-e - Visualization and statistical literacy (knowing about which visual representations for which purposes to use, ability to quickly interpret various visualizations)	13	5

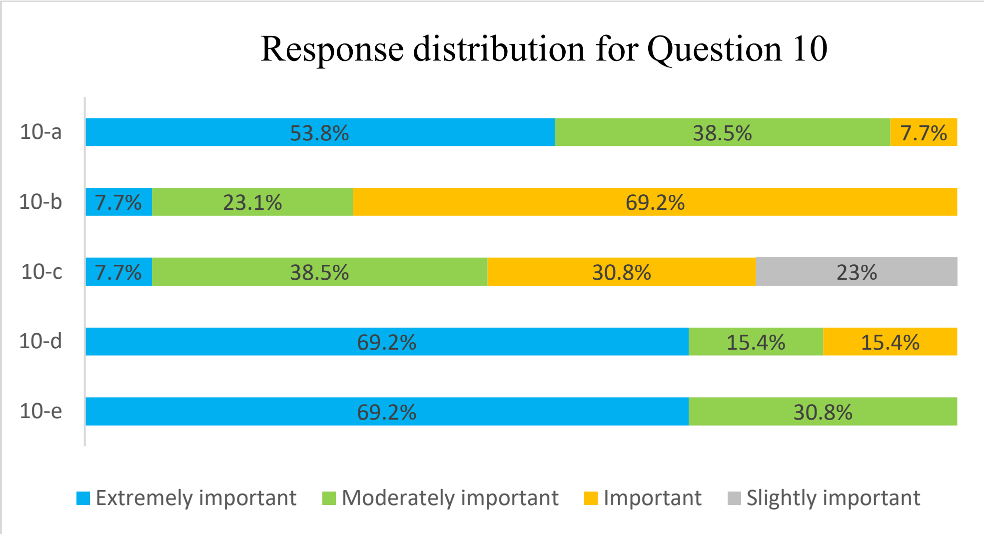


Figure 16 Response distribution for Question 10

The findings for Question 10 illustrate that the majority of the respondents selected domain knowledge and expertise along with soft skills and visual literacy as extremely or moderately important competences of an analyst for driving knowledge discovery. Whereas the importance of experience with the tool and hard skills were chosen by the minority and demonstrates the less degree.

The next question investigated the importance of the particular features of VA tools provided in Table 9.

Q11 - How important are the following features of VA tools for facilitating the effective knowledge discovery process?

Table 9 Features of VA tools included in the questionnaire

11-a - Connectivity (easy to connect to a large variety of data sources and easy to integrate with other information systems)
11-b - Simplicity (easy to use even for non-technical users)
11-c - Visual competence (provides appropriate graphics by default)

11-d - Sharing (facilitates sharing of insight)
11-e - Scalability (handles large data sets in terms of either the number or the dimension of individual data elements)
11-f - Interaction (provides maximized utility and usefulness)
11-g - Integrated visualization and mining

The frequency of each feature choice is demonstrated in Figure 17. The legend does not include the item "Not important", as it was not chosen by the respondents.

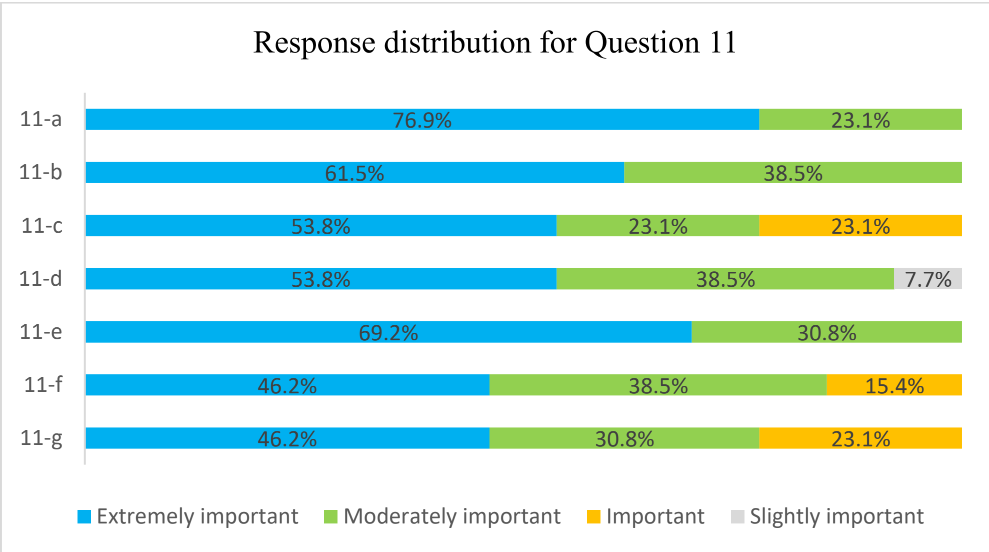


Figure 17 Response distribution for Question 11

Overall, the response distribution illustrates that all the respondents selected *Connectivity, Scalability and Simplicity* as being either extremely important or moderately important. The different degree of importance is demonstrated for such features as *Sharing, Visual competence, Interaction and Integrated visualization and mining.*

The next section included the statements aimed at exploring opinions regarding visual culture of data discovery and the concept of visual discovery by design.

Table 10 Descriptive statistics for Statements 12, 13, 14 and 15

	N	Sentiment score (Median)

S12 - Organizational culture significantly affects the process of knowledge discovery.	13	4
S13 - Collaborative and sharing environment of visual data discovery, which implies involving both analysts and other people within the company, could significantly drive the effective knowledge discovery process.	13	5
S14 - Effective visual discovery by design requires that all involved users (both analysts and business people) possess similar solid visual literacy and competence.	13	4
S15 - Building efficient visual culture is challenging due to security issues. Providing access rights, even limited, to as many employees as possible could be daunting.	13	5

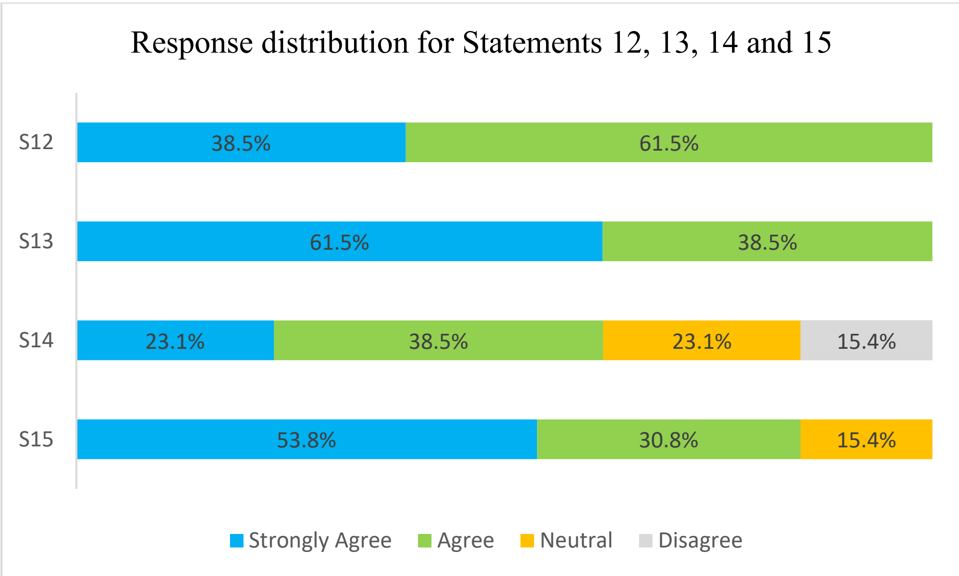


Figure 18 Response distribution for Statements 12, 13, 14 and 15

Statements 12 and 13 received the total agreement of the respondents while indicating the strong agreement in the significance of collaborative and sharing environment in a company for driving knowledge discovery. Slightly more than 60 percent of the respondents expressed their agreement regarding the requirement in possessing similar visual literacy for both analysts and business people. Approximately 8 percent of the survey participants agreed to the statement that creating a single visual and analytic environment by engaging as many people in a company as possible could be challenging due to access rights issues.

5 DISCUSSION

This chapter discusses the results from the interviews and questionnaire as well as corresponding findings from the literature review answering the research questions.

5.1 Analysis of the findings

The interview participants described the drivers of the knowledge discovery process from different aspects, which can be attributed to the fact that the interviewed business analysts have different levels of experience in the role of an analyst and particularly different experience with VA tools and use them for different purposes along with various areas of business. However, when analyzing the responses to the interview questions, several commonalities can be derived. First of all, the interviews indicate the significance of VA tools features including simplicity, interactivity, visual competence, connectivity, scalability and more advanced data mining techniques. Further, the conversations reveal that the critical role in the effective process of knowledge discovery pertains to the analyst, who possesses such crucial competences as business understanding and domain knowledge in addition to visual literacy and soft skills. The findings from the quantitative method reflect the findings from the interviews, as they also reveal that advanced VA tools offering diverse features can significantly drive the effective process of knowledge discovery. Based on the results from quantitative data, the analyst contributes to this process by providing their business knowledge and visual literacy. Soft skills along with communication and collaboration skills are also regarded as important facilitators in knowledge discovery. For instance, the results from the survey illustrate that all the respondents consider soft skills to be important, albeit to different extent. Whereas the questionnaire reveals that technical skills are of a less importance. A discrepancy in opinions regarding the importance of experience with a tool is noticed in the findings of the qualitative and quantitative data: the results from the questionnaire indicate that the respondents regard experience with a tool as insignificant for the process. Presumably, it can be attributed to the fact that the question included several competences and when choosing, the respondents had to quickly decide on the importance of a skill in relation to other competences and skills offered for this question. Whereas during the conversation in the interviews, the participants were asked clarifying and follow-up questions, which led to broadening replies and made them rethink and consider. Both the qualitative and quantitative data reveal the importance of organizational culture, particularly in the context of visual culture. Access rights are considered to be possible issues in establishing visual culture of data discovery. The concept of visual discovery by design is regarded as a significant facilitator of the process of knowledge discovery. The findings from both methods depict that this concept does not necessitate the requirement of possessing strong visual and data literacy by all people within an organization, including analysts and business people. The literature review indicates that currently it is not an absolute requirement in those competences in order to establish a visual culture of data discovery and engage many people into the process of knowledge discovery. However, the cultivation of such competences is of an utmost importance for the future development in terms of the knowledge discovery process. Undeniably, possessing data literacy has become crucial in the data-driven world and it is obvious that data and business analysts, who deals with data on another level in comparison to other business people, should have wider visual and data competences and skills.

From both the interviews and the questionnaire, a conclusion can be drawn that the user and a tool imply a synergy in order to drive the process of knowledge discovery. In this sense, the role of a VA tool is supportive, which also is identified in the literature review. Table 11 illustrates the complementary strengths of the user and VA tools in the process of knowledge discovery, which is derived from the literature review, adapted from Bertini & Lalanne (2010) and adjusted from the empirical findings.

Table 11 Complementary strengths of the user and the VA tool

Analyst	**VA tool**
define strategies	cleanse & reduce data
examine & understand data	arrange & transform data
interpret & explain	build models & visualizations
judge & measure interestingness	extract patterns, trends & visualize
generate hypothesis	verify

explore & derive knowledge	select optimal solution and configuration

As also concluded from the literature review (Keim 2010), the strengths of the user and the computer are combined in visual analytics by using their respective capabilities. However, it is an analyst who is engaged in the analytical process by interacting with the interface. From the business analyst perspective, their background knowledge and knowledge about the context play a crucial role in the effective process of insight and knowledge discovery, since it allows them to make conclusions and estimations regardless of the completeness of information available. While VA systems can provide analysts with diverse functionalities to handle massive datasets and apply various visualizations to extract patterns and trends, it is an analyst who defines strategies and methods for analysis of data. Qualitative and quantitative data depict the critical contribution of an analyst to the process, who possesses necessary competences such as subject matter knowledge, visual and data literacy, communication and collaboration skills, while advanced VA tools significantly boost the analytic process by providing diverse features. Interactive features of VA tools enable analysts to actively determine which information to display and how it should be displayed. Thus, from the perspective of VA tools, the design of their features is important for facilitating the process of insight and knowledge discovery. The success of information visualization is determined by the interactive features of VA tools, which drive interaction with the user and connect to the user's analytical process. The information, which should be perceived and displayed, is determined not only by top-down or user-driven processes (e.g., seeking relevant information), but also by bottom-up or visualization-driven processes (e.g., Gestalt principles). While advancements and analytic techniques in VA technologies aid the analysis pipeline by automating specific tasks, the increasing complexity and uncertainty in data demands the human engagement into the analytic process.

Evaluation of the strategies or procedures, which the analysts mostly employ in the explorative analytic process resulted in insight or knowledge discovery, is difficult to perform. Although the interview participants provided relatively common schema for an analytical process, it is difficult to precisely distinguish between the steps in the exploration process in terms of the use of prior knowledge or information provided. It is obvious that the steps performed in the explorative data analysis for insight and knowledge discovery are soundly connected and the analysts skip from one step to another, while formulating or altering hypothesis or searching for supporting information. They might refer to their prior knowledge (e.g. previous reports as revealed from the interviews) or they might apply information provided in the dataset at hand. This remains unclear. However, the initial steps in transforming data into information are identical for all the interviewed analysts and are displayed in the Figure 19 (adopted from Sosulski 2019):

Figure 19 Steps in transformation of data into information

The process commences with understanding the data by viewing, which is done through exploiting such visualization tools as Tableau, Power BI, R or Excel. The initial viewing enables to understand the meaning of the data and its attributes, which, in turn, helps in better formulation of questions for further visual analytical exploration. The findings from both qualitative and quantitative data indicate the significance of several features in VA tools which support this process of visual exploration eventually leading to an insight or knowledge discovery. The foremost features comprise simplicity, visualization capabilities, interactivity, that is, if the tool is easy and intuitive to use along with the provision of visual competence and interactivity, it ensures the smooth dialogue between a user and a system so that the principle of transparency is achieved, and an analyst is able to concentrate on the task at hand. Simplicity also implies ease of learning and alignment with skill level of a user. The importance of scalability and connectivity or interoperability of VA tools is emphasized, which supports the literature review.

The findings regarding the concept of visual discovery by design depict that it could significantly facilitate the process of knowledge discovery by creating collaborative and sharing environment for analytic process. However, interview participants and respondents of the questionnaire argued that building a visual culture with the emphasis on visual discovery by design can be daunting due to issues associated with providing access to data to various people levels in a business organization. The literature review reveal how knowledge discovery is lifting nowadays while demanding from companies more independence and improved agility to acquire, analyze and maintain new insights and knowledge from data, and how self-service is evolving to self-sufficiency while forcing companies to shift their organizational culture to a new level. Business users are becoming more data savvy and are constantly developing further data competences. The close work between various business users across a company could provide a deeper understanding of the data and business. Visual analytics simplifies collaboration and engagement by enabling self-sufficient users to explore the data by means of visualization and share insights, which, in turn, can decrease time to knowledge discovery.

5.2 Answers to the research questions

This section summarizes the findings and provides answers to the research questions.

RQ1: What drives insight and knowledge discovery in visual analytics?

Under this research question, the study attempted to explain the role of visual data discovery for data-driven business environment and analyze the core facilitators in the process of knowledge discovery while utilizing VA tools. As the study aimed at understanding of a holistic picture of different drivers in the effective process of knowledge discovery, the answer to this question included consideration of the analyst perspective, the VA tools perspective and the organization perspective. Hence, visual culture of data discovery is explained to understand its role for knowledge discovery.

Visual culture of data discovery within an organization consists in harmonizing design, information and organizational culture with the emphasis on gaining insights, discoveries and ability to visually communicate. The significance of data visualization is increasing, since organizations are demanded to make faster data-driven decisions and quickly deliver data-driven solutions, wherein abilities in visualizing huge volumes of multidimensional data stimulates insight and knowledge discovery. Thus, VA tools play a significant role, where KDD process requires from organizations to implement effective organizational and technological practices in order to make the process of knowledge discovery more effective. Analysts play a central role in constructing knowledge from information derived from visual data analysis. According to recent studies in visual analytics theories, the concept “human in the loop” has shifted to the concept “human is the loop”, which illustrates the merge of human work processes with analytics. The analyst’s prior knowledge about the data, the problem domain, visual analytics tools and methodology impact the analysis process. Additionally, it is a user who decides on whether the collected evidence is sufficient to trust a new insight or discovered knowledge or whether further investigation is required.

From the tool perspective, various data mining techniques and visualization methods are investigated, and their enhancement is proposed in order to facilitate the process of explorative data analysis. In sum, from the theory can be concluded that both enhanced visualization and data mining techniques equally contribute to the knowledge discovery process. The techniques for knowledge extraction and visualization vary according to data type, the objective of the analysis as well as tool design. However, the utilization of interactive techniques and how visualizations are connected is significant for the process of knowledge discovery.

The interviews and the questionnaire reveal that the analyst and VA tools are complementary drivers in the effective process of knowledge discovery. Analysts contribute to the process by providing their competences such as business knowledge, visual and data literacy, soft skills, while VA tools facilitate the process by offering such features as simplicity, scalability, interactivity, connectivity, advanced visualization capabilities and data mining techniques.

With regard to the findings from qualitative and quantitate data collection in addition to the results from the literature review, the following conclusion can be summarized by displaying in Figure 20.

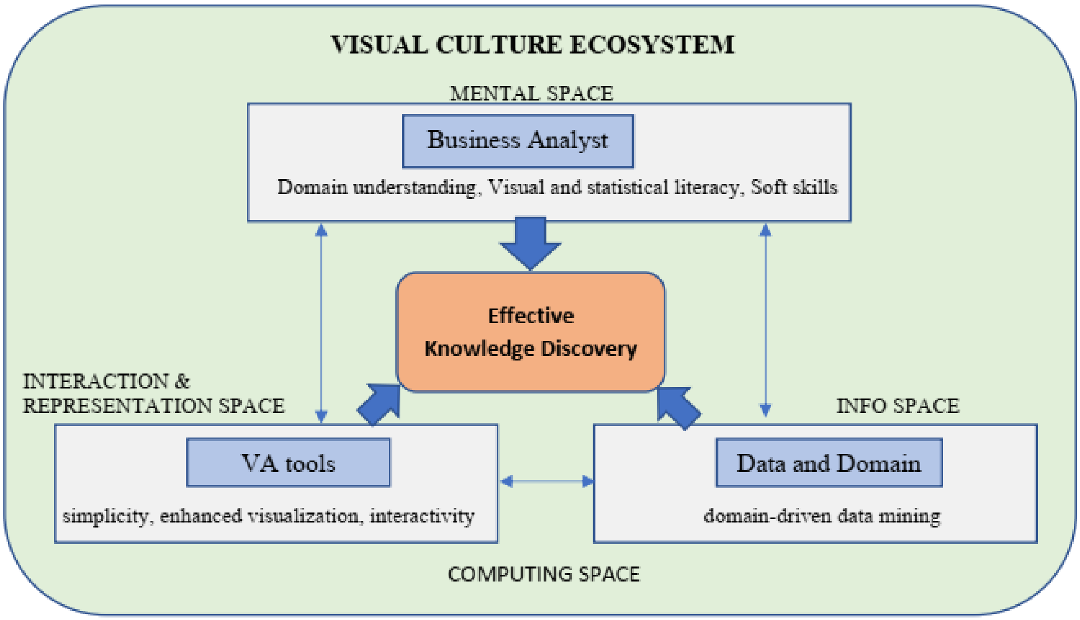


Figure 20 Driving elements in effective process of knowledge discovery

Visual culture embraces different spaces: business analyst as a mental space, VA tools as a computing space, interaction space and representation space, Data and Domain as information space, wherein computing space interlinks VA tools and data. When viewing a visual analytics environment in a comprehensive way, driving forces of effective knowledge discovery are comprised of several elements, which are interlinked with each other. First of all, efficient knowledge discovery can be achieved by involving domain knowledge and understanding of the business analyst in addition to their visual literacy and soft skills. The supportive role of VA tools in the efficient process of human reasoning is undeniable and the following features of a tool facilitate the process of insight and knowledge discovery: simplicity implying that it is exploited intuitively as well as enhanced visualization and interactivity, which provide the analyst not only with the possibility to visualize information in accordance with the type of data and the purpose of visualization, but also to explore data on different levels and display granular data while providing the possibility to alter analysis on the fly if needed. The third element consists in data and, specifically, in domain-driven data mining, which represents information space. Accumulating data and first and foremost providing access to these data in order to explore and gain insights from them is another facilitating term for effective knowledge discovery.

Those driving forces determine the effective process of knowledge discovery, provided that they function in the synergy within the single seamless analytical environment, which embraces enablement technology and visual culture of data discovery.

RQ2: Why and how does visualization simplify cognitive tasks?

Under this research question, the study attempted to understand why and how visual representation supports and facilitates human cognitive capabilities, particularly in the application of VA tools. Different frameworks, models and theories were described to answer the question. Thus, the core of the effectiveness of visualization consists in human perception and abilities in pattern recognition coupled with human eagerness to discoveries and serendipities. Human's internal representation of visualizations and external representations are explained by mental models, which are constructed and simulated in working memory for reasoning, when particularly dealing with interactive visualizations. Different design considerations have been developed in multiple previous studies to leverage human cognition in the visual analytics process. They include Gestalt principles, visual imagery, cognitive fit and preattentive attributes. In order to improve knowledge extraction in visualization, the following leveraging points have been proposed in previous studies: capturing exogenous attention, directing endogenous attention, facilitating chunking, aiding reasoning with mental models, aiding analogical reasoning, encouraging implicit learning. Interaction is considered to be an amplifier of human cognition in active data exploration. Good interaction implies simplicity, directness and "naturalness" of the dialogue between human and data, while enabling the user to easily and smoothly explore and map the data and its visualization so that the principle of transparency is achieved, by which the tool itself seems to disappear and the user ostensibly gets a direct control and thus is able to focus on the task. Various concepts and principles for interfaces have been developed to make interaction more effective. They include, for instance, Gibson's concept of affordances, the Hick–Hyman law, Fitt's law.

RQ 3: What are the implications of the visual discovery by design for knowledge generation?

Under this research question, the study endeavoured to understand the concept of visual discovery by design and explain its implications for knowledge discovery.

The concept of visual discovery by design implies the deliberate planning and designing of single analytic environment while embracing visualization, mobility and collaboration. Moreover, this concept entails engaging as many users as possible within the organization to participate in the discovery process. Visualization in discovery by design is a critical amplifier and facilitator of human cognitive capabilities, it enriches the discovery process by allowing technical as well as non-technical users to access data, collaborate and transfer knowledge. Undeniably, the concept requires consideration of security and privacy issues. The findings from the interviews and the questionnaire also reveal that organizational culture affect the knowledge discovery process and collaborative environment facilitate the analytic process leading to insights and knowledge discovery.

The concept of visual discovery by design is aimed at increasing the utilization and the business value of visual analytics in a company while building single smooth analytic ecosystem, which empower users to contribute, discover and share their insights. According to the findings from the interviews, the engagement of other business people into the process would enable analysts to focus on more complex tasks and would accelerate the effective process of knowledge discovery. Therefore, the concept of visual discovery by design has enormous implications for the process of knowledge discovery, as it requires from both analysts and users with business and specifically domain expertise to become partners in the discovery process by collaboratively engaging in the analytical processes of huge amounts of diverse data. Business organizations would significantly benefit from encouraging cooperation and sharing between users from cross-functional units in that it could boost creativity, business knowledge and awareness which, in turn, could lead to collective innovation.

6 CONCLUSION

This chapter provides concluding and final marks for the research in addition to limitations in the sturdy.

6.1 Summary

The study was aimed at describing and explaining why visualization is critical in today's business organizations for knowledge discovery and what facilitates knowledge discovery in the context of visual culture in organizations by distilling the most important findings in different standpoints of visual analytics, particularly with the interplay between analysts and VA systems. At the beginning, the paper introduced the concept of visual culture of data discovery to explain the critical role of visual analytics in the data-driven context. The research questions were answered using mixed methods approach including both qualitative and quantitative methods.

The research question regarding supportive role of visualization for cognitive tasks was investigated through literature. To answer other two research questions along with literature review, four professional business and data analysts were interviewed with the goal of clarifying the facilitators in the knowledge discovery process, the role of an analysts in this process and features of VA tools particularly needed for effective knowledge discovery. The experts' ideas about driving aspects of knowledge discovery by means of VA tools were collected in addition to their opinions about the notion of visual discovery by design. The online survey was conducted afterwards, in which 13 analysts participated and provided their responses regarding the competencies of an analyst and features of VA tools required for the effective process of knowledge discovery. Besides, the questionnaire also included the section on the concept of visual discovery by design.

By and large, visualization, specifically in visual analytics, amplifies human cognitive capabilities, which is explained through various theories and principles, such as sensemaking, cognitive fit theory, Gestalt principles, theory of distributed cognition, etc. The core driving elements in the effective process of knowledge discovery in the context of data-driven business environment were provided in the illustration in the discussion part and include such drivers as, first, the analyst possessing domain knowledge, soft skills and visual and data literacy, which represent the mental space; second, data and domain, which represent information space; third, VA tools with diverse features including interactivity, advanced visualization and data mining along with intuitiveness, which represent interaction and representation space. All these space function within a single analytic ecosystem. The concept of visual discovery by design is considered to be an essential aspect for facilitating knowledge discovery, as it emphasizes collaboration, visualization and mobility. Collaboration is regarded as a constituting component of visual analytics according to the theory of distributed cognition.

6.2 Limitations

The study has several limitations. First of all, the low number of participants in both the interviews and the questionnaire is one of the limitations. Larger sample of participants would be more representative, which in turn could provide more insights and ideas about the investigated phenomenon. The study has also limitations common for qualitative methods conducted in the form of in-depth or semi-structured interviews (Yin 2015). That is, regardless of efforts put into analyzing of qualitative data, there might be relevant findings, which were not identified. Hence, the subjectivity of interpretation of data is another limitation in this study.

REFERENCES

Ackoff, R.L. (1989). From data to wisdom. *Journal of applied systems analysis.* Vol. 16, 3-9. Retrieved from http://www-public.imtbs-tsp.eu/~gibson/Teaching/Teaching-ReadingMaterial/Ackoff89.pdf

Alspaugh, S., Zokaei, N., Liu, A., Jin, C., Hearst, M.A. (2019). Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices. *IEEE Transactions on Visualization and Computer Graphics*. Vol. 25/1. Retrieved from https://ieeexplore-ieee-org.ezproxy.vasa.abo.fi/stamp/stamp.jsp?tp=&arnumber=8440815

Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. In T. Azzam & S. Evergreen (Eds.). *Data visualization, part 1. New Directions for Evaluation*, Vol. 139, 7-32. Retrieved from http://web.a.ebscohost.com.ezproxy.utu.fi/ehost/pdfviewer/pdfviewer?vid=1&sid=4c9fa16b-b0ed-4340-a1c2-c42d7c029706%40sdc-v-sessmgr02

Bačić, D. & Fadlalla, A. (2016). Business information visualization intellectual contributions: An integrative framework of visualization capabilities and dimensions of visual intelligence. *Decision Support Systems*. Vol. 89, 77-86. Retrieved from https://www-sciencedirect-com.ezproxy.utu.fi/science/article/pii/S0167923616301038?via%3Dihub

Baker, J., Jones, D.R., & Burkman, J. (2009). Using Visual Representations of Data to Enhance Sensemaking in Data Exploration Tasks. *Journal of the association for Information Systems.* Vol. 10/7, 533-559. Retrieved from http://web.a.ebscohost.com.ezproxy.utu.fi/ehost/pdfviewer/pdfviewer?vid=1&sid=e418be87-4e00-4e1a-92f7-0c96a2a91376%40sessionmgr4007

Begoli, E. & Horey, J. (2012). Design Principles for Effective Knowledge Discovery from Big Data. *Joint Working Conference on Software Architecture & 6th European Conference on Software Architecture*. Retrieved from IEEE Xplore Full-Text PDF: (abo.fi)

Bertini, E. & Lalanne, D. (2010). Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery. *SIGKDD explorations*. Vol.11/2, 9-18. Retrieved from https://dl.acm.org/doi/abs/10.1145/1809400.1809404

Cao, G., Duan, Y., Li, G. (2015). Linking Business Analytics to Decision Making Effectiveness: A Path Model Analysis. *IEEE Transactions on engineering management*. Vol. 62/3, 384-395. Retrieved from https://ieeexplore-ieee-org.ezproxy.vasa.abo.fi/stamp/stamp.jsp?tp=&arnumber=7132744

Chang, R., Ebert, D., Keim, D. (2014). Introduction to the Special Issue on Interactive Computational Visual Analytics. *ACM Transactions on Embedded Computing Systems*. Vol. 9/4, Article 39. Retrieved from tiisintroduction.pdf (dbvis.de)

Chen, M. & Florodi, L., (2013). An analysis of information visualization. *Synthese.* Vol. 190, 3421-3438. Retrieved from https://link-springer-com.ezproxy.vasa.abo.fi/content/pdf/10.1007/s11229-012-0183-y.pdf

Cui, W., (2019). Visual Analytics: A Comprehensive Overview. *IEEE Access*, vol. 7, pp. 81555-81573. Retrieved from https://ieeexplore.ieee.org/document/8740868

Dar, P., (2020). Gartner's 2020 Magic Quadrant is Out! Check out the latest developments in Best Analytics Tools. *Analytics Vidhya.* Retrieved from https://www.analyticsvidhya.com/blog/2020/02/gartners-2020-magic-quadrant-bi-analytics-tools/

Daradkeh, M.K., (2018). Determinants of visual analytics adoption in organizations: knowledge discovery through content analysis of online evaluation reviews. *Information Technology & People.* Vol. 32/3, 668-695. Retrieved from https://www-emerald-com.ezproxy.vasa.abo.fi/insight/content/doi/10.1108/ITP-10-2017-0359/full/html

Davenport, T.H. & Bean, R. (2018). Big Companies Are Embracing Analytics, But Most Still Don't Have a Data-Driven Culture. *Harvard Business Review.* Retrieved from https://hbr.org/2018/02/big-companies-are-embracing-analytics-but-most-still-dont-have-a-data-driven-culture#:~:text=Analytics-,Big%20Companies%20Are%20Embracing%20Analytics%2C%20But%20Most%20Still%20Don%27t,According%20to%20a%20new%20survey.&text=Print-,For%20six%20consecutive%20years%20NewVantage%20Partners%20has%20conducted%20an%20annual,in%20large%20corporations%20view%20data.&text=The%20survey%20includes%20senior%20executives%20from%2057%20large%20corporations.

Dewani, R., (2020). 6 Top Tools for Analytics and Business Intelligence in 2020. *Analytics Vidhya.* Retrieved from https://www.analyticsvidhya.com/blog/2020/08/6-top-tools-for-analytics-and-business-intelligence-in-2020/

Dimara, E., & Perin, Ch., (2020). What is Interaction for Data Visualization? *IEEE Transactions on Visualization and Computer Graphics.* Retrieved from https://bib.dbvis.de/

Elmqvist, N., Moere, A.V., Jetter, H-Ch., Cernea, D., Reiterer, H., Jankun-Kelly, T-J., (2011). Fluid interaction for information visualization. *Information Visualization.* Vol. 10/4, 327-340. Retrieved from https://journals-sagepub-com.ezproxy.vasa.abo.fi/doi/full/10.1177/1473871611413180

Endert, A., Hossain, M.S., Ramakrishnan, N., North, Ch., Fiaux, P. & Andrews, Ch. (2014). The human is the loop: new directions for visual analytics. *J Intell Inf Syst.* Vol. 43, 411–435. Retrieved from https://link-springer-com.ezproxy.vasa.abo.fi/article/10.1007/s10844-014-0304-9

Fekete, J-D., van Wijk, J.J., Stasko, J.T., North, Ch., (2008). The Value of Information Visualization. *Information Visualization: Human-Centered Issues and Perspectives.* Springer, 2008. Retrieved from https://link-springer-com.ezproxy.utu.fi/book/10.1007%2F978-3-540-70956-5

Ferster, B. & Shneiderman, B. (2012). *Interactive Visualization : Insight Through Inquiry.* MIT Press. Retrieved from https://ebookcentral.proquest.com/lib/kutu/reader.action?docID=3339532

Gartner PeerInsights (2021). Analytics and Business Intelligence (BI) Platforms Reviews and Ratings. Retrieved from https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms

Gartner IT Glossary (2020). Big Data. Analytics. Advanced Analytics. Retrieved from https://www.gartner.com/en/information-technology/glossary

Green, T. M., Ribarsky, W. & Fisher, B. (2009). Building and applying a human cognition model for visual analytics. *Information Visualization*. Vol. 8/1, 1–13. Retrieved from https://journals-sagepub-com.ezproxy.vasa.abo.fi/doi/pdf/10.1057/ivs.2008.28

Heer, J. & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information Visualization*. Vol. 7, 49-62. Retrieved from https://journals-sagepub-com.ezproxy.utu.fi/doi/pdf/10.1057/palgrave.ivs.9500167

Hollan, J., Hutchins, E.., Kirsh, D., (2000). Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. *ACM Transactions on Computer-Human Interaction*. Vol. 7/2, 174–196. Retrieved from https://dl-acm-org.ezproxy.vasa.abo.fi/doi/pdf/10.1145/353485.353487

Isenberg, P., Fisher, D., Paul, Sh. A., Morris, M. R., Inkpen, K., Czerwinski, M. (2012). Co-Located Collaborative Visual Analytics around a Tabletop Display. *IEEE Transactions on Visualization and Computer Graphics.* Vol. 18/5, 689-702. Retrieved from https://ieeexplore-ieee-org.ezproxy.vasa.abo.fi/stamp/stamp.jsp?tp=&arnumber=6104042

Kandel, S., Paepcke, A., Hellerstein, J.M., Heer, J. (2012). Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18/12. Retrieved from https://ieeexplore-ieee-org.ezproxy.vasa.abo.fi/stamp/stamp.jsp?tp=&arnumber=6327298

Keim, D.A., Andrienko, G., Fekete, J-D., Görg, C., Kohlhammer, J., Melangon, G. (2008). Visual Analytics: Definition, Process, and Challenges. *Information Visualization: Human-Centered Issues and Perspectives.* Springer, 2008. Retrieved from https://link-springer-com.ezproxy.utu.fi/book/10.1007%2F978-3-540-70956-5

Keim, D.A., Mansmann, F., Thomas, J. (2010). Visual analytics: how much visualization and how much analytics? *ACM SIGKDD Explorations Newsletter*. Vol. 11/2, 5-8. Retrieved from https://dl-acm-org.ezproxy.vasa.abo.fi/doi/pdf/10.1145/1809400.1809403

Kim, M.C., Zhu, Y. & Chen, C. (2016). How are they different? A quantitative domain comparison of information visualization and data visualization (2000–2014). *Scientometrics.* 107:123–165. Retrieved from https://link-springer-com.ezproxy.vasa.abo.fi/article/10.1007/s11192-015-1830-0

Kirin, D. (2017). Lessons from Becoming a Data-Driven Organization. *MIT Sloan Management Review*. Cambridge Vol. 58/2. Retrieved from https://search-proquest-com.ezproxy.utu.fi/docview/1875399253?pq-origsite=primo

Larsen, D. & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management.* Vol. 36, 700–710. Retrieved from https://abo.finna.fi/PrimoRecord/pci.cdi_gale_infotracacademiconefile_A469808101

Leece, D. (1999). Applying Data Visualization and Knowledge Discovery in Databases to Segment the Market for Risky Financial Assets. *Managerial and Decision Economics.* Vol. 20, 267–280. Retrieved from https://web-b-ebscohost-com.ezproxy.vasa.abo.fi/ehost/pdfviewer/pdfviewer?vid=1&sid=e8fccf7e-59fd-420a-ba82-e51a8bfca294%40pdc-v-sessmgr03

Li, Y., Thomas, M.A., Osei-Bryson, K-M., (2016). A snail shell process model for knowledge discovery via data analytics. *Decision Support Systems.* Vol. 91, 1-12. Retrieved from https://www-sciencedirect-com.ezproxy.utu.fi/science/article/pii/S0167923616301233?via%3Dihub

Liu, Z. & Stasko, J.T. (2010). Mental Models, Visual Reasoning and Interaction in Information Visualization: A Top-down Perspective. *IEEE Transactions on visualization and computer graphics*, Vol. 16/6, 999-1008. Retrieved from IEEE Xplore Full-Text PDF: (abo.fi)

Ltifi, H., Mohamed, E.B. & Ayed, M. (2016). Interactive visual knowledge discovery from data-based temporal decision support system. *Information Visualization*. Vol. 15/1, 31–50. Retrieved from https://journals-sagepub-com.ezproxy.utu.fi/doi/pdf/10.1177/1473871614567794

Lycett, M. (2013). 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems*. Abingdon Vol. 22/4, 381-386. Retrieved from https://search-proquest-com.ezproxy.vasa.abo.fi/docview/1365931330/fulltextPDF/3C583AF7EACA4D09PQ/1?accountid=8166

Murray, D.G. & Chabot, Ch. (2013). *Tableau your data: fast and easy visual analysis with Tableau software.* John Wiley & Sons, Inc. Retrieved from https://ebookcentral.proquest.com/lib/kutu/reader.action?docID=1486345

Myatt, G.J. & Johnson, W.P. (2011). *Making Sense of Data III : A Practical Guide to Designing Interactive Data Visualizations.* John Wiley & Sons, Inc. Retrieved from https://ebookcentral.proquest.com/lib/kutu/reader.action?docID=697915

Nguyen, T.H., Taylor, J., & Franks, B. (2016). *Leaders and Innovators: How Data-Driven Organizations Are Winning with Analytics*. John Wiley & Sons, Inc. Retrieved from https://ebookcentral.proquest.com/lib/kutu/reader.action?docID=4659730

Patterson, R.E., Blaha, L.M.,Grinstein, G.G., Liggett, K.K., Kaveney, D.E., Sheldon, K.C., Having, P.R., Moore, J.A. (2014). A human cognition framework for information visualization. *Computers & Graphics*. Vol. 42, 48-58. Retrieved from https://abo.finna.fi/PrimoRecord/pci.cdi_gale_infotracacademiconefile_A370810929

Pohl, M., Smuc, M., Mayr, E. (2012). The User Puzzle—Explaining the Interaction with Visual Analytics Systems. *IEEE Transactions On Visualization And Computer Graphics*. Vol. 18/12, 2908-2916. Retrieved from IEEE Xplore Full-Text PDF: (abo.fi)

Ryan, L. (2016). *The Visual Imperative: Creating a Visual Culture of Data Discovery*. Morgan Kaufmann, 1st edition. Retrieved from https://www-sciencedirect-com.ezproxy.vasa.abo.fi/book/9780128038444/the-visual-imperative#book-description

Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.Ch., Ellis, G. & Keim, D.A. (2014). Knowledge Generation Model for Visual Analytics. *Data Analysis and Visualization Group, University of Konstanz*. Retrieved from https://bib.dbvis.de/

Saulles, M.D. (2019). What exactly is a data-driven organization? *CIO*. Framingham. Retrieved from https://www-proquest-com.ezproxy.utu.fi/docview/2309566094/21B7ABCB2CB84B08PQ/9?accountid=14774

Saunders, M., Lewis, Ph., Thornhill, A., (2012). *Research methods for business students*. Pearson Education Limited, Sixth Edition, 2012.

Sedig, K., Parsons, P., Babanski, A., (2012). Towards a characterization of interactivity of visual analytics. *Journal of Multimedia Processing and Technologies, Special Issue on Theory and Application of Visual Analytics*. Vol. 3/1, 12-28. Retrieved from https://www.researchgate.net/publication/235782439_Towards_a_characterization_of_interactivity_in_visual_analytics

Simon, P. (2014). *The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions*. John Wiley & Sons, Inc. Retrieved from https://ebookcentral-proquest-com.ezproxy.vasa.abo.fi/lib/abo-ebooks/reader.action?docID=1638155

Stodder, D. (2013). Data visualization and discovery for better business decisions. *TDWI best practices report*. Retrieved from *Data Visualization and Discovery for Better Business Decisions (solutiondesignteam.com)

Sun, G.D., Wu, Y.C., Liang, R.H., Liu, S.X. (2013). A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges. *Journal of Computer Science and Technology*. Vol. 28/5, 852-867. Retrieved from https://link-springer-com.ezproxy.vasa.abo.fi/article/10.1007/s11390-013-1383-8

Tableau (2020). What is business intelligence? Your guide to BI and why it matters. *Tableau*. Retrieved from https://www.tableau.com/learn/articles/business-intelligence

TIBCO (2020). What is Business Intelligence? *TIBCO Software.* Retrieved from What is Business Intelligence? | TIBCO Software

Tsai, Ch-W., Lai, Ch-F., Chao, H-Ch., Athanasios V., V. (2015). Big data analytics: a survey. *Journal of Big Data.* Vol. 2/1, 2-21. Retrieved from https://search-proquest-com.ezproxy.vasa.abo.fi/docview/1987954913?pq-origsite=primo

Venkatesh, V., Brown, S.A., Bala, H. (2013). Bridging the qualitative–quantitative divide: guidelines for conducting mixed methods research in information systems. *MIS Quarterly*. Vol. 37/1, 21-54. Retrieved from https://www-jstor-org.ezproxy.vasa.abo.fi/stable/43825936?seq=1#metadata_info_tab_contents

Yin, R.K. (2015). *Qualitative Research from Start to Finish*, *Second Edition.* Guilford Publications. Retrieved from https://ebookcentral-proquest-com.ezproxy.vasa.abo.fi/lib/abo-ebooks/detail.action?docID=2008479

Ylijoki, O. (2019). Big data – towards data-driven business. *Dissertation.* LUT University Press. Retrieved from *Ossi Ylijoki A4.pdf (lut.fi)

Z: Why Vis? Why vis? - Recent changes (dbvis.de)

APPENDICES

Appendix 1. Interview questions

Interview questions: Knowledge discovery in Visual Analytics in a data-driven business world

1. In which area of business do you conduct data analysis? (Financial, Marketing, R&D, etc.)

2. Do you use specific visualization tools? (Tableau, Power BI, etc.) Why?

3. For which purposes do you use visual analytics tools? Is it effective?

4. What do you think facilitates knowledge discovery in visual analytics most?

5. What features must a visualization tool have to drive the process of knowledge discovery?

6. In what way does the analyst impact the process of knowledge discovery?

7. Does the organizational culture impact the process of knowledge discovery? In what way?

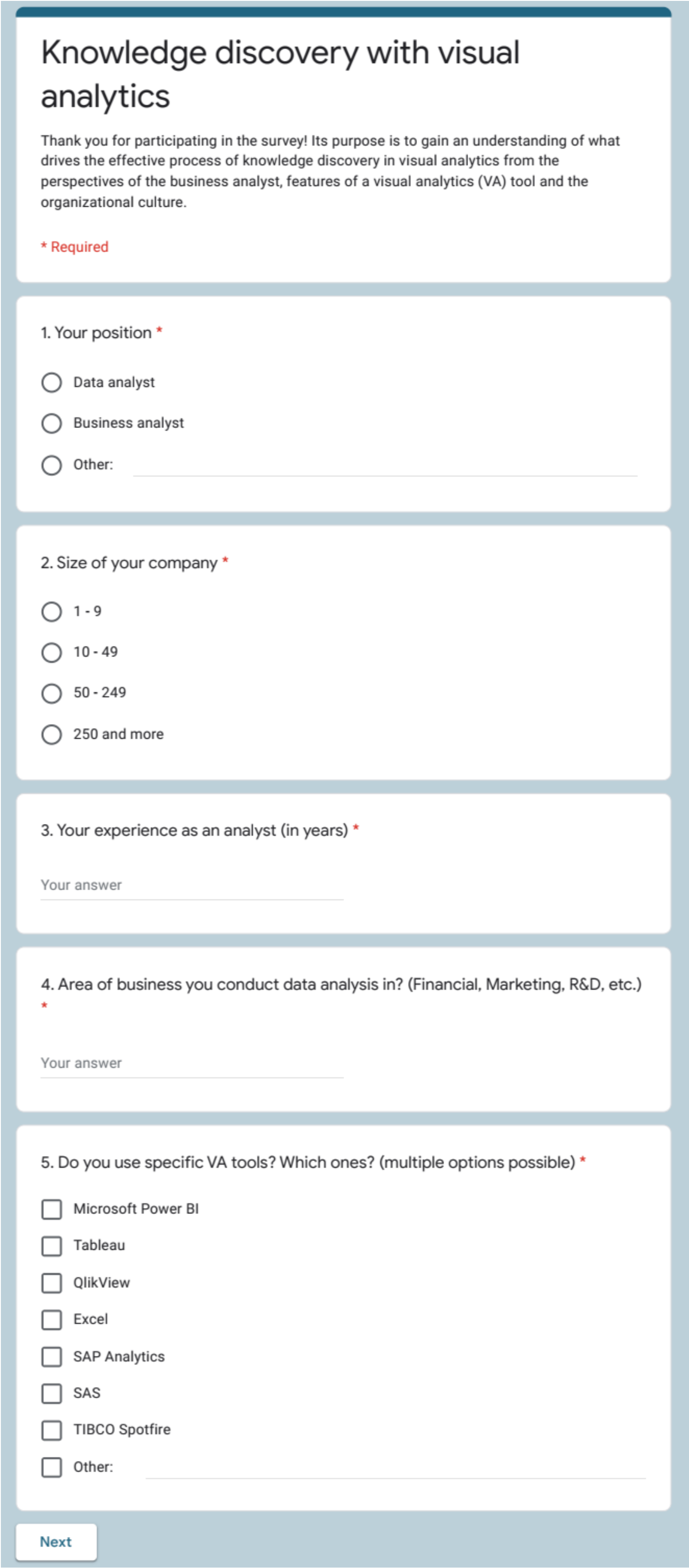
8. Could you recall the procedure of data analysis when you gained an insight or discovered knowledge (your strategy)? Which steps did you take? What led you to an insight? What was the role of visualization in this process?

9. What do you think about visual discovery by design (the intentional process of data discovery, which combines visualization (visualization capabilities of tools + visualization competence of employees), mobility (cross-platform capabilities enabling analysis on-the-fly) and interactive and collaborative sharing environment (involvement of all employees, both technical and non-technical)? Does your company implement this concept?

10. Please rank the following features of a visual analytics tool in terms of the effectiveness for knowledge discovery.

a. Simplicity (easy to use even for non-technical users).
b. Connectivity (easy to connect to a large variety of data sources and easy to integrate with other information systems).
c. Visual competence (provides appropriate graphics by default).
d. Sharing (facilitates sharing of insight).
e. Scale (handles large data sets).
f. Interaction (provides maximized utility and usefulness).
g. Computationally enhanced visualization (contains techniques which are fundamentally visual but contain some form of automatic computation to support the visualization)
h. Visually enhanced mining (contains techniques in which automatic data mining algorithms are the primary data analysis means and visualization provides support in understanding and validating the result)
i. Integrated visualization and mining (contains techniques in which visualization and mining are integrated in a way that it's not possible to distinguish a predominant role of any of the two in the process)
j. other?

Appendix 2. Online questionnaire



Knowledge discovery with visual analytics

Thank you for participating in the survey! Its purpose is to gain an understanding of what drives the effective process of knowledge discovery in visual analytics from the perspectives of the business analyst, features of a visual analytics (VA) tool and the organizational culture.

* Required

1. Your position *

- ◯ Data analyst
- ◯ Business analyst
- ◯ Other: ______

2. Size of your company *

- ◯ 1 - 9
- ◯ 10 - 49
- ◯ 50 - 249
- ◯ 250 and more

3. Your experience as an analyst (in years) *

Your answer

4. Area of business you conduct data analysis in? (Financial, Marketing, R&D, etc.) *

Your answer

5. Do you use specific VA tools? Which ones? (multiple options possible) *

- ☐ Microsoft Power BI
- ☐ Tableau
- ☐ QlikView
- ☐ Excel
- ☐ SAP Analytics
- ☐ SAS
- ☐ TIBCO Spotfire
- ☐ Other: ______

Next

Knowledge discovery with visual analytics

* Required

Business analyst and VA tools

6. Only the analyst can significantly facilitate the process of effective knowledge discovery regardless of VA tools? *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly disagree

7. Advanced VA tools with diverse features can considerably drive the effective knowledge discovery process, even if the analyst is new to the industry and to the company. *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly disagree

8. For the effective knowledge discovery process, it is important for the analyst to have a solid experience with VA tools. *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly disagree

9. VA tools can significantly facilitate the process of knowledge discovery regardless of the analyst's knowledge about and experience in the business domain. *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly disagree

10. How important are the following competencies of an analyst in terms of facilitation of the effective knowledge discovery process?

a) Knowledge and expertise in business and processes *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

b) Experience with the tool *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

c) Solid technical skills (diverse programming languages) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

d) Soft skills (communication, influencing skills) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

e) Visualization and statistical literacy (knowing about which visual representations for which purposes to use, ability to quickly interpret various visualizations) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

f) Other competencies

Your answer

11. How important are the following features of VA tools for facilitating the effective knowledge discovery process?

a) Connectivity (easy to connect to a large variety of data sources and easy to integrate with other information systems) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

b) Simplicity (easy to use even for non-technical users) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

c) Visual competence (provides appropriate graphics by default) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

d) Sharing (facilitates sharing of insight) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

e) Scalability (handles large data sets in terms of either the number or the dimension of individual data elements) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

f) Interaction (provides maximized utility and usefulness) *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

g) Integrated visualization and mining *

	1	2	3	4	5	
not important	○	○	○	○	○	very important

h) Other important features of VA tools

Your answer

Please share your comments on the previous questions.

Your answer

Back Next

Knowledge discovery with visual analytics

* Required

Visual culture of data discovery and discovery by design

Visual discovery by design implies deliberate planning and designing of the knowledge discovery process and engaging as many users within the company as possible (both technical and business people) into this process. Main characteristics of such organizational culture are visualization (use of advanced VA tools), collaboration (engagement of many employees across the company) and mobility (use of VA tools on mobile devices for on-fly analysis).

12. Organizational culture significantly affects the process of knowledge discovery. *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly disagree

13. Collaborative and sharing environment of visual data discovery, which implies involving both analysts and other people within the company, could significantly drive the effective knowledge discovery process. *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly Disagree

14. Effective visual discovery by design requires that all involved users (both analysts and business people) possess similar solid visual literacy and competence. *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly Disagree

15. Building efficient visual culture is challenging due to security issues. Providing access rights, even limited, to as many employees as possible could be daunting. *

- Strongly agree
- Agree
- Neither agree or disagree
- Disagree
- Strongly disagree

Please share your comments on the previous questions.

Your answer

Back Submit