FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Supporting sense-making and insight models for visual analytics

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktorin der Technischen Wissenschaften

by

MSc Johanna Doppler BSc
Registration Number 0626206

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Ao.Univ.Prof. Mag. Dr. Margit Pohl

The dissertation has been reviewed by:

Prof. Dr. Andreas Kerren | Prof. Dr. Achim Ebert

Vienna, 11th January, 2019

Johanna Doppler

Erklärung zur Verfassung der Arbeit

MSc Johanna Doppler BSc
Vally-Weigl-Gasse 2/3/19, 1100 Vienna

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 11. Jänner 2019

Johanna Doppler

Acknowledgements

My sincere thanks goes to my friendly advisor Margit for her constant support, encouragement and engagement, which made this thesis possible in the first place. When I write *we* in this thesis I most notably refer to the close collaboration with her that I was fortunate to enjoy during the last couple of years. I appreciate her feedback and the time she made available for me very much. I also thank my colleagues Simone Kriglstein, Guenter Wallner, Bettina Gastecker, Valerie Riegler and Isabel Dallinger, with whom I intensively worked with and from whom I could learn a lot over the years.

I am deeply thankful for the experience within the huge project, making it possible to work together with so many great people, especially William Wong, whom I thank for his repeated input and inspiration, as well as Neesha Kodagoda, Patrick Seidler, Roger Beecham and Jason Dykes whose close collaboration taught me so much. I further thank Simon Attfield, Eva Hillemann, Alexander Nussbaumer, Chris Pallaris and Peter Passmore, who I had the honour to work and publish with. It was quite a journey.

In the course of the doctorate I participated in inspiring doctoral colloquia that helped in finding the right direction in my research. I am very grateful for these brilliant opportunities and all the feedback I received from people interested in our work. The readiness to help others and energy of fellow researchers amazes me time and again. Notably, I thank Geraldine Fitzpatrick for her support and kind words as well as my colleagues at the institute who made the office an inspiring fun place where volunteers for pre-tests and feedback rounds were not hard to find.

I will always be grateful for the study opportunity and the working environment the TU Wien offered me in the last thirteen years. It brought me friends and family, whom I want to thank at last. I am very thankful for all those sympathetic ears who heard so much about the up's and down's that came along over the years. Above all it was Jörg's unstinting support that spurred my research – his critical perspective added value to my work, as did his devoted proofreading. Last but not least, he saved me countless hours with coding problems and kept my spirits up. Without your love this would not have been possible.

Kurzfassung

Die maschinelle Verarbeitung digitaler Daten stellt ein nie dagewesenes Informationspotential dar, das eine Vielfalt an neuen Möglichkeiten eröffnet. Ziel und Zweck von *Visual Analytics* ist die Verbindung von automatischer Datenverarbeitung mit interaktiven Visualisierungen zur Gewinnung von Einblicken in komplexe Sachverhalte. Der Mensch, als sogenannter *human in the loop*, ist unabdingbar um resultierende Cluster, Trends und Abweichungen zu analysieren und um semantische Informationen zu ergänzen. Der Prozess des Sensemakings, der Sinnstiftung, an sich und wie dieser am besten unterstützt wird, ist bisher jedoch weitgehend ungeklärt. Die Frage ist, welche Sensemaking-Strategien in zunehmend komplexeren Visual Analytics-Systemen angewendet werden. Empirische Untersuchungen zeigen auf welchen Einfluss die Art der Darstellung auf den Erkenntnisgewinn hat und wie der Mensch bestmöglich unterstützt werden kann. Der interaktive Umgang des human in the loop mit dem System gibt Aufschluss über seine analytische Arbeitsweisen und sein Sensemaking, wodurch er für Designempfehlungen herangezogen werden kann. In den empirischen Erhebungen der vorliegenden Doktorarbeit werden daher interaktive Prozesse während explorativer Analysen beobachtet. Das Ziel ist, dadurch ein besseres Verständnis vom Sensemaking-Prozess und dessen Strategien zu erlangen.

Die Analyse der vorliegenden Doktorarbeit beinhaltet zum einen die Untersuchung von Sensemaking-Modellen und deren Auswirkungen auf die Gestaltung von Visual Analytics-Systemen, und zum anderen ein Set von Sensemaking-Strategien, welches aus fünf empirischen Untersuchungen hervorgegangen ist. Die qualitative Analyse der Untersuchungen liefert konkrete Gestaltungsrichtlinien dafür, wie Sensemaking durch Visual Analytics unterstützt werden kann. Darüber hinaus leisten diese empirischen Studien einen Beitrag zum besseren Verständnis der untersuchten Modelle und demonstrieren zugleich die Verallgemeinerbarkeit des Strategien-Sets durch die Verwendung unterschiedlich komplexer, vor allem realistischer Systeme.

Abstract

Digital data has the potential to inform us in novel ways as computing allows us to process and analyse vast amounts of data which would not be possible otherwise. The aim of *visual analytics*, combining automatic computation and interactive visualisation, is to enable insight into complex data. The inevitable human in the loop provides contextualised human judgement to reason about discovered clusters, trends, and outliers in the data. The process of sense-making, i.e., giving meaning to one's experience, and how design influences insight, however, is not understood very well. It is an open question which sense-making strategies get employed while working with increasingly complex visual analytics systems. Empirical investigations on system design can show up how to support the human in the analysis process and how to present the available data appropriately. Investigating how users interact with a system provides an understanding of how they work and make sense of data, and consequently, informs system design with regard to the human sense-making and insight process. We, therefore, investigate interaction processes to improve the understanding of how people make sense and to describe the observed sense-making strategies.

First, this thesis includes an analysis of sense-making and insight models and their implications for the design of visual analytic tools in the form of design guidelines. Secondly, we describe a framework of sense-making strategies derived from five investigatory studies closely exploring user interaction. The qualitative analysis of our investigations yields detailed design implications for the support of sense-making and insight models in visual analytics. The empirical work not only contributes to a better understanding of these models but also demonstrates the generalisability of the framework by using systems of different, notably realistic, complexities.

Contents

CHAPTER 1

Introduction

To see what's in front of one's nose needs a constant struggle.

George Orwell [Dag]

Humans ask questions because we want to understand the world. Many efforts in the psychological and neuroscience discipline have been made to study the human mind and human thinking processes but the questions remain: How do we make sense? Where do insights come from? In the following I will introduce the terms sense-making and insight and set the scope of this thesis by describing the characteristics of Visual Analytics (VA) and basic models of sense-making and insight before the research questions and the domain of criminal intelligence analysis are discussed. Finally, I will present the methodological approach and main contributions of this thesis.

With sense-making we mean the processing and understanding of information, i.e., giving meaning to what we see, perceive and further reason with. Sense-making is related to many research areas, such as problem-solving, memorisation and learning, the visual processing system, and last but not least visualisation. Visualisation in the field of Human Computer Interaction (HCI) is concerned with supporting the human to understand and explore multidimensional datasets and helping them to gain valuable insights into their data by designing user-friendly systems, i.e., systems adapted to the users' needs. To ensure this the human visual system and information processing was the focus of many user studies, which showed up individual variabilities in our population that range, for example, from an approximately 4.5% colour vision impairment (around 8% male, 1% of the female) to issues that come up with age, such as a reduced contrast sensitivity and useful field of view [WGLL98, WGK10, War04, Sch09]. There are many more aspects to the human perception and perceptional phenomena are fairly complex - moving objects draw our attention even in the peripheral vision, whereas change blindness occurs right in front of our eyes.

Effective visual analytics systems consider not only the users' needs, but also user experience and expectation of what they want to accomplish with the tool. This is addressed in the research on sense-making and insight and, therefore, benefits utilisation of the tool and outcome of the analysis.

> Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces.
>
> Thomas and Cook [TC05, p.4]

Visual analytics offers visual representations and interaction techniques that make it possible to reach a conclusion in human judgements when dealing with complex data [KKEM10, TC05]. The task of the analyst is to solve specific problems and gain *insights* with the data at hand. The moment of insight is often described as a sudden *Aha!* experience when we get an answer to a problem we are currently working on. This light bulb moment is the point in thinking when all pieces *fall into place* and the answer suddenly becomes clear [DS03, Kle13]. A single representation can answer a small set of questions only, which is why such interfaces include various representations of the information objects, e.g., network graphs (node-link diagrams), timelines, spatial or multiple views. Fewer error rates and saving time are typical economic goals but a well designed system can do much more – especially in analytics tools the user experience influences imagination and motivation. Consequently, a tool that manages to answer various kinds of questions is crucial to produce better results.

The need to analyse large amounts of disparate data emerged from the security domain to understand catastrophic events better, such as, e.g., natural phenomena like the occurrence of hurricanes. Analysts gain insight by repeatedly looking at vast amounts of data in space and time. They iteratively explore and refine the view on the data to get new insights, thus, insight is the result of meaningful analysis. The train of thought that enables creativity and sense-making is motivated by interactivity, enabling the exploration of data in a natural manner. Therefore, user interfaces need to be specifically designed to support the interactive dynamics of real-time analytic interaction [HS12, KKEM10, PSCO09].

A considerable amount of research in cognitive psychology and HCI might guide the design of analytical systems. Unfortunately, the utilisation of already available results is not a straightforward process. Sometimes, it is not clear how basic research from cognitive psychology might be applicable due, e.g., to its abstract nature. Another problem is that psychological research is often formulated in ways accessible only by specialists [War04]. Therefore, a process of translation of this research is needed [Spe11]. Spence argues for a structured approach to translate the existing information from cognitive psychology and HCI to interaction design. In the systematic form of design actions, empirical research from cognitive psychology and HCI can be made useful for the purpose of information visualisation.

1.1 Sense-making

To understand how our brain works a body of research on human perception and cognition exists [War04]. The main input stream for information processing is the visual perception. Many characteristics of the visual processing system can be explained with evolution. We can see the most number of shades of the colour green simply because green is most dominant in nature. Our brain *sees patterns* because of evolutionary necessity. In order to survive in the wild we have to detect palatable fruits like red berries and moving prey, which we attend to immediately. These quick, *pre-attentive* processes are also responsible for phenomena such as change blindness or optical illusions, which are effective even if we strain ourselves to see "the real world" [NHT01]. However, psychological theories can not comprehensively describe sense-making in naturalistic settings. There is some research to support human sense-making with visualisation in general that brought forth recommendations on how to use colour, when to animate data or present small multiples, for example, but research specifically focusing on the sequence of interaction and on sense-making as a process is quite scarce[PDH17].

Sense-making in HCI was characterised by Russell et al. [RSPC93] as the search for information to answer task-specific questions. Russell's model of sense-making consists of a learning loop and encodings that yield a good representation. In sense-making research making inferences is described as a central factor for the interaction with visual analytics tools as they are essential to generate hypotheses. Furthermore, the process of abduction plays an important role for sense-making in visual analytics. Abduction is used by humans to find explanations for perceived phenomena [AHW10].

Our work focuses on sense-making in the individual during visual analytics working on *wicked* problems, which are ill-defined, evolving, multi-factored situations [RW73]. Most importantly, they have no stopping rules so a clear and evident solution is lacking. In intelligence analysis, for example, time is a crucial factor, however, checking all facts and suspects as they evolve is a must. There is also no immediate solution but usually more than one valid answer and it is not easy to distinguish between correct and wrong by the time. In retrospect one can also only assume if there might have been better solutions.

The theory of mental models assumes that we try to build a representational model with the information we have at hand [Joh80]. With this model in mind we then reason about, ask questions and make inferences. Gaining an understanding from data is supported by computation and visualisation of the data. A successful representation triggers questions to be answered by the representation so that we can add information to our mental model and consequently be able to gain new insights. Thus, different representations can support building mental models. A simple rearrangement of data can already lead to new insights if, for example, new groups show up in the data, where the basic interaction of sorting supported sense-making [Spe14]. On the other hand, background knowledge and individual differences in mental models of different people might lead to different insights using the same visualisation.

1.1.1 Insight

Insight is not a clearly defined term in the literature. Currently, two meanings of insights co-exist – a short moment of a new understanding and a longer process incorporating several steps towards gaining the insight. Both include the aspect of unexpectedness, in a way that it is impossible to predict when insights will occur. Chang et al.[CZGR09] discuss these two parallel meanings of insight. On the one hand, there is a notion of a sudden move from a state of not knowing how to solve a problem to a state of knowing. The common understanding of insight in psychology is the short, spontaneous moment which is referred to as "Aha!", "Eureka!" or the "light bulb moment". On the other hand, Chang et al. contrast a broader notion of insight occurring within the visualisation literature, which is regarded as equivalent to the acquisition of knowledge, which they refer to as model-confirming insight.

Many accounts consider a step-change in understanding as an insight process [May95]. North, for example, argues that insights *build up over time*, and often raise new questions that lead to further insight. He further describes insight as qualitative on multiple levels of resolution [Nor06]. It is suggested that information visualisation ought to promote both kinds of insight.

For the purpose of our research, however, choosing a definition is necessary and we think the psychological view of insights – the quick shift in the way we understand things [KPRP07] is appropriate when sense-making processes are considered as precursors. This kind of insight has also been defined as ”sudden unexpected thoughts that solve problems” [Hog01, p.251]. Insight in that sense provides a comprehension of a situation by the unconscious synthesis of prior knowledge and experience with newly collected data to create an unexpected, dramatic realisation. This means that we know more after the moment of insight than before. A common agreement on insight characteristics is that they are useful, which signifies a quantifiable measure. North describes this characteristic as relevance in the domain of the data [Nor06].

The short moment is a somewhat pragmatic notion because it can be better observed and measured. This notion is, for example, used in neuroscience, because specific cortical activity changes can be observed via functional Magnetic Resonance Imaging (fMRI). In this domain Aha!-moments are described as the feeling of relief and the sudden emergence of a solution to a problem occurring without any conscious forewarning. Tik et al. [TSL+18] distinguish a low feeling of insight from a high, emotional Aha!-moments. They interpret insight as the reward during exploitation, which follows from exploration and consequently describe sense-making tasks in two phases: the exploration phase costing cognitive demand and the rewarding phase when a solution is found.

The specific research area of problem-solving discusses insight a lot and distinguishes non-insight problems, that do not require insights to get to an understanding to insightful problems. The difference here is that two people, given the same information, come to the same conclusion following common understandings, whereas insight problems may only be solved by persons getting the insight [GM05].

Therefore, the meaning of insight is a critical factor. Ranging from a simple understanding to a ground breaking breakthrough, insights can be described on different levels, highlighting the necessity for qualitative research and high-level analyses.

1.1.2 The Notional Model of Sense-making for Intelligence Analysis

In 1999 Pirolli and Card [PC05] published the information foraging loop, incorporating seeking, searching and filtering, and, finally, reading and extracting information. Humans do not only perceive information by user-driven processes, like searching for information, but also by visualisation-driven processes, like salience or Gestalt laws. The information processing theory of Pirolli and Card's model supports that, so that processing can be driven from data to theory (bottom-up) or from theory to data (top-down).

This framework evolved into the Notional Model of Sense-making for Intelligence Analysis that organises the process of foraging and sense-making in two major loops. The foraging loop incorporates search activities whereas the sense-making loop tries to interpret the gathered information and make sense out of it to develop a consistent mental model. Pirolli and Card point out that the model brings a cost structure with it which results in a trade-off between wide exploration and detailed exploitation of the information.

The notional model may be criticised in several ways. It describes task-oriented processes via stages that build on one another. This is accomplished without the distinction between processes that are driven by the information provided and those that are driven by the user's prior knowledge. From a cognitive perspective, processes like *schematise* or *build case*, for example, should rather be classified as knowledge-driven (top-down) processes than bottom-up processes [PSM12].

Furthermore, the categorisation of the sense-making effort is problematic. Green and Fisher [GF09] argue that search tasks, which are put on the lower effort side in the model, already require complex cognition since reasoning and decision making strategies need to be applied from the very beginning. Finally, the model is not suitable to explain user interaction with information visualisations because it misses the tool and perceptual interaction that are relevant in this context.

To close this gap, the Data-Frame Theory was proposed by Klein et al. [KPRP07] to explain real-world decision making appropriately (see description in subsection 2.1.1). On this basis a new model was developed by Klein, the Triple Path Model of Insight [Kle13].

1.1.3 The Triple Path Model of Insight

Klein is a psychologist who played a crucial role in establishing the field of naturalistic decision making. The Triple Path Model of Insight [Kle13] is grounded on the Data-Frame Theory and evolved from Klein's analysis of 120 solutions for problems in demanding situations of different orders of magnitude. He observed five cognitive strategies that change the way we think and ultimately lead, often in combination, to insights:

- Connections
- Coincidences

- Curiosities
- Contradictions
- Creative desperation

These cognitive strategies form the triggers in this model which combines *Connections, Coincidences* and *Curiosities* as a trigger which changes our activities to add new information and to create a new anchor in a story. The connection path is the main pathway to insight as 85% of Klein's problems showed this kind of activity, where an implication was spotted and a new anchor was added to one's understanding. The other two paths occurred less often but *Contradiction* and *Creative desperation* were observed when a weak anchor was either used to rebuild the story or totally discarded to escape an impasse. The paths are not mutually exclusive, i.e., they can all be part of the outcome, which is a solution to a problem or a new understanding of the problem. In this sense our activities change the way we act, see and interpret the data, and also feel a greater sense of urgency or desire.

Making new connections may create new ideas or open up new pathways to consider or investigate. Klein points out that connecting the dots is easy if we know which dots need to be connected. Therefore, in a first step, analysts have to identify dots from so called non-dots, anti-dots and other ambiguities. Dots are information relevant to the problem and non-dots are irrelevant. With anti-dots he refers to information that is different, contradicting or conflicting. An anti-dot could be, for example, situational information that makes an idea seem less likely, though perhaps plausible.

A critical aspect of Klein's theory of how insights are generated is that they are up to the individual. People with the same information and background knowledge might still not gain the same insight, as it is up to the individual to identify which *dots need to be connected.* Person-related factors that might become barriers to insight are *flawed beliefs, insufficient experience, a passive stance,* and *a concrete reasoning style.* The activities described in his Triple Path Model are no guarantee for insight and the main question of why one gets an insight and the other does not remains open.

It is an open question how to design a system to overcome the above-mentioned person-related factors. Current person-related factors that might hinder insight are yet too vague and need to be further investigated. Building upon Klein's model, it is thus the main goal of this thesis to shed more light onto this grey area. Furthermore, this thesis aims to contribute further research on cognitive strategies to encourage insight, as expressed in more detail in the research questions below.

1.2 Research questions

Sense-making processes of users of visualisations are still not understood very well in general. The above-mentioned models provide a basis for empirical studies in that area.

There are still many open questions regarding the analysis of sense-making strategies. Individual differences and the generalisability to other domains are examples for such open questions. Providing additional interaction techniques such as filtering, highlighting, etc., possibly influences these sense-making strategies which is worth to investigate in an empirical way.

There is a gap in the state-of-the-art research concerning realistic sense-making with visual analytics tools. Some research on graph comprehension, e.g., Kosslyn [Kos89], Tversky [Tve05], deals with sense-making processes using visualisations. In general, generic and fairly simple tasks, such as path finding or look-up tasks are used. Differences between experts and novices, for example, lie in the former's ability to focus their attention to relevant areas. However, this applies only to tasks from their domain of expertise and are not universally transferable [GS13]. It is, therefore, crucial to determine if the task falls into the domain of the expert and, additionally, how novices can be trained to gain expertise.

To get a greater picture of sense-making processes, realistic, more complex tasks are useful for studying thinking steps and complex reasoning. Friel et al. [FCB01] developed a model consisting of three different levels:

1. reading the data (extracting data, locating data)
2. reading between the data (finding relationships, integrating data)
3. reading beyond the data (extrapolating from the data, generating hypotheses)

We make use of this model to study insights on different levels as we research sense-making in realistic settings and complex problems to reach better decisions. In collaboration with various co-authors we conducted multiple investigatory studies regarding the work of *visual analysts*. We address the following research questions:

1. How can the design of visual analytics tools support cognitive processes for intelligence analysis?
   a) Should the emphasis be on the perception of details or rather on a comprehensive mental model [KYW$^+$07, Joh96]?
   b) How should a system be designed to support connection strategies [Kle13]?
   c) How should a system be designed to support contradiction strategies [Kle13])?
   d) How should a system be designed to support creative desperation situations [Kle13]?
2. What kinds of sense-making processes get employed in visual analytics?
   a) Can sense-making strategies be observed on a general level or are they rather task and/or domain specific?
   b) Is it more helpful to adopt visualisations that are already in use in the context of intelligence analysis, like network visualisations, timelines and geographical visualisations or to develop new visualisations specifically addressing the needs of intelligence analysts?

3. What kinds of interaction processes should be supported by the system (e.g., scroll, zoom, pan, filter, reconfiguring elements on the screen)?
    a) Which of these interaction processes is especially helpful in the context of sense-making processes in intelligence analysis?
    b) How can these interaction processes be supported?
4. How useful are models of insight generation? Does the representation of the problem influence insights [HS74, ZK10]? Does using different representations lead to different kinds of insights and which visualisations are especially useful in the insight generation process?
    a) How should maps be designed to support the analysts in their work?
    b) How should social networks be designed to support the analysts in their work?
    c) How should time-oriented data be designed to support the analysts in their work?

The next section will give an introduction to the domain of criminal intelligence analysis, as well as a brief overview of analysis techniques common in this community.

1.3 Criminal intelligence analysis

Intelligence analysis is seen as a combination of sense-making and expert knowledge [PC05]. Criminal Intelligence Analysis covers sense-making for the purpose of investigation and making sense of the crime situation in general. The terminology crime analysis is used for the latter, general crime and disorder problems, rather than investigation support of single crimes. Crime analysis has been described as the "systematic study of crime and disorder problems and police-related issues – including socio-demographic, spatial, and temporal factors – to assist the police in criminal apprehension, crime and disorder reduction, crime prevention and evaluation" [BS13, p.22].

This research was undertaken in the course of the European funding programme FP7 Research & Development (R&D) project Visual Analytics for Sense-making in Criminal Intelligence Analysis (VALCRI) [VAL18], funded by the European Commission (EC). The project was carried out over the span of four years from June 2014 to June 2018 at TU Wien (Vienna University of Technology) in cooperation with sixteen academic and industry partners across Europe. The result of VALCRI is an integrated software support system for police analysts to analyse crime reports and crime-related behaviour.

1.3.1 Ethical considerations in VALCRI

Ethical considerations are crucial when algorithms are used to facilitate human reasoning and analytic discourse. In VALCRI interactive visualisations are tightly coupled with semi-automated human-mediated semantic knowledge extraction computation. The project worked with guidelines on ethics concerning the transparency of system operations from the very beginning and is, thereby, picking up the notion of transparency of Moor's [Moo85] unique characteristic of computer technology that requires the consideration of ethics in technology design and development, and provides the rationale behind the computer ethics movement.

Another ethical concern relevant to how the system operates is that of privacy, which is recognised as a fundamental human right and further recognised as vital to the pursuit of human autonomy – the very characteristic of humanness that enables *ethics* through human free will and the ability to think and act in an uncoerced and undetermined way. Regarding data protection VALCRI addresses the national specifics of core data protection principles for both end user countries, namely Belgium and the United Kingdom. Although the principle of purpose limitation is manifested in all national data protection laws of the EU member states, the wording and interpretation of this principle based on the previous European data protection Directive 95/46/EC [Dir95] can differ significantly per country. The EC triggered the process of shaping a new, coherent and common legal data protection framework, acknowledging this lack of harmonisation with regard to central principles of data protection. According to the European Directive 2016/680 [Dir16] regarding data protection in the area of law enforcement and criminal justice, law enforcement agencies are required to analyse the risks and consequences of technological advancements and personal data processing in order to implement adequate mitigation techniques. It is, however, not specified how such an impact analysis should be performed by police actors. One of the goals of VALCRI was to draft an innovative methodology and overview of this process, yielding the very first guidelines to exist specifically for law enforcement under this new legislation.

Criminal intelligence in general is a controversial topic for ethical discussions. On the one hand, pre-empting crime is a driving factor to improve security for people while, on the other hand, privacy and legal issues are crucial to protect people's rights and to prevent the misuse of personal data. Klein's sense-making research addresses a similar problem with a performance equation comprising error prevention and insight generation [Kle13]. He argues that a good balance is necessary as too much error prevention might block insight generating and that making errors is a valuable contribution. The privacy impact arising from the algorithmic processing of various data and datasets is discussed within the project, but not part of this doctoral thesis.

The emphasis of this project was to design the technology from cognitive, legal, privacy and ethical perspectives to provide a support tool for fluent analysis that enables law enforcement agencies to make their activities more transparent, while at the same time respecting the rights of the individual with respect to security and liberty. The processes by which conclusions are reached should be made easy to inspect since the collection and processing of data underlies legal, privacy and ethical guidelines. The differing national laws constitute the conditions under which personal data processing can happen in VALCRI. These guidelines along with human factors that influence the analysis were addressed in the design of the system. Our work focused on cognitive issues within the *Human Issues* work package. Analysts have to stay creative as well

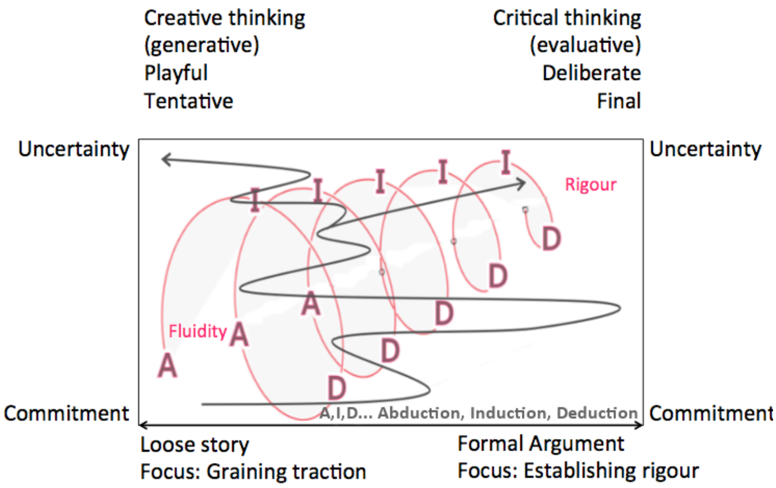


Figure 1.1: How analysts think: Inference making is not uni-directional [Won14]

as critical and reason analytically and inferentially throughout the whole thinking and reasoning process. The reasons for the analysis may change from establishing crime patterns to assembling evidence. Also the kind of analysis may change from statistical analysis to constructing an argument, which further changes the rules regarding the access and holding of data versus evidence. From a cognitive perspective the goal is to rapidly provide access to information necessary to help the analyst structure and explain a series of events.

1.3.2 Intelligence analysis

The intelligence and investigation analysis process is seen as a continuum where the purpose changes from the support for thinking and reasoning about anticipating and pre-empting crime, to support for thinking and reasoning about specific investigations. Analysts form hypotheses and build up from a *loose story* to a *formal argument* by changing the focus from playful, creative thinking to deliberate, critical thinking again and again until they reach a rigorous insight, see also Figure 1.1.

Strategic analysis of criminal behaviour aims to monitor the implementation of intervention strategies and re-evaluate changed behaviour [Cla13, KGS11, Kri99]. In that sense, intelligence-led policing focuses on three major fields: 1) the identification of short and medium term risks in form of harm for the community, 2) the identification of opportunities to improve its well-being, and 3) the evaluation of intervention strategies [SHK+16].

An analytical method is to identify patterns and anomalies in the collected evidence. This approach is called Comparative Case Analysis (CCA) and is used by law enforcement professionals for sense-making and evidential reasoning. CCA focuses on series of linked events which can be distinguished from other events to evaluate crime that has already happened, hoping to prevent future crime.

1.3.3 Crime pattern theory

According to crime pattern theory criminals stick to their Modus Operandi (MO), i.e., proven successful the perpetrator's criminal strategy is unlikely to change. The Major Incident Manual 2006 [Ass06] describes Crime Pattern Analysis as a method to identify such patterns and trends in data on crime and incidents to prevent future crime. Among other methods, the report particularly suggests visualisations like, e.g., bar charts or maps. The analyst adopts an iterative process of manipulating, searching and sorting the characteristics of crimes and trying to recognise differences and similarities in the cases of interest. The United Nations Report on Drugs and Crime 2011 [Nat11] distinguishes between four different analysis techniques: link analysis, event charting, flow analysis, and telephone analysis. The authors also point out that links between entities (e.g., persons, locations, etc.) should be represented graphically to clarify their relationships.

In the context of our work, link analysis is of special interest since a disproportionate amount of harmful impact comes from people engaging in co-offending, which is readily identified through the analysis of crime networks [AF11]. Insights on their evolution are obtained by monitoring the temporal development of co-offending networks. Boba Santos [BS13] also points out that crimes often follow distinct patterns, especially geographical patterns. Using maps to visualise crimes helps to uncover these patterns because the spatial and temporal dimensions of criminal activity can be clarified easily through such visualisations.

Crime analysis is typically done with volume crimes, i.e., crimes that occur in quantity (e.g., burglaries or petty theft) by areas, days of the week or time of the day. More advanced analytics correlate crimes by similarities of key features such as specific attributes of the MO across crime categories, which might allow to connect them to a gang or organised crime group. Crime and crime pattern analysis, therefore, concerns sense-making of space and time, including hotspot analysis, statistical process control charting and crime profiling as well as network analysis of offenders [HP14]. Hotspot analysis maps crimes to geographical areas in *space*. Statistical process control charting compares crimes across periods on a timeline and calculates statistically significant differences over *time*. In general, a hotspot is described as a popular location. Hotspot analysis is an adopted method in crime analysis to map areas of interest where crime is concentrated [BDM+17]. The interpretation of hotspot maps can be difficult when aggregation is used or occlusion occurs, design considerations which needs to be taken care of in the map design [Rot13, BDM+17].

To compare crime over time several timelines can be juxtaposed (small multiples approach), superimposed (stacked approach) and integrated, i.e., combined time steps that are inseparable without changing the layout [BBDW14]. Horizontally stacked timelines, for example, can be used to colour-code time varying weighted digraphs to provide an overview of dynamic graphs [BM15].

1.4 Methodological approach

Visual analytics systems should be specifically designed to support users in perceiving the essential information at a glance and, additionally, allow an intuitive interaction with the system at hand. Empirical research can help with this task by studying the usage behaviour and showing up where improvements are necessary. To evaluate a system with regard to sense-making we use the theory of distributed cognition and the idea that a system is easy *to make sense* with if it is seen as intuitive, i.e., when the system is designed in a way that its conceptual structure overlaps with the users conceptual structure. The analysis of sense-making and reasoning processes can help to develop systems that are adapted to specific user concepts like the introduced theories of intelligence analysts. This work includes an analysis of the Triple Path Model of insight by Klein [Kle13], to investigate how well it reflects the work of analysts and the implications of the model for the design of VA tools.

In the course of the VALCRI project we developed a set of empirically based guidelines and recommendations for the design of visual analytics tools for criminal intelligence based on empirical research. Based on these guidelines we conducted a series of exploratory studies on sense-making strategies within this framework. The studies were designed to answer specific questions of the analyst's work. The result of these investigations supported the development of design guidelines again, and thus, informed the design of the VALCRI system. The examination of the connection path of Klein's model and our first research question, for example, identified factors that support to see connections in the crime data of CCA.

To measure insight in visualisations and VA we take the granularity of insights into account. Gotz and Zhou [GZ08] developed a multi-tier model of insights which emphasises the fact that the process of insight generation can be, on the one hand, generic, and, on the other hand, task-dependent. They argue that insight generation on a more general level is task dependent. Lower level tasks, which are more generic, can be generalised over many domains. Choosing the right task is an important factor to represent the rich exploration possibilities of interactive information visualisations [GB98].

Our research focuses on rich insights, which can be studied using domain dependent, realistic tasks and real data. North [Nor06] argues that using simple benchmark tasks can only lead to identifying low-level effects and can not be used to study insight. Therefore, we developed challenging tasks for our evaluations. For some studies we consulted police analysts during task development to ensure realism. The datasets underwent a process of depersonalisation to protect the privacy of the used data. In the course of our user investigations we carefully chose which data to collect for the analysis and only collected data necessary to the purpose of fulfilling the aim of the investigations. To conform to the code of ethics we informed the participants about the goals and asked for their consent to publish depersonalised findings.

In our investigations we evaluate the use of visualisation on an action level. We derive strategies based on exploration actions which get interpreted through thinking aloud protocols and screen captures. In the thinking aloud protocol participants are asked to *think out loud* while working with a visualisation. Fleck and Weisberg [FW04] could show that verbalisation did not overshadow the problem solving process. The advantages and disadvantages of the employed methods are discussed in detail in chapter 3.

1.5 Overall impact and associated publications

The main contribution of this thesis lies in a systematic analysis of sense-making and insight models and the empirical research in the area of visual analytics, yielding a generalisable framework of sense-making strategies. We adopted a novel approach to describe cognitive sense-making processes as, to the best of our knowledge, no existing framework can describe sense-making in visual analytics comprehensively. Our research differs from related work in its level of detail and the employed methodology, i.e., applying thinking aloud in extensive, realistic prototype evaluations. Our investigations show that the strategies are transferable from rather simple to complex systems by re-using the sense-making strategies to evaluate different kinds of systems.

In addition, we report specific design recommendations for various visualisations derived through qualitative analyses from our investigations. Taking existing literature into account, we condensed our results into compact, ready-to-use design recommendations, to facilitate their usage in practical applications. They are listed at the beginning of each study in chapter 5 and also summarised in chapter 6.

The results from the empirical work, guidelines from literature (part A) and five user investigations (part B), were published in the form of peer-reviewed journal articles and appeared in conference proceedings. The following papers build the basis for this doctoral thesis, which were partly published under the author's former name Haider, J.:

1. **Haider, J.**, Pohl, M., Pallaris, C., & Wong, W. B. L. (2015). Supporting Sense-Making and Insight Processes in Visual Analytics by Deriving Guidelines from Empirical Results. In Schulz, H.-J.; Urban, B. & von Lukas, U. (eds.), Proceedings of the International Summer School on Visual Computing 2015. (pp. 59–68). Rostock, Fraunhofer Verlag, ISBN: 978-3-8396-0960-6.

2. Pohl, M., **Haider, J.**, Pallaris C., & B.L. William Wong. (2015). Guidelines for Sense-Making in Intelligence Analysis. In *European Intelligence and Security Informatics Conference* (p. 177). `https://dx.doi.org/10.1109/EISIC.2015.45`.

3. **Haider, J.**, Pohl, M., Hillemann, E.-C., Nussbaumer, A., Attfield, S., Passmore, P., & Wong, B. L. W. (2015). Exploring the Challenges of Implementing Guidelines for the Design of Visual Analytics Systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59(1).* (pp. 259–263). `https://doi.org/10.1177/1541931215591053`.

4. Kriglstein, S., **Haider, J.**, Wallner, G., & Pohl, M. (2016). Who, Where, When and with Whom? Evaluation of Group Meeting Visualizations. In M. Jamnik, Y.

Uesaka, & S. Elzer Schwartz (Eds.), Diagrammatic Representation and Inference: 9th International Conference, Diagrams 2016, Philadelphia, PA, USA, August 7-10, 2016, Proceedings (pp. 235-249). Cham: Springer International Publishing. `https://doi.org/10.1007/978-3-319-42333-3_19`.

5. Seidler, P., **Haider, J.**, Kodagoda, N., Wong, B. L. W., Pohl, M., & Adderley, R. (2016). Design for Intelligence Analysis of Complex Systems: Evolution of Criminal Networks. In *2016 European Intelligence and Security Informatics Conference (EISIC)* (pp. 140–143). `https://dx.doi.org/10.1109/EISIC.2016.036`.

6. **Doppler Haider, J.**, Seidler, P., Kodagoda, N., Adderley, R., Wong, B. L. W., & Pohl, M. (2017). How Analysts Think: Sense-making Strategies in the Analysis of Temporal Evolution and Criminal Network Structures and Activities. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61:193–197*. `https://doi.org/10.1177/1541931213601532`.

7. Kriglstein S., Pohl M., & **Doppler Haider, J.** (2018). How Users Transform Node-Link Diagrams to Matrices and Vice Versa. In P. Chapman, G. Stapleton, A. Moktefi, S. Perez-Kriz, & F. Bellucci (Eds.), Diagrammatic Representation and Inference (pp. 526–534). Cham: Springer International Publishing. `https://doi.org/10.1007/978-3-319-91376-6`.

8. **Doppler Haider, J.**, Gastecker, B., Seidler, P., Kodagoda, N., Wong, B. L. W., & Pohl, M. (2018). Sense-making Strategies in Explorative Intelligence Analysis of Network Evolutions. In *Journal of Behaviour & Information Technology*. `https://doi.org/10.1080/0144929X.2018.1519036`.

9. **Doppler Haider, J.**, Seidler, P., Kodagoda, N., Wong, B. L. W., & Pohl, M. (2019). Sensemaking Strategies vs. Quality of Insights: Investigating Analysis Processes of Multiple Tiles. (Manuscript submitted in 2019)

10. **Doppler Haider, J.**, Pohl, M., Beecham, R., & Dykes, J. (2019). Understanding Map Comparison: Strategies for Detecting Difference in Map Line-up Tasks. (Manuscript submitted in 2019)

Papers 1, 2 and 3 are a result from both a literature review and the development of guidelines for the design of visual analytics systems in the starting phase of the VALCRI project. We adopted a systematic approach for the development of guidelines, including the steps 1) requirements analysis, 2) the analysis of related work and design recommendations, and 3) translating this knowledge into applicable guidelines using Spence's framework of design actions [Spe11]. This approach yielded 17 guidelines to support the sense-making process, concentrating on insight generation, reasoning, visualisation principles, and recommended techniques. Papers 5-10 deal with our second empirical part. We conducted five studies on sense-making processes in visual analytics with the goal of analysing cognitive strategies during realistic tasks. Aiming for a general framework of sense-making strategies we evaluated visualisations of different complexities and contexts. We investigated systems in the context of intelligence analysis, ranging from static representations in the first investigation (paper 4), to an interactive system of two visualisations, a matrix, and a node-link diagram (papers 5-7) and, finally, to a complex prototype including multiple visualisations (paper 9). In addition, we analyse sense-making strategies in the context of geographical maps (paper 10).

The results of this empirical work are design recommendations based on the results of the investigations, as well as the framework of sense-making strategies, developed in investigations 3-5. In the last two studies we could observe that some strategies were more successful than others, but further investigations are necessary to generalise these results. My contributions and responsibilities for each paper are declared and referenced separately at the end of each section in chapter 4 and chapter 5 respectively.

Additionally, my dissemination includes the following papers and peer reviewed posters, which are not incorporated in this thesis:

1. **Haider, J.**, Pohl, M., & Fröhlich, P. (2013). Defining Visual User Interface Design Recommendations for Highway Traffic Management Centres. In *Information Visualisation (IV), 2013 17th International Conference* (pp. 204–209). `https://doi.org/10.1109/IV.2013.27`.

2. Pohl, M., **Haider, J.** (2015). Guidelines for Sense-Making in Intelligence Analysis. Poster at the *European Intelligence and Security Informatics Conference*, Manchester, United Kingdom. "Best Poster Award"

3. **Haider, J.**, Pohl, M., & Wong, W. B. L. (2015). Supporting Sensemaking and Insight in Visual Analytics. Poster at the International Summer School on Visual Computing 2015, Rostock, Germany.

4. **Haider, J.**, (2016). Understanding Sensemaking Strategies in Criminal Intelligence Analysis Looking at User Interaction. Participation in the IEEE VIS Doctoral Colloquium 2016, Baltimore, United States.

5. Pohl, M., **Doppler Haider, J.** (2017). Sense-making Strategies for the Interpretation of Visualizations – Bridging the Gap between Theory and Empirical Research. In *Multimodal Technologies and Interaction 1(3)* 16. `http://dx.doi.org/10.3390/mti1030016`.

The research of this doctoral thesis follows the work on design guidelines developed in the course of my master thesis (main results published in paper 1). Early work of the project was also published in the form of two posters (paper 2 and 3). Research results, furthermore, got published within the IEEE VIS Doctoral Colloquium in 2016 (paper 5) and have been incorporated in a journal article (paper 6).

1.6 Structure of the thesis

In chapter 2 I will give an overview of the current state-of-the-art and describe existing approaches that are relevant for the present work. Then I will describe our research questions and the conceptual framework of this thesis in chapter 3 where I also discuss the methods that were used in the empirical parts of this thesis: the development of guidelines and the execution of five studies.

The first empirical part, chapter 4, focuses on guidelines for sense-making derived from existing literature and the second empirical part, chapter 5, provides details on the investigatory studies that we conducted and published in the course of this work. The main outcomes are described in chapter 6 and a critical reflection of the results and the challenging task of formulating successful guidelines are discussed in chapter 7. To conclude my thesis the last chapter, chapter 8, summarises our research on sense-making and design recommendations for visual analytics tools.

The catalogue of guidelines in the form of design actions can be found in the Appendix.

CHAPTER 2

Theoretical background

> The important thing is not to stop questioning. Curiosity has its own reason for existing.
>
> Albert Einstein [Ein55, p.64]

Sense-making can be studied on different levels, e.g., individual or collaborative sense-making in groups, which draw on different theories and models of how sense-making works. For consultants sense-making constitutes a business model as they promise insights on how to optimise their customers' processes. Early on Weick raised the awareness for a sense-making perspective by shifting the focus from how we shape an organisation with decisions to how meaning drives our organising behaviour. He points out that meaning is derived from the cognitive activity of framing experienced situations which depends on various interests of the individual. In 1995 Weick [Wei95] described seven properties of sense-making that interact and intertwine as individuals interpret events. The first property is especially relevant for individual sense-making: *Identity* – who people think they are in their context shapes the interpretation of events.

In 2005 Weick described sense-making as "the ongoing retrospective development of plausible images that rationalize what people are doing" [WSO05, p.409]. Sense-making on an organisational level is different to individual sense-making, as it means a collaborative process of developing a shared understanding out of different individuals' perspectives.

In our work we focus on individual sense-making in decision contexts, a research field heavily influenced by Russell [RSPC93], Pirolli and Card [PC95], as well as Klein, Moon & Hoffman [KMH06a, KMH06b].

2.1 Individual sense-making

How do we make sense of the world? A common theory is that we devise several strategies to answer questions about the world around us.

Most immanent is our gut feeling, which is very fast and out of our mental control. The opposite is a brain decision based on elaborate thought processes, which takes a lot of effort and therefore can be quantified via measuring physiological symptoms such as dilation of pupils or increased heart rate. In psychological research the former is called system 1 (fast thinking), while the latter is referred to as system 2 (slow thinking). The repercussions of intuition and impulsive decision making [Kah11] make us believe that slow thinking results in better decisions with less biases and a fewer error rate. However, this view is questioned by Gigerenzer [Gig07], who argues that inferences based on heuristics can lead to better results than slow and deliberative reasoning (system 2). There are situations where the gut feeling is better than elaborate thinking and we do not know why. One explanation for the gut feeling is the use of heuristics – simple rules of thumb leading to inferences that are automatic, rapid and involuntary. Johnson-Laird describes the two systems as unconscious and conscious processes [Joh96]. An advantage of unconscious reasoning is that it can process large amounts of information at once, whereas conscious processes are restricted by the working memory, which can process only a few elements at a time [Mil56].

Sense-making has been discussed widely in the information visualisation community but there is no generally accepted model of how sense-making with visualisations works. There are serious concerns about this kind of research and its generalisability due to the used tasks, e.g., Kahnemann's critics see specifically designed tasks that will generate errors in laboratory settings only [Wol01].

The identification of sense-making processes is quite a challenging issue in this context. Blandford et al. [BFA14] and Sedig et al. [SPLM16] have addressed this issue in great detail. The concept of cognitive strategies, parted into general and domain specific strategies, is discussed by Lemaire and Fabre [LF05] who argue that many reasoning processes are a combination of the two. Sedig et al. [SPLM16] point out that complex cognitive activities can be described on different levels of abstraction. In information visualisation and visual analytics such issues have been discussed within the framework of sense-making theories. In this context, Klein's approach has been especially influential [KMH06a, KMH06b, Kle13].

2.1.1 The Data-Frame Theory of Sense-making

The sense-making model described by Klein et al. [KPRP07] claims to be able to represent real-world decision making in a more appropriate way. Here, sense-making is defined as the "deliberate effort to understand events" [KPRP07, p.114]. This model is based on the concept of frames, which stand for hypothesised connections in the data which get elaborated on iteratively, questioned and re-framed (if necessary). Klein's Data-Frame Model [KPRP07] is used as a conceptual basis to conduct empirical research, such as Attfield et al.'s [AHW10] research on fraud investigation. On top of this model Attfield et al. [AHW10] developed a model of sense-making where sense-making processes are set within a context of goals, interests and values. Interactions can occur between the significant elements of the model (frame), information (data) and semantic knowledge. They discuss the distinction between naturalistic and normative sense-making and its implications for intelligence analysis, where rigid arguments need to be developed. Kodagoda et al. [KPS+17] developed a coding scheme influenced by the Data-Frame Model to classify the activities of intelligence analysts.

Klein's research in sense-making and insight evolved from the Data-Frame Model into the Triple Path Model of Insight [Kle13]. This model incorporates three processes: *Connection, Contradiction* and *Creative desperation*, which were described in the introduction of this thesis. According to Klein we have to shift an anchor in our prior knowledge to make sense of events and get new understandings. This can sometimes be a central belief that needs to be changed to alter the direction of thinking, and to get new ideas about the kinds of actions we take. Klein distinguishes in total between five different processes that enable people to gain insights: making connections, finding coincidences, emerging curiosities, spotting contradictions, and being in a state of creative desperation.

More research is needed to be able to make conclusive statements on sense-making strategies and how to improve decision making with interactive visualisations and inform system design in general.

2.1.2 Heuristics

Interacting with visualisations is often interpreted as a sense-making process in which users adopt different strategies or heuristics. The idea of heuristics has received a considerable amount of attention in recent years. Although the usefulness of heuristics is controversial, there is some agreement that they reduce the mental load during reasoning and decision-making processes and that they can lead to useful results in many cases. Newell and Simon [NS72] describe how heuristics are used to cut down the large problem space to manageable dimensions. They describe two heuristics in particular – hill-climbing and means-end analysis. Hill-climbing deals with local search where small changes are made to see if they improve the solution or not. With these little steps one climbs to the top of a set of possible solutions – leading to the best solution. Means-end analysis is a problem solving technique that envisions the end, i.e., the ultimate goal, to find a strategy that will lead there in a sensible way. Gigerenzer [Gig07] argues that reasoning processes in everyday situations are often based on a specific heuristic – gut feeling. Based on empirical research he could show that this heuristic can be very efficient.

2.1.3 Storytelling

The activity of telling a story is characterised by occasional improvisation, theatrics, or embellishment. These elements can have a great effect in entertaining others but also play a role in memorisation techniques, since a coherent story is easier to recall than a series of random events or arbitrary numbers – a short storyline is faster to learn than a telephone number. It has been argued that the development of a storyline supports users to form connections between disparate facts to make them memorable. Similar issues have been discussed in cognitive psychology (see, e.g., [ZMG95]).

In cognitive psychology, the emphasis is on the activities of the study participants and how they make sense of the material that is being presented to them. Based on this material participants construct coherent models. Using storytelling as a way to design visualisations has been discussed in the visualisation community [KM13, LRIC15].

2.1.4 Distributed cognition

The theory of distributed cognition in contrast to most psychological theories sees knowledge embodied in artefacts and humans instead of only in the human brain alone [HHK00]. In this view cognitive processes which are distributed among users and computers are carried out with the help of artefacts. Knowledge about computing, for example, is embodied in a computer. People using computer systems are often times not competent in the knowledge that is embodied in the system. Nevertheless, they can use this knowledge as it is accessible through the system.

Humans excel at providing contextual meaning and outperform computers in reasoning tasks where background information plays a crucial role [KKEM10]. In contrast, computers are better at formal reasoning processes and can process large amounts of information. Therefore, visual analytics works best if system design and tasks support the strength of the user. The computer can take over parts of the information processing and, therewith, reduce the cognitive load of the user.

Cognitive load theory suggests that we can use external structures to change the way we think. By reducing the need to keep information in memory we can improve thinking and reasoning [Swe94]. Clark describes *scaffolded minds* as a decentralised problem solving technique which describes human sense-making taking the help of external aids into account [Cla98].

In the sense of this theory, the analysis of user interactions with the system gives access to the sense-making processes that are going on in one's mind. It further indicates that a mental model of how things work emerges while using technology [OD92, PSM12]. A mental model provides some kind of representation of the world, blending visual characteristics with conceptual and physical features. The abstraction of a mental model can be used as a template for understanding [Pyl03, Rap05].

There are several ideas of how we construct and work with mental models. A visualisation is, on the one hand, a representation of a mental model (the one from the designer of the visualisation), and, on the other hand, a means of externalisation or communication (looking at a visual representation might involve insights when we get a new understanding). Early research based on the cognitive load theory suggests that can hold seven plus or minus two elements at a time in memory [Mil56]. The human mind, however, finds ways to be capable of more. Psychological researchers began to describe the use of *chunks* first when studying the game chess, to explain how chess master are able to remember the location of 20 or even more pieces on the game board [CS73].

The construction of mental models for complex visual representations enables the filtering and abstraction of the information of interest and it is suggested that visualisations should only change in a way that allows to form a coherent mental model. When animation is used for transformations, for example, the transformation should happen at the speed of perception. Furthermore, techniques to preserve the mental map in animations are suggested [DGK01, AP16].

2.1.5 Gestalt psychology

> The whole is something else than the sum of its parts.
>
> Kurt Koffka [Kof35, p.176]

Gestalt psychology is well known for its research in the area of perception. Gestalt theory tries to explain how we organise individual elements into groups in the way we perceive and recognise patterns and is used for the design of visual displays. The idea that the combined perception can mean something else than the single parts plays an important role for insight problem solving as well. The representational aspect of problem solving is emphasised in this theory. When the structure (Gestalt) of a problem is made clear, a solution can be found easily. Problem solvers redefine the representation of the problem to find a solution. Insights are a result of creative thinking going beyond the given information. Gestalt psychologists are interested in short "Aha" insights, which involve the activation of less clearly relevant information corresponding to a paradigm shift (or reframing) with respect to a problem. However, researchers from the field of Gestalt psychology argued that "Aha" moments often come from a long-term iterative reasoning process [DD95]. It is still controversial whether gaining an insight is a specific process compared to other problem solving activities or not, as there is no commonly agreed definition of insight [FW04, PSCO09].

In problem solving psychology, there is a renewed interest in the concept of insight. According to Novick and Bassok [NB05] this is related to a trend toward examining more complex, ill-structured problems. Everyday reasoning is often confronted with ill-structured problems and trade-off decision making and, therefore, Gestalt psychology can be considered highly relevant for sense-making research.

2.2 Guidelines for visualisation

Visualisation is key for sense-making as humans quickly interpret visual information with seemingly little effort. In recent years human centred visualisation and the perspective on interaction were increasingly discussed in the information visualisation community, e.g., by Kerren et al. [KEM07] and Ward [WGK10]. Guidelines for the design of graphical displays exist for various contexts, such as guiding a process, suggesting general and specific principles, communicating good solutions or capturing previous experiences as best practises [Nes05]. Interaction is seen as the solution to working with big data, i.e., vast amounts of data that need preprocessing prior to visualisation. The information needs to be aggregated to be visualised at once, or filtered to display parts in detail. Interaction patterns have been studied to support different kinds of tasks, such as comparison tasks [Tom15], because interaction is seen as the key to manage cognitive overload. Interactive visualisation is a challenging research field as sense-making processes are still not understood very well and how analysts derive insights is current state-of-the-art research.

Graphical displays use space to organise information and to facilitate memory by externalising information and simultaneously using the power of spatial inference making for other domains [TMB02]. The reuse of pre-designed solutions is an approach to utilise and apply design knowledge. However, it is often difficult to transfer such a solution to another domain with different requirements. Thus, related taxonomies of cognitive work systems are developed instead and appropriate design methodologies encourage the utilisation of existing design knowledge across different work domains [PDE10].

Sutcliffe [Sut00] describes design advice for specific scenarios (claims) as an alternative to the representation of HCI knowledge, which can be generalised for future reuse. Appropriate recommendations from theory can be drawn from bridging representations that build on models of interaction or claims with a task-artefact approach. However, when applying cognitive theories directly to visualisation design we face scalability problems. Another problem is that cognitive insights acquired by psychologists are too generic or difficult to understand for the direct application by interaction designers. The knowledge has to be presented in a way to improve design guidance without the need for the extensive background knowledge [Spe11].

2.2.1 The visualisation - interaction gap

Not only is there a need to transfer knowledge from psychological research to HCI, but there is even a communication gap within the visualisation community, between the developer (interaction designers) and visualisation designers [Tom15]. Designers with an HCI background tend to focus more on the human, whereas developers traditionally focus on the computer and the interaction interface. Only by bringing interaction and visualisation research together a good experience with visualisation tools can be achieved. The goal is to develop systems that integrate effective visual representations and intuitive, enjoyable interfaces in order to generate insight.

Challenges in this area become more and more relevant with the design of novel systems and innovative controls, such as advances in augmented reality or robotics. Already widely used interaction techniques are joysticks, motion sensors or trackpads, e.g., in the domain of gaming and mobile visualisation, e.g., the interaction with events or Point of Interests (POIs) in maps on mobile devices [SSS10]. New developments should be based on empirical results; basic principles are especially important because of the fast technological progress which widens the gap between research and development. In the following I will discuss some of the basic principles which apply to the design of (interactive) visual analytics tools.

A considerable amount of guidelines is proposed in the literature. The human visual system, for example, is well investigated having established guidelines for perceptually salient graphics [EBB05, CN06, HE12] which are supported by empirical research [FNPS99a, SC99, NSH02, CFL10]. A general short list for interface design, e.g., comprises the eight golden rules by Shneiderman [Shn10]:

1. Strive for consistency.
2. Cater to universal usability.
3. Offer informative feedback.
4. Design dialogues to yield closure.
5. Prevent errors.
6. Permit easy reversal of actions.
7. Support internal locus of control.
8. Reduce short-term memory load.

This short list is an example for guidelines on a very general level with a big margin in the application. These guidelines do not tell designers how to apply them in a specific implementation. We can find lengthy sets of such advices for different design aspects in the literature, such as the organisation of displays, navigating the system, getting the user's attention and how to facilitate user entries. They need to be treated with caution, however, as not nearly all of them are based on empirical evidence. In the second best case they are based on theory, but many are derived from experience or personal preference instead. Efforts to retrieve knowledge from experts through workshop discussions on their domain can yield a valuable collection of advice, such as when and how to use multiple views [WBWK00] but guidelines for specific trade-off situations need further empirical evaluation.

2.2.2 Minimise cognitive load

> Contrary to common expectation, static displays are generally more effective in terms of both time and accuracy.
>
> Farrugia and Quigley [FQ11, p.79]

One finding from literature is that visualisation should prefer schematic graphics that are reduced to the essential information instead of rich realistic ones [TMB02, Bar97].

Cognitive overload occurs easily in static representations but has even more dramatical effects with dynamic visualisations, such as change blindness. Animated data is an intriguing technique to show change over time because a lot of information can be conveyed quickly but the problem is that not all the information will be perceived at the first sight. Animations often times are just too fast to be accurately perceived and overwhelm our mind as they present increased amounts of information in larger time-series [AMST11]. Actively paying attention to the animated data is no guarantee to not miss part of the information, because of disappearance and split attention [ON11, HF08]. Disappearance occurs when the user misses information because of time, e.g., when the user does not pay attention from the beginning or starts watching at a later point of the animation's time span. Split attention occurs when the user focuses on a specific part of the screen, while other things happen in the other areas. The change-blindness effect is a robust effect, i.e., we miss big changes even if we are aware of the problem and despite our efforts to see them [NHT01].

Several studies compared animation to alternative, static representations and showed that despite its enjoyable nature it is not as effective for analysis as other techniques. Now the quasi-consensus in the visualisation community is that animation is more prone to participant errors and is therefore not the recommended choice for trend analysis [AP16]. Kriglstein et al. [KPS12] systematically reviewed the literature on animation with the result that there are few evaluations of animation for time-oriented data. Hence, the first question should be if animation is the only option and, secondly, how it can be designed to make it work. Some studies, on the other hand, suggest that in certain situations visualisations can benefit from animation, e.g., animated transitions in evolving graphs that support maintaining the mental model or real time reorientations in time and space [RFF+08, TMB02]. When animation is used for complex data it is generally agreed upon that interaction possibilities need to be provided. The general belief is that with the options to halt, jump back and watch a sequence again, it is possible to support the *Apprehension Principle* of good graphics, according to which graphics should be accurately perceived and appropriately conceived [TMB02].

A comparison of animation and small multiples with temporal graphs has shown that small multiples produce less cognitive load [FHQ11]. Small multiples are limited by the size of the screen, because the essence of the idea is to see all the data at once. This means a sharp constraint on the level of detail that can be shown [RTJ+11]. The analysis of eye gazes using small multiples suggests that computationally equivalent animations can only be provided when backwards animation and the option to reverse sequencing is available [AA08].

Another technique that comprises all the data in an integrated summary view, is to use different encodings in one layout. Interaction is the key to counteract the limitations of these techniques. The goal is to enable analysts to make new connections and show up missing data at the same time.

Empirical results can provide valuable insights on the influence of graphical features. Rodgers et al. [RSC15], for example, extensively studied linear diagrams and derived design principles, which were useful for the design of the visualisations we evaluated in our first investigative study. I figured that it is hard to find empirical data for specific use cases when we were working on guidelines relevant for monitoring floating car data. In the course of a project about traffic management systems we gathered recommendations for static and dynamic data, the representation of change and interaction [HPF13].

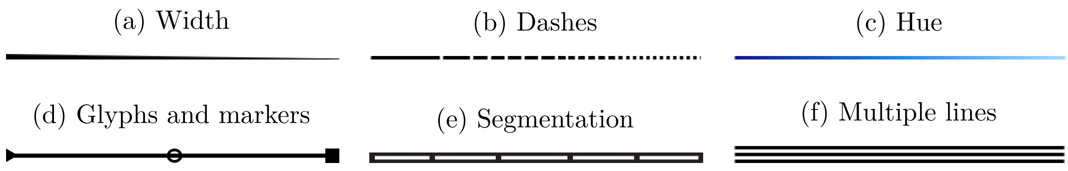


Figure 2.1: Continuous (a-c) and discrete (d-f) line features.

Chang and Nesbitt [CN06] suggest that structuring existing guidelines on the basis of Gestalt-based principles can be useful to support the development of new guidelines. In the domain of criminal intelligence analysis the visualisation of event data and social networks are relevant. In the following I will discuss the related work of graph design and how time can be incorporated into graph visualisations.

2.2.3 Graph design

Graphs are traditionally represented in node-link diagrams and sometimes literature on graph visualisation implies that kind of representation. Entities are represented as nodes and the relationships between them as links (lines). Lines offer a limited number of encodings; properties that span over the entire line length are: a) width, b) style, and c) colour. Additional features can be used, such as: d) markers or glyphs, e) segments and f) multiple parallel lines, compare Figure 2.1.

Nevertheless, graphs can also be represented in a strictly structured way – matrices – which in general, as studies have shown, are more efficient than node-link diagrams [GFC04, ABHR$^+$13]. The nodes are represented in rows and columns and their relationship is depicted in the respective cell. Node-link diagrams and matrices have been compared several times to identify their advantages and disadvantages. Node-link diagrams are well-suited for path related tasks [GFC04, KEC06] and matrices are more efficient for large dense graphs. Drawing abstract versions of such diagrams make up an own discipline – graph drawing – which is concerned with rules for graph layouts that produce representations that are clear to understand.

A common rule to reduce the cognitive cost in understanding graphs are, for example, to minimise edge crossings and bends and maximise symmetry [Pur97, WPCM02]. More recent research, however, suggests that edges are not as important for graph comprehension [BPWv17].

The readability of two graph representations, a matrix and a node-link diagrams, was evaluated by Ghoniem, Fekete and Philippe [GFC04] using seven generic tasks which yielded recommendations for graph representation depending on graph size and density. They report that matrices are especially efficient for larger, denser networks.

To leverage the strength of both representations some visualisations combine them, such as the MatrixExplorer [HF06] and Nodetrix [HFM07]. Both juxtapose matrices with node-link diagrams. MatrixExplorer shows the whole data in both styles and Nodetrix uses matrices as nodes for dense sub-networks which are linked, representing the sparse connections in the graph network. Henry and Fekete [HF07] developed another hybrid tool, MatLink, that shows curved links for the connected nodes in lines and columns to support path-finding tasks in the matrix. They could show that their combination of both tools is superior to using either node-link diagram or matrix alone.

More specific research on the design of matrices was done by Liu and Shen [LS15]. They show that square matrices lead to improved pattern recognition in comparison to triangular matrices and propose a compact juxtaposition of many matrices, as well as three design recommendations for matrices:

1. Triangular representation does not hamper graphical perception of adjacency matrices.

2. Symmetric rather than translational juxtaposition should be preferred for detecting changes of structures and patterns.

3. Complementary juxtaposition is beneficial for optimising utilisation of display space.

Graphs incorporating temporal developments – dynamic graphs

Research on evolutionary graph visualisation is state-of-the-art in the Visualisation community. Network visualisation faces the challenge of scalability and a major factor to provide a successful visual analysis tool is the interaction design [PAKC15]. Most commonly, a series of diagrams gets animated or is shown next to each other as small multiples [AAK+14, BBDW14]. To be able to follow and understand changes over time, mental representations need to be preserved, which is especially relevant in animation as it depends on the ability to memorise [AAK+14]. This is problematic due to a high cognitive demand on the working memory and perceptual effects like change blindness [NHT01]. The benefit of animation is not conclusive for real world visualisations that depict many changes [AA11, AAK+14, AP16, BPF14a].

Other methods to visualise network evolution are either based on small multiples or on summary views, incorporating a time-to-space mapping within a timeline [BHD+15, BBDW14]. Using the third dimension Matrix Cubes, stack adjacency matrices in a space-time cube [BPF14b]. However, the general problem with Three-dimensional (3D) layouts is that they do not only suffer from occlusion but also from perspective distortion and are, therefore, difficult to comprehend [CS98, KDA+09, RTJ+11].

Unrolling the dynamic network in space is the alternative. Views that attempt to incorporate all data points in time usually use a timeline on the x-axis or z-axis in 3D space (Space-Time cubes[AAD+10]). Another type of visualisation are summary views that integrate or superimpose the temporal information, e.g., by encoding recency by transparency (older fading out) or layer ordering (newer on top of older ones).

New approaches and novel designs get developed and evaluated as well. The analysis of trends in NetEvViz [KNC+11] offers a timeslider to select two time points in a network to show their differences in a node-link representation.

Other research proposes to use multiple links between nodes. Node-link diagrams with multivariate edges [KAW+14] were presented as multiple threads depicted through parallel, multi-coloured lines. Although there are many application areas for multiple links, the problem is that sometimes they can be difficult to distinguish and that an appropriate design is needed. Beck et al. [BBDW14] also discuss multiple links for dynamic graphs.

Some alternative solutions to node-link diagrams for dynamic graphs have been evaluated as well. Rufiange and Melancon's [RM14] taxonomy of dynamic networks includes glyphs that show small summaries of the evolution of edges. They propose extended glyphs for the visualisation of multivariate edges, e.g., by using animation or stacked bar charts, and introduce AniMatrix, an animated matrix-based software evolution representation. Other techniques are timelines in edges for ego-networks [Rei10] and parallel edge-splatting [BVB+11]. The first technique focuses on the representations of single actors, but has the constraint of reading directions for the temporal interpretation and, therefore, a sharp layout limitation. The latter visualises communities in an aggregated overview and consequently impedes the view on individual nodes and relationships.

2.3 Evaluation of sense-making and insight

The research community of information visualisation is well established and it is a common practice to assess new solutions with user studies. The call for empirical studies remains open [KPSL08, CFL10, KP15].

In recent years an increasing amount of research addresses sense-making and insight [Nor06, Won14, SPLM16, GF09, PC05, Kle13, CZGR09]. Nevertheless, there are still many open questions. One open question concerns the identification of insights and how insights can be measured. It is extremely difficult to solve this problem on a general level, but it is probably manageable for defined application areas where the definition of insights follows from the task description.

Recent research in sense-making focuses on everyday insight experience that also includes examples of negative insights for the first time [HK18]. This kind of research addresses insight generation beyond cognitive problem solving including elements related to applied psychology. Pohl et al. [PWK16] combined interaction log analysis with thinking aloud to understand how analysts interact the system. They studied patterns from interaction sequences on the basis of Yi's taxonomy [YaS07]. The analysis of interaction logs alone is difficult as it requires an interpretation of the interaction sequences to connect user actions to intentions and effectively gained insights. The insight-based method [SND05] is an alternative to the traditional task-based method to evaluate visualisations by different characteristics of insights and will be discussed later in more detail.

Connecting neural processes to insight moments is current state-of-the-art research. The idea of stimulating certain regions of the brain to promote *insightful thinking* is intriguing. To investigate insights and problem solving an experimental setup was developed. The Remote Associates Test asks participants to find an associated noun when being presented a sequence of three words. The used datasets are constructed so that the search for the solution is hard because it is not closely associated, which would be obvious answers. The mental work is to break through this habitual mode of thinking and find a remote associate. From the sequence *house–bark–tree*, for example, the remote associate word would be *tree*. This kind of test is suitable for controlled experiments because once the solution is found it can be reported immediately. Tik et al. [TSL+18] describe neural processes of two types of insights through an fMRI study. On the one hand, they observed at the Aha!-moment as an emotional short moment, and, on the other hand, they found low insight events during analytical solutions. They showed up in the same areas of the brain but differ in the signal strength – the Aha! moment having a bigger impact.

Laboratory studies, however, are not suited for the analysis of naturalistic sense-making. Nevertheless, results from such controlled experiments are interesting and can spark up new directions for other kinds of sense-making research.

2.3.1 Choosing the right method

Formative usability evaluation of visualisations aims to identify and solve problems in the tool use. Typically used qualitative methods that go beyond time and error are interviews, observation, field studies, and diaries. Interviews on sense-making processes are unlikely to yield new results due to the obscure nature of insight and because humans interpret their reasoning while they think of *how they think* and likely idealise their thinking steps by emphasising rationality and logical thinking. The observation of participants as they perform given tasks using a thinking aloud protocol deems better as cognitive processes can be observed in a direct way.

These qualitative methods are practical as they are cheap, easily accessible and transferable to natural working environments compared to technical equipment that only works in laborious settings, e.g., eye-tracking devices or Magnetic resonance imaging (MRI) scanners. The main drawback is that the analysis process is very time-consuming and the study design as well as the data collection require careful planning as, e.g., asking the "wrong" question can lead to unsatisfactory results. Qualitative methods are used in the social science because quantifiable measures, such as log data, typically open up more questions than they can answer. Looking at response times with regard to sense-making, for example, does not reveal much about the thinking and it is not possible to say why the tool was used in certain ways. Therefore, sense-making and insight generation are often studied using thinking aloud protocols and, sometimes adapted, methods for the qualitative content analysis, such as Integrative or Emerging Themes Analysis, Cognitive Task Analysis or Critical Decision method [May03, HCS98, FW04, HK18, KCM89, WB02]. In the analysis of the verbal reports it is necessary to development an appropriate coding scheme to get systematic results from thinking aloud protocols. This can either be done in a top-down approach using literature or by starting with the content in a bottom-up manner through a repeated analysis of the protocols [Sn15].

The thinking aloud protocol asks a great deal from the participants because they should speak out loud what they think during fulfilling a task in an experiment [ES93].

Thinking aloud gives additional insights on the tool use but the method induces an unnatural feeling for the participant and reduces the realism of the study. With regard to sense-making research the method is criticised because it increases the cognitive load for the participant during problem solving in the tasks. With thinking aloud protocols alone it might not be possible to understand the user behaviour in complex tasks, when the cognitive load is too high for the user to describe his or her inference making [CFL10]. The advantage of hearing at least some of the participants' thoughts and plans, however, often out-weighs these disadvantages.

Eye-tracking is often used to study insight problem solving. This is fitting because the tasks in such studies have clear answers and the components of the problem are well defined. Eye movements alone, however, do not explain thinking steps of the users. Results from the eye-tracking (scanpaths and fixations) need to be interpreted, which is not a straightforward process and the tasks need to be designed carefully to support the interpretation as well as to avoid, for example, precision problems. To get a more diverse picture of the user's behaviour Opach and Nossum [ON11], for example, suggest to combine the eye-tracking technique with other evaluation methods. They conclude from their experiments with animated maps that eye-tracking is valuable for evaluating the user's attention in a way that is not easily accessible with other methods. One might assume that thinking aloud can help to interpret eye movements, but the combination of thinking aloud and eye-tracking is rather difficult. The concurrent recording of eye-gazes and verbal protocols yield unsatisfactory results because participants should not explain actions and manipulation of the software and verbalise cognitive processes instead. The process of verbalisation might also influence fixations, when the thinking aloud protocol is used in a conversational manner. Thus, verbal protocols can not be linked directly with eye movements.

Retrospection with the help of recordings of the users' eye-movements proved to yield better results than cueing without eye-movements because more comments and explanations of the analytical process can be retrieved. The problem with retrospection in general is that it is not suited for long exploratory studies because participants easily forget their thinking steps, even when they see their own eye movements. As a rule of thumb short sessions of about ten minutes should be cued. For the overall duration of the session this procedure means a doubled amount of time is needed, or rather half of the tasks can be performed in one session. This needs to be considered if timing is an issue. On the other hand, an altered performance of the main task because of the thinking out loud task is impossible [HNA+11]. Ericsson and Simon, however, argue that concurrent thinking aloud may, if anything, slow down performance but does not alter it [ES93]. Even more problematic for studying sense-making, however, is that retrospective cues are an interpretation of the reflecting participant whereas the goal with concurrent thinking aloud is to observe immediate thoughts.

Gegenfurtner and Seppänen study expert performance in a thinking aloud and eye-tracking study and address another important issue for evaluations – choosing the right sample. By using two different user groups in their sample they could show that results depend on the participants' level of expertise and reveal specific differences in the area of interest. Experts focused on task-relevant areas in familiar tasks and were able to transfer these underlying processes to semi-familiar tasks, but not to tasks from an unfamiliar domain [GS13]. The importance of the user sample and tasks is discussed in the next section.

The insight-based method

The insight-based method follows an unguided data analysis protocol and thus, enables an open-ended analysis of insight generation by not restricting the users analysis to a set of preplanned benchmark tasks. Saraiya et al. [SND05] point out that a higher level analysis of reported insights which are domain dependent is needed, such as grouping them into different categories. They quantified the insights gained from the exploratory use of visualisations in the field of bioinformatics and grouped them into four categories: overview (overall distributions of gene expression), patterns (identification or comparison across data attributes), groups (identification or comparison of groups of genes), and details (focused information about specific genes). They also distinguish between unexpected insights, like the Aha-moment described by Klein, and directed insights, that answer specific questions of the user.

Due to the unguided protocol, it is possible that two participants may analyse the data in different ways and report different insights. With a sufficiently large number of participants this can cover several different types of possible insights for the visualisation tools. Thus, investigations using this kind of method may often provide feedback that the evaluators may not have thought about earlier.

It is possible that two different evaluators may analyse the reported insights differently resulting in different conclusions about the visualisation tools. Therein lies the strength and the weakness of the insight-based method. The subjective group makes the insight method less uniform, however different insights derived by different participants suggest that different kinds of insights are possible and they potentially depend on how or which tools are used. Also, this method can help to determine if participants' with different backgrounds may use the visualisation tool in different ways.

2.3.2 Users and tasks

An empirical evaluation of visualisations should ideally be conducted with potential users to reveal problems and show up possible improvements in the quality of the visual representation. In situ observations are field studies where activities can be observed by the experimenter as they take place *in situ* and simultaneously are not altered through the observation. An experimenter, however, only remains truly unobtrusive if not present, i.e., using a hidden camera. In an actual setting this is impossible most of the time and the trade-off for designing a user study is to make sure that the observation feels natural and that the observer fades into the background. Ideally participants will forget about the observer after some time when they engage in the task itself. To reduce the "eye over the shoulder" effect certain measures can be taken, such as to position the observer in some distance or video record the activities of interest, when a camera is suitable. When eye-tracking is used a high realism might be reached, but data collection and analysis is expensive and there is no guarantee for a clear interpretation of the data. Participatory observation, laboratory observational studies and contextual interviews are good alternatives to study user activities [Car08]. The difficulty of evaluating information visualisations lies in the generalisation, precision and realism of the results, which cannot be addressed equally in one study with current methodologies. This is why the choice of the research method needs to be appropriate for a specific context and a particular goal. Also choosing the right sample for a user study is important to produce valid results [Car08, Maz09, KP15].

Describing the user sample is especially important to support reproducibility. The renewed interest in reproducible methods was made very evident in the broadening of the scope of the BELIV workshop in 2018 and the complementary focus topic: *replication practices in visualisation* that gives the rare opportunity for reflective assessment even of unsuccessful replication efforts. The workshop also changed its name from "Beyond Time And Errors: Novel Evaluation Methods For Visualisation" to "evaluation and BEyond – methodoLogIcal approaches for Visualisation".

The expertise of participants plays a role depending on the tasks that get studied. However, experts are not always easy to reach. They can be difficult to recruit as participants for controlled user studies due to various reasons. Sometimes they are just a very rare resource that is hard to access, and if available, they are expensive and often better exploited through other methods. Analytic evaluations include reviews by experts, such as with heuristics [Nie92]. Inspection methods are inexpensive evaluation approaches and use a set of heuristics as a method of focusing attention in important aspects of the visualisation which need to be considered. Several experts develop the heuristics and the inspection is done from an individual expert. While heuristics may represent a valuable tool for improving the quality of information visualisations the application process and taxonomies are yet to be defined.

2.4 Limiting the scope

Research in this area necessarily includes perceptual issues. The bandwidth of perception, however, is broad. Decades of research concern the physiology of the human visual system, the theory of optics, colour and the saliency of visual properties. Ware summarises perceptual fundamentals in his information visualisation book "Information Visualization: Perception for Design" [War04] . We do not address attention or biases with our studies, as each would constitute a research field on its own.

We focus on sense-making processes during visual analytics, i.e., analytically solving tasks with the help of interactive visualisations. The domain of intelligence analysis imposes certain requirements and at the same time delimits the field of application and visualisations of interest. Police analysts need tools which are intuitive and ready to be used in practice. Visualisations, on the one hand, need to be flexible for open exploration and, on the other hand, need to communicate findings unambiguously to be able to act as rigorous evidence.

These requirements set the boundaries on technically feasible types of visualisations, which is reflected in our studies. We evaluate basic, tried and tested visualisations, and adapted them as needed. Thus, we did not consider techniques that are famously difficult to comprehend, such as animation or 3D [CS98].

CHAPTER 3

Methodology

To study sense-making and insight generation processes with visual analytics we studied the interaction with tools using the distributed cognition approach [HHK00]. The basic assumption is that interaction processes are an indication of the sense-making processes users engage in while they work with the system. It implies a tight relationship between the users' activity and their thought processes. Hence, interaction processes help to understand problems in the tool use and to identify design recommendations, but also reveal general sense-making strategies. Distributed cognition has been suggested as an appropriate basis for modelling reasoning processes of analysts using visualisations [PSM12]. The *Triple Path Model* of Klein [Kle13] fits in this concept because the activities mentioned in the model are typically supported by interactive visualisations.

In the next chapter I will describe the results of my research, the guidelines and the empirical work. Here I will describe my approach, how I addressed the research questions and which methods I used.

3.1 Approach

In the initial phase of this work I covered the topics *sense-making, insight, visual analytics, visualisation, intelligence analysis, decision-making* and *problem solving* in an extensive literature review.

This was the basis to formulate the research questions:

1. How can the design of visual analytics tools support cognitive processes for intelligence analysis?

2. What kinds of sense-making processes get employed in visual analytics?

3. What kinds of interaction processes should be supported by the system (e.g., scroll, zoom, pan, filter, reconfiguring elements on the screen)?

4. How useful are models of insight generation? Does the representation of the problem influence insights [HS74, ZK10]? Does using different representations lead to different kinds of insights and which visualisations are especially useful in the insight generation process?

We attempt to answer these questions with empirical research and conducted several user studies with end users and students as participants. The goal of the studies with end users as participants is to clarify the more task specific aspects of the sense-making process. Additional goals are to investigate which kinds of visualisations are advantageous for the tasks of end users or which cognitive processes in the sense of the Triple Path Model of Klein [Kle13] are most valuable to solve problems occurring in intelligence analysis. The selection of methods will take into account the constraints of the research with the end users (e.g., studies will not exceed certain time limits, experimentation in a narrow sense will not be possible).

We further conduct more general research on sense-making issues with students as participants. This allows to recruit a larger participant sample in order to conduct quantifiable research in a mixed methods approach. In these studies, the specifics of the tasks of intelligence analysts will not play a major role. The emphasis will be on the investigation of interaction processes with visual analytic tools for sense-making and their design and also the analysis of the Triple Path Model of insight developed by Klein [Kle13]. We want to investigate how well this model reflects the work of analysts and what the model implies for the design of visual analytic systems. We believe that basic sense-making strategies are independent of domain knowledge in the same manner as the fast thinking (System 1) can not be controlled, i.e., trained or altered through experience. We, therefore, think it is feasible to conduct studies with non experts as participants, which is an important factor for successfully recruiting participants in large quantities. Experts in the field of criminal intelligence analysis are a very limited resource and are, therefore, not accessible at this rate. Hence, we argue that the best approach is to combine the resources in that way. Finally, we want to stress that the results of our studies are necessarily tentative because of the exploratory nature of the investigations, but they can act as a stepping stone for future work.

3.1.1 RQ 1: How can the design of visual analytics tools support cognitive processes for intelligence analysis

To address this research question we first had to get a comprehensive picture of the intelligence analysis process and the workflows in law enforcement agencies. The characteristics of this kind of analysis were investigated within the requirement analysis of the R&D project VALCRI. The requirement analysis started with observations and interviews with intelligence analysts from the law enforcement agencies. It was essential to get to know current policies in Belgium and the United Kingdom and how these analysts work and how their tasks look like. Furthermore, we had access to the transcripts of interviews with experts, conducted by project partners, to get detailed requirements of the end users who are experienced in the field of CCA. With this knowledge we could plan studies to answer the sub research questions, i.e., how to design visualisations to support the individual sense-making strategies and, if the emphasis should be on the perception of detail or a comprehensive model, respectively.

The *Triple Path Model of Insight* by Klein [Kle13] introduced in chapter 1 was the main inspiration for the analysis of sense-making strategies. We see a challenge in Klein's theory of how insights are generated in the context of criminal intelligence. He points out that people with the same information level might still not be able to gain the same insight, as insight is up to the individual to identify which dots need to be connected. So the pathways are no guarantee for insights and the main question of why one person gets an insight while another does not remains an open question.

In study 1 and 2 we evaluate basic visualisations and the transformation of one visualisation to another to better understand cognitive processes of the respective visualisations. The overall goal of the remaining studies is the analysis of sense-making processes on a general level. The assumption is that such investigations can assist the formulation of guidelines for the design of visual analytics tools for criminal intelligence (being thus specifically useful for the VALCRI project), but also apply to a broader range of the visual analytics field. We use the model of *How Analysts Think* by Wong [Won14], which points out that inference making is not uni-directional and that uncertainty keeps playing a role within this process. It better suits naturalistic decision making because it represents potential spontaneous jumps in the thinking and reasoning strategies compared to Pirolli and Card's sense-making model [PC05], which suggests a rather linear model of data foraging and synthesis to arrive at an insight.

We investigated visualisations that are currently used in the context of intelligence analysis, e.g., maps, node-link diagrams and timelines. For the exploratory studies we used datasets from the domain of visual analytics and tasks related to criminal investigations, e.g., the Global Positioning System (GPS) data from the Visual Analytics Science and Technology (VAST) Challenge 2014 [CGW14], see Figure 3.1. The visualisations for the study materials were developed using Tableau Desktop [TAB14], a tool that focuses on business intelligence and interactive data visualisations, and R, a software environment for statistical computing and graphics [R C13]. For the data analysis we used R.

3.1.2 RQ2: What kinds of sense-making strategies are adopted in visual analytics?

It is increasingly difficult to derive valid inferences from the growing amount of information that is available [TC05]. The goal of our user studies is to get a better understanding of sense-making processes while users work with visualisations, such as maps, timelines, or networks (compare Figure 3.2) and how analysts derive knowledge from information systems.

Visualisations are often evaluated with regard to effectiveness through rigorously controlled experiments, where aspects of the tools, tasks, data, and participant are controlled (making up the independent variable) and accuracy and performance are typically measured (the dependent variable) [Pur12]. Measures for accuracy include the

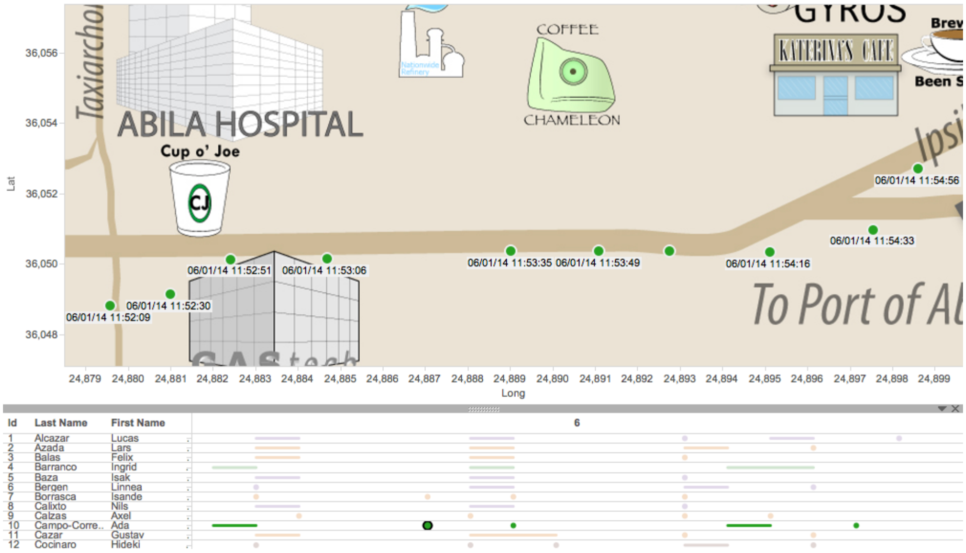


Figure 3.1: The interactive analysis of movement data in Tableau showing timestamps of the GPS dataset from the VAST Challenge 2014 [CGW14] overlaid on a schematic map.

number of correct and incorrect responses, as well as precision or error rates. Regarding performance the time to complete predefined benchmark tasks is measured. Such controlled experiments have a limited range of applicability. When comparing two or more tools in controlled experiments, for example, tasks need to fit the measure requirements and the performance assumption needs to be met, i.e., faster responses are better. Both of these aspects are problematic for wicked problems, that have no clear answer and ending point per se, thus, it is not clear when a solution is reached and further analysis might be preferred over fast responses.

Therefore, we conducted thinking aloud investigations to study how visual analytics systems are used. We investigate if it is more helpful to adopt visualisations that are already in use in the context of intelligence analysis, such as network visualisations, timelines and geographical visualisations or to develop new visualisations that specifically address the needs of intelligence analysts. For study 3 we developed two visualisations

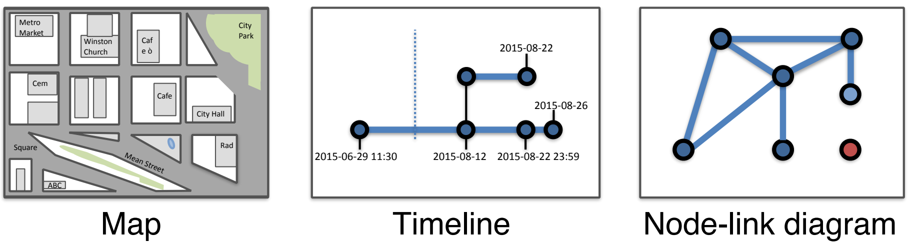


Figure 3.2: Examples of common visualisations.

on the basis of traditional techniques, the matrix representation and the node-link diagram. The prototype in study 4 comprises traditional as well as specifically designed visualisations for the intelligence domain. Study 5 also compares a traditional with a newly developed visualisation.

3.1.3 RQ3: What kinds of interaction processes should be supported by the system?

Nowadays a variety of innovative controls for the interface are available, like, for example, touch screens, 3D gesture, speech or gaze recognition. Yet basic techniques are not fully researched yet, although their study can be very rewarding, as it could be shown, for example, that scrolling and tooltip usage depends on the user group [PWM+12]. Traditional interaction techniques for visualisations are, e.g., scroll, zoom, or pan for navigation, select and filter to reduce the search space, and sort and reconfigure to structure elements on the screen.

Studying interaction processes is a laborious work. Log data can lead to insights about interaction patterns [FW04, PSM12, PWK16] but it has to be cast into meaningful actions first and is not well suited for exploratory studies and prototypes like in our case. We, therefore, take a qualitative approach through the analysis of screen captures and thinking aloud protocols and look at basic interaction techniques in study 3, such as hover and zoom as well as the switching behaviour between two visualisations. In this investigation we compare two visualisations which both offer a reduced set of interaction possibilities: hovering to show details, zooming and scrolling/panning, which was only necessary in a zoomed state. The system was kept simple on purpose to reduce confounding factors when studying how users work with a visualisation and which visualisation they favour for which kind of task. As a result we can show that one of the visualisations has a much higher demand for interaction than the other and that hovering can be supported by a visual highlighting.

In study 4 we address interaction patterns on the basis of tool use as we investigate the use of ten visualisations within the VALCRI prototype. We also describe the sense-making strategies during tool used and, furthermore, assess the quality of self-reported insights using the different visualisations.

An outcome of both studies is that the superficial use of a visualisation, i.e., seeing if something pops out by changing the view (switching tools) when no thoughtful interaction within a tool could be observed and when insight generation was low. This activity can be interpreted as confusion and decreased satisfaction. We could further observe that more diverse strategies were employed, which we interpret as the heuristic method of trial and error. This indicated that the applied strategies are either not well supported or the expectations on the visualisations support are not fulfilled to one's satisfaction and, therefore, caused a desire to switch. We take a close look at this switching behaviour to analyse which visualisation is used for different sense-making strategies and analyse the relation of applied sense-making strategies and generated insights. In one specific case, the interaction of switching the whole visualisation can seen as favourable. The

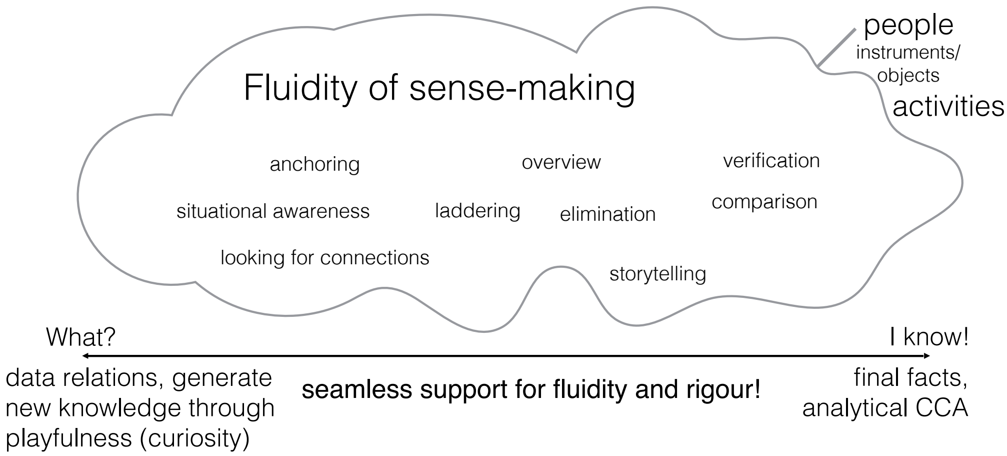


Figure 3.3: Overview of the aspects of the analysis process considered in study 3. In this investigation we address the question of how to make sense of events in dynamic, incomplete crime networks.

strategy of verification could be exposed through the experimental design in study 3, where participants could only see one visualisation full screen at one time. Switching to the other visualisation for verification purposes in this case led to concluding a task.

3.1.4 RQ4: How useful are models of insight generation?

Does the representation of the problem influence insights [HS74, ZK10]? Does using different representations lead to different kinds of insight and which visualisations are especially useful in the insight generation process?

It is suggested that representing the problem in a different way can support insight as, e.g., rearranging data might produce new groups [HS74, ZK10, Spe14]. Based on the results of study 3 we discuss how the observed sense-making strategies fit to the model of fluidity and rigour (How analysts think, Figure 1.1). Optimal experience and performance can be reached by providing support for a variety of human thinking processes in order to cope with complexity, compare Figure 3.3.

In study 4 we evaluate a complex system with an exploratory open-ended study design. We ask the participants to report any insight they find relevant and then study the quality of the insights in a bottom-up approach. We assessed how well the task was fulfilled with a qualitative analysis and grouped the results to see when richer insights were gained.

To analyse insights the multi-tier model by Gotz and Zhou [GZ08], which consists of tasks, sub-tasks, actions and events, is helpful to address different granularities of insights. Applied to intelligence analysis the overall goal (*task*) might be, for example, to find out more about the network of criminal gangs. A *sub-task* might be to identify the gang member with the highest activity or the newest gang members or if old members change their crime behaviour. In study 3 we use several of such sub-tasks which focus either on temporal changes or groups to identify if the evaluated visualisations are more suited for certain tasks. For an open exploration we use the task *"Identify key network insights to generate strategic recommendations."* in study 4. The action tier is a generic set of user actions, like query, filter, sort, etc., that allow to accomplish sub-tasks. Gotz and Zhou identified 20 actions and assigned them in three categories that describe the intention of the action: exploration, insight and meta actions. A meta action would be one with regard to a users own analysis history like undo or redo. Insight actions can be identified by user annotations and recordings of findings, e.g., bookmarks. Exploration actions are all those that have the intent to change the view on the data.

Three levels for sense-making of graphs were introduced by Friel et al. [FCB01]: (1) reading the data (2) reading between the data, and (3) reading beyond the data. Similarly, we may describe insights on different levels. The result of an action is the outcome, e.g., getting a list of criminals sorted by their activities. This would be a so-called *low-level insight* resulting from interacting with the dataset. Looking at the phone calls of one criminal might lead to new connections, and further low-level insights like these might add-up to a *higher level insight* if the connections made are of special relevance, e.g., a new gang member can be identified as a consequence. An overview of this example is given in Table 3.1.

Our evaluation is mainly based on the action level. We particularly analyse exploration and insight actions. In study 3, 4 and 5 insight actions constitute the answers to the tasks. In study 3 the experimenter prompted and recorded the final answers after each exploration phase. In study 4 and 5 we asked participants to document their insights.

Table 3.1: The granularity of insights gained through network analysis in criminal intelligence.

	Detail	Connections	Inference
Tasks	Identify pattern	Identify gang	Predict development of a gang/identify possible mitigation strategies
Sub-tasks	Identify single offender	Identify phone calls between two offenders	Predict successor of imprisoned offender
Actions	Identify data point	Identify connections/contradictions	Predict/identify causal relationship
Events	Hover	Click	Filter

3.2 Methods

We used a mixed approach for user studies, combining qualitative and quantitative methods, like thinking aloud, observation of behaviour, interaction log data and written reports/diaries. An overview of the used methods is given in Table 3.2. RQ1 was addressed with observations and interviews, the remaining research questions RQ2, RQ3 and RQ4 were addressed with exploratory studies where we manually code the usage behaviour via screen captures and thinking aloud protocols for sense-making strategies.

With the qualitative approaches we could emphasise the human reasoning processes as a whole to show up new patterns within the data, in contrast to quantitative measurements like task performance. The thinking aloud protocols provided explanations on why the tool was used in a certain way, raised new questions or confirmed results from quantitative data. In the thinking aloud experiment sessions we used screen and audio recordings as well as retrospective cueing as a basis for a thorough content analysis.

We use empirical research methods with exploratory protocols such as thinking aloud [ES93, BR00], interaction analysis [JH95] and the insight-based method [SND05]. These methods are prevalent in the field of HCI and have a broad range of application as they can be adapted for different domains. The data analysis is more contextual and two different experimenter will most likely produce different evaluations. The issue of reproducibility of results in this context is an ongoing discussion in the community, as no solution to the problem is in sight, and, currently, the upsides of exploratory protocols prevail [DL94].

3.2.1 Questionnaires

In our first study we compare which group meetings are perceived best in a variety of visualisations. Here it was possible to collect standardised information from 21

Table 3.2: Overview of methods used in my research.

Research	Data collection	Analysis
Guidelines	Interview, Literature review	Requirement analysis, User stories
Investigation 1	Questionnaire	Mixed*
Investigation 2	Pen & paper questionnaire	Mixed*
Investigation 3	Thinking aloud observation, Interview	Qualitative content analysis, Interaction sequences
Investigation 4	Thinking aloud observation, Insight-based method, Questionnaire	Mixed*, Interaction sequences
Investigation 5	Thinking aloud observation, Interview	Mixed[1]

*Combination of qualitative and quantitative methods participants via an online questionnaire (an excellent method to reach a large number of people to collect quantitative data). A big advantage in our case is that participants do not have an "eye over the shoulder" feeling and can work on the tasks without pressure or time constraints. A disadvantage of questionnaires is that one might not reach the level of detail in qualitative responses compared to, e.g., an interview. Free form fields are one option to collect thoughts in a more open manner. To answer the research question of study 1 we combined pre-defined answers (checkboxes) with free form fields.

Another option is to use pen & paper, which is an effective method that allows for creative freedom and helps to express, develop, and communicate concepts with a low entry barrier [Bux10, TS09, Tve11, WHC15]. In study 2 we collect hand drawn sketches using this method in a questionnaire. In both studies we use a within-subject approach so that the sample size is not reduced due to the number of evaluated visualisations. Another advantage is that inter-personal differences have no effect on the results, but this design also imposes severe constraints on the length of the questionnaire and, thus, the number of questions. Participants are likely to get tired after some time and about one hour is the maximum time before participants require a break. The result of surveys can, on the one hand, be influenced by learning effects or, on the other hand, by symptoms of fatigue, which both should be counteracted [YS10].

Extra efforts to the questionnaire design need to be made to avoid memorising answers while testing different visualisations with the same dataset. Using the same dataset is necessary to remove a confound - different datasets could lead to artefacts that might favour one visualisation over the other. Hence, to use the same underlying dataset without being recognised by the participants the data labels were altered to avoid repeating answers. We can not eliminate the chance of recognising patterns in the answers entirely, but by choosing a challenging setting chances are less likely than they are with simple tasks.

3.2.2 The thinking aloud protocol

Human-computer interactions are often evaluated with direct observation, video analysis, key stroke logs or pen & paper questionnaires [CFL10]. The thinking aloud protocol aims to reveal the intentions of the interaction, which often can not be interviewed directly as participants tend to forget them until they can be asked, e.g., via retrospective interviews.

The original thinking aloud protocol by Ericsson and Simon [ES93] suggests that participants verbalise their thoughts on their own during an experiment, where certain tasks are solved, and the observer is only allowed to remind the user to *think out loud* at times in a standardised, unobtrusive way. This should reveal aspects of the user interactions that are not accessed easily by other methods. The behaviour of talking out loud is very unnatural, especially, in the presence of an observer. Ericsson and Simon suggest that the successful application of the method is a matter of training to get used to verbalising thoughts and not actions. In study 3 we let our student participants practise for 10-15 minutes with pre-defined tasks giving them the chance to not only get used to thinking out loud but also experience the side of the observer. As observers they could get a feeling of which kind of information is more interesting – comments on the users' thinking steps rather than comments on actions that can be observed anyway.

Nevertheless we used the thinking aloud protocol in an engaging, conversational manner, suggested by Boren and Ramey [BR00] rather than the strict way that was envisaged originally. This way the resulting data is more valuable for HCI because the experimenter can instruct the verbalisation process and unnecessary information can be filtered out immediately. Furthermore, no training of the participants to use the method is needed, which is essential when dealing with experts. In study 3, for example, we could not ask the group of experts to train a method because of heavy time constraints. Last but not least, this engaging protocol is especially suited for evaluating prototypes and innovative technologies, which might need assistance in difficult situations [BR00, Tom15]. In study 4, for example, the prototype needed to be restarted at times when components remained static or crashed.

3.3 Analysis

Laboratory studies often contain qualitative methods, such as experimenter observations, thinking aloud protocols or surveying a participant's opinions, to enrich the quantitative results of the experiments and, in the best case, be able to explain them [Car08].

3.3.1 Qualitative analysis

For the interpretation of the thinking aloud data we used qualitative content analysis [Sch12, May03]. This is an empirical methodology for the systematic investigation of textual data that preserves some of the advantages of quantitative content analysis but still yields rich qualitative information about textual communication. Qualitative content analysis can either adopt a bottom-up or a top-down approach. The top-down approach requires developing a frame of categories before the study, the bottom-up approach consists of a repeated processing of the material with the goal to structure the material in a way to derive the categories from the material itself. Hence, to ensure that we cover all kind of sense-making strategies of the individual analysts we adopt a bottom-up approach but we are also inspired by the Triple Path Model of Klein, suggesting some categories like looking for *Connections*, showing up *Contradictions* and experiencing *Creative desperation*.

For the analysis we use the *Count Coding System* as a systematic observational measurement. This system helps to count key observations as number of instances or duration of instances [YS10]. The number of specific behaviours, the duration of behaviours or the time between observed behaviours can be of interest.

In the course of the VALCRI project we observed how police intelligence analysts work in their working environments and which systems and tools they use. A focus group and one to one interview findings helped define the requirements for the VALCRI system. During the initial stage we used Cognitive Task Analysis to understand the end user requirements. The requirements were collected from a focus group with 20 intelligence analysts. This was followed by one to one interviews with seven intelligence analysts using a retrospective interview technique, the Critical Decision Method [HCS98]. From the requirements and the literature review we derived guidelines for the VALCRI prototype, which is described in chapter 4.

We also adopted the insight-based method [SND05] to count insights in the high-level analysis of study 4 and categorised insights into low-level, moderate and high-level insights. We assessed the quality of the insight based on the following criteria: whether the participants provided a textual explanation of their insight and whether the text was only an explanation or contained an additional recommendation. We also counted the number of these insights. The task required that the insights were all of a similar abstraction level. The research questions for this kind of investigation was: "How do users generate insights with intelligence specific visualisations and which tools do they use to gain insights?". We, thus, use the following measures in study 4:

- Number of insights: How many insights could be gained during the analysis?
- Quality of insights: On which level are the reported insights? How many conclusions can be made for future actions?
- Tool use for insight reporting: Which visualisations are used to convey the insights gained during analysis (screenshots in report)?

The coding scheme for sense-making strategies was developed in an iterative approach combining bottom-up and top-down elements, where two authors studied the content together and refined a coding scheme to eleven content-specific sense-making codes through repeated in-depth discussions. The resulting coding scheme employed in Doppler-Haider [DHSP+17] and [DHGP+18] is in turn based on the work of Klein [Kle13]. The aim of the development of the coding scheme was to find categories that were distinctive and on the same level of abstraction. Consequently, the entire thinking aloud protocols are coded with these categories by one researcher. A second researcher independently codes parts of the protocols randomly. This approach is a common technique to verify qualitative results [KPS+17].

This, of course, is a trade-off of the mixed approach. To be able to do quantitative analysis of qualitative data a minimum of 15-20 participants are needed to fulfil requirements for statistical testing. For the qualitative analysis the data needs to be prepared for the analysis of two independent researchers. Verbal protocols usually get transcribed manually, which is a very time-consuming process because the material needs to be watched more than once. In our case approximately 150-200 hours of experiment sessions where analysed per study.

Elaborate discussions during the development of the codes typically result in a good understanding of the coding scheme. We have high agreement rates of two independent coders in our studies (2, 3 and 4) and the inter-coder reliability is very high too, e.g., Cohen's Kappa $\kappa = .871$ in study 3 ($N = 115$).

In study 4 both researchers fully coded the verbal protocols of each participant (with high agreement rates). In study 3 a second evaluator coded 12.4% of the protocols, choosing two minutes each from the first and the second half of the screen capture at random. In study 2 a second evaluator coded an entire task per participant at random.

3.3.2 Quantitative analysis

We use quantitative analysis in the mixed approach studies to compare groups from the qualitative results in order to look for significant differences between them. The comparison of means is appropriate for the deemed-to-satisfy provision of small populations in qualitative research. Statistical tests, such as the Kruskal-Wallis test can be used for that purpose. However, this test shows only if there is a significant difference between groups. This does not mean that there is a significant difference between each pair of groups. For this purpose a post-hoc pairwise comparison of the groups is necessary. We use the Wilcoxon rank sum test with the Bonferroni adjustment method for that post-hoc analysis.

Furthermore, we use inferential statistics as the American Psychological Association (APA) promotes reporting of intervals since 1999 [Wil99]. The HCI community also criticises significance testing due to the replication crisis in social sciences and the fact that the practice of "p-picking", fishing for significant results, is theoretically possible [Cai07, Dra16]. Inferential statistics uses a small sample of the population to make some inference about a population. By using confidence intervals we can make educated predictions about the approximate value of the population parameter instead of one point estimate [Urd11]. P-values capture only the strength of evidence against the null hypothesis that groups are about the same (no difference in the means). The downside of null hypotheses testing is that the p-value as a single value gives not much information about the distribution. Confidence intervals are based on the same statistical theory but are much more informative because the interval indicates the extent of uncertainty and enables a nuanced interpretation [Cum14, FL09]. Another advantage is that the number of participants looses importance [Dra16]. We, therefore, report effect sizes and interval estimates for our observations. We provide 95% confidence intervals, which describe values that contain the population parameter, e.g., the population mean, with a probability of 95%.

To address the question which sense-making strategies participants use most, the following measuring are appropriate for quantitative analysis:

- Frequency: How often is a sense-making strategy used?
- Duration: How long does a sense-making strategy take?

Such results will then be interpreted using the qualitative data, as our focus is on the thinking aloud protocols. Furthermore, we assess the correctness of tasks, where possible. In open exploration tasks there is not always a clear cut answer, but one can assess plausible answers. In study 5 there is only one statistically correct answer per line-up. Here, we have to look at the results per condition to analyse and interpret the sense-making strategies of the users with varying geometry and map type, compare Figure 3.4.

Figure 3.4: The interpretation of quantitative results requires a detailed analysis of the study material. In study 5 we used 24 conditions which we printed and spread out on a desktop for comparison and discussion during the analysis process.

In study 5 we use the line-up method, which is an established tool for visual testing to evaluate the validity of graphical findings in an objective manner, similar to standard statistical inference tests [WCHB10, HFMC12]. The line-up is a visual inference technique that got its name from the police line-up (or identity parade) where a suspect is placed within random persons in order to be identified by a witness. Similarly, the plot of the data is placed next to random null data, and an impartial observer is asked to identify the most different plot. The probability that the observer identifies the plot of the real data is called the *power* of the test. In null hypothesis testing the power is the probability to reject a (true or false) null hypothesis. Low power indicates an inconclusive result as it does not distinguish between a true null and the case that the experiment just could not detect the null. A problem for replicating results is that most published papers lack power at all [RK16]. Hofmann et al. show that although power is difficult to calculate analytically because it is requires an alternative hypothesis, the power of line-ups can be estimated as "the ratio of correct identifications x out of n viewings" [HFMC12, p.2]. Line-ups, therefore, can be used to evaluate the effectiveness of plot designs.

We analyse sense-making strategies during geographical line-up tasks for the first time to better understand how the line-up task is solved. We investigate, on the one hand, the influence of artefacts that are incidental to the statistical structure under investigation.

Visual artefacts are especially relevant in geographies, as humans tend to perceive Gestalt-like structures in visualisations [Kof35]. This may induce an interpretation of false structure that affects the original task of the line-up. On the other hand, we shed light on the interpretation of spatial autocorrelation in geographic maps. We use the *Moran's I* statistic to quantify the amount or *intensity* of spatial association – the degree to which one object is similar to other objects nearby [Mor50]. The quantitative analysis here reveals strategy usage per line-up and the success per strategy, i.e., if certain strategies lead to correct or wrong answers in the autocorrelation task.

3.3.3 Research ethics

The research conducted in the course of this doctoral thesis follows the code of ethics of the faculty of informatics and the code of conduct at Vienna University of Technology [COD07].

As our research includes human participants we were concerned about the impact and their well-being during our investigations. Ethical considerations include potential risks and discomforts as well as privacy and data protection. To avoid discomforts we planned breaks whenever necessary and that the whole process including introduction and follow-up questions take no longer than one hour. Furthermore, participants were informed that they could stop and discontinue if they wished to and that their data would be destroyed on request. The data collected in the course of our user investigation was chosen carefully to the purpose of their aims. The privacy of participants was maintained through an anonymous evaluation as the results were analysed independently of any personal information.

In all user investigations we informed the participants about the aim and procedure of the investigation, the estimated time planned and provided contact information for further questions. We also asked for their agreement to publish potential findings of the collected data and let them sign a confidentiality agreement. In the thinking aloud investigations we recorded their voice and tool use via screen capture software to ensure that nothing important was lost in the process. These recordings did not contain any personal information and were stored for the evaluation process only.

3.4 Summary

I use tried and tested research methods of the HCI field, so that we can rely on the results. I choose a mixed approach to get a complete picture of the tool use and be able to analyse sense-making strategies with qualitative content analysis. The thinking aloud protocol is also an established method, but different variations in the strictness level (from reminding to think aloud as the only interruption by the experimenter to answering questions and helping with technical issues of a prototype) are used in the community. A key challenge here is to get meaningful verbal reports of what the participants think.

Research on sense-making deals with theoretical concepts of thinking [PC95, PC05, KPRP07, KMH06a, KMH06b, Won14] and the analysis of real world decision making

[Kah11, Kle13]. This kind of research is challenging because human behaviour is so complex that even one oneself is not able to explain every idea or insight. This is typically referred to as a *Gut feeling*. Other types of insight describe an *Eureka!* or *Aha! moment*, when a new understanding is suddenly available or the solution to a problem is finally found. Using the think aloud protocol in a "soft" approach makes the unnatural situation more comfortable through answering questions, and, therefore, potentially improve the likeability of people to talk. It further enables the experimenter to ask questions about particular insights right in the moment, where the probability is highest that the participant remembers her thoughts. Here, retrospective cueing is not an option because people most likely forget the seemingly unnecessary steps in between and focus on the main insights they gained during the analysis. Thinking aloud is a common method applied in usability studies because a few participants can already show up the main issues and the peak of insight is gained quickly (after a few participants answers tend to repeat and new insights are gained rarely). With regard to sense-making, however, we are interested in commonly developed strategies, and are, therefore, especially interested in such repetitions. Hence, we do qualitative studies in larger quantities than usual and aim rather for 15-20 participants.

Asking about sense-making processes in an interview will not yield a complete picture. In the interviews with experts we asked about procedures which have to be followed and routines that were developed and what they do differently because of their experience (as opposed to their formal training). We are aware that even these answers are skewed because individuals like to think of themselves of being rational and not biased. Biases are well-known examples of unconscious steps in our thinking processes and specific training and methods are used to counteract on them as they are critical for the outcome of the analysis. Bias mitigation is a whole research area on its own and is not part of the research questions. However, some kinds of biases, such as the confirmation bias could be observed in the thinking aloud protocols.

CHAPTER 4

Empirical part A: Guidelines based on literature

In the course of the R&D project VALCRI on sense-making in visual analytics we retrieved results from empirical research to derive 17 principle guidelines for the development of the system prototype. Our work within the project was located in the human issues work package, including a literature review and the development of guidelines, which was primarily conducted by the author of this thesis.

We conducted a literature review on the topics sense-making and insight and insight generation with visual analytics tools with a focus on empirical results. On this basis we carried out user studies aiming for more detailed information and to extend this knowledge base for the design of systems. Through these studies, we put forward recommendations for specific contexts (to be described in the second practical part).

The following guidelines were developed on the basis of an extensive requirement analysis with 12 intelligence analysts from three law enforcement agencies of two different countries. In the following I will first introduce the requirements of the VALCRI project that built the basis to develop a human issues framework with respect to evidential structuring and reasoning, sense-making, and cognitive bias mitigation. Then I will elaborate on the guidelines regarding sense-making, which were our contribution to the human issues framework of the project.

4.1 Requirement analysis of the VALCRI project

The goal of the VALCRI system is to facilitate human reasoning and analytic discourse, tightly coupling semi-automated semantic knowledge extraction with human mediation. The goal of the requirement analysis was to capture the different workflows in a few use case scenarios and identify the analysts' goals. Employing a use case scenario, interviews and questionnaires requirements from twelve end users were collected. Shortcomings of current systems reported by the analysts highlight where analysts would like to be supported and how they think of an optimal analysis process.

The requirements from the end users were described as user stories, which contain the *role*, the *desire/tool*, the *benefit* and optional *comments* of the end users. As an outcome of the analysis 500 user stories were collected and structured in to 28 themes that cover, e.g., the user interface and system functionalities, intelligence processes, such as tactical, strategic and operational planning, or data presentation and reporting. The resulting extensive list of user stories should provide a comprehensive overview in a way that is useful for the design of the prototype. The list was audited from each partner of the human issues work package with regard to their field of expertise. Based on their feedback comments were added on evidential structuring and reasoning, sense-making and insight, bias mitigation as well as ethical, legal and privacy issues.

We contributed 94 aspects that need to be considered for sense-making and insight generation. There were several user stories addressing similar issues with regard to sense-making. We note that some user stories were not covered by current literature since they were too specific. The literature review showed up several gaps where no research results could be found. In our investigations (described in part B) we could address some sense-making aspects in visual analytics, which were not addressed previously in the literature. Before that, we could derive 17 guidelines for visual analytics systems from a sense-making and insight perspective.

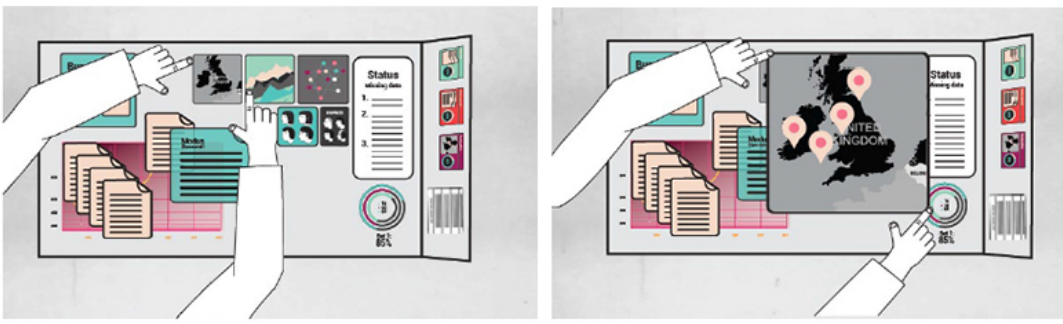


Figure 4.1: Early sketch of the user interface for the VALCRI prototype based on the requirements analysis using a large canvas, tangible multiple views and two screens: one for interaction and one for presentation.

4.2 Using design actions as a framework

Based on the requirement analysis of the VALCRI project we created a guideline catalogue within the framework for theory-based interaction design by Spence and De Bruin and their concept of design actions [DBS08]. As Spence [Spe11] points out, knowledge from cognitive research is often times not directly applicable for system design and a process of translation is needed to *broker* the insights to system designers and developers. Similarly, this approach was already employed in Haider et al. [HPF13] focussing on the representation of floating car data of traffic management systems.

Consequently, the project's development partners drafted a first design for the user interface based on the requirements and the guideline catalogue, compare Figure 4.1. They further applied the guidelines in the incremental implementation phase, which included formative evaluation with the analysts to ensure that the developed system is useful and stays relevant to work.

In the following I will describe the literature review, analysis of previous research and the derived guidelines in more detail. To target developers the guidelines were prepared in the form of design actions, which supposedly deliver only the important gist of theories and empirical results. The original format suggests a long tabular format with ID, title, description, effect, upsides, downsides, issues, and theory fields as rows. The title and description explain the meaning of the guideline in short and in long. The effect describes its consequences. Advantages and disadvantages inform the developer about the main factors allowing trade-off decisions, which might be necessary. Finally, the underlying theory and an example are given as a reference. We made minor changes to these fields as we summarised the up- and downsides in one field which we call *trade-off*. An additional field for the context refers to the theme of a user story. We think the context is especially relevant to make informed decisions in the domain of visual analytics. The original catalogue for the VALCRI project included an optional field for solution approaches to refer to practical application examples. This information is omitted here as they are not part of this thesis.

The adapted template of a design action can be found in Table 4.1.

Table 4.1: Guideline template based on Spence and De Bruin's framework of design actions [Spe11].

ID	Gx short title
Title	
Description	
Trade-off	(optional)
Context	Within the user requirements of VALCRI
Theory	Reference
Example	

4.3 The guideline catalogue

The first guidelines (G1-G7) are mainly influenced by Klein's research on naturalistic decision making [Kle13] and comprise recommendations of what should be supported by the visualisation on a general level. The next eight guidelines address the human reasoning processes and basic visualisation principles that can be found in the literature. The remaining three guidelines concern specific visualisation techniques which are useful in the domain of intelligence analysis. The design actions can be found in the Appendix, while the investigations are described in chapter 5.

In the following, the seventeen guidelines G1-G17 from literature will be discussed. They are grouped in themes addressing insight generation and reasoning, finally moving on to practical applications from basic principles for visualisations to specific application examples.

- Insight generation: Representations should show up
  - G1 Contradictions
  - G2 Clustering
  - G3 Connections
  - G4 Coincidences
  - G5 Deviations or unexpected ideas
- Reasoning: Users should be supported in taking an active stance
  - G6 Motivate curiosities
  - G7 Provoke questions
  - G8 Holistic processes
  - G9 Open exploration
  - G10 Locking-in assumptions
- Visualisation principles:
  - G11 Graphs
  - G12 Time
  - G13 Clutter
  - G14 Interaction
- Recommended techniques:
  - G15 Node-link diagrams
  - G16 Repeated search
  - G17 Multiple views

4.3.1 Insight generation

G1 Contradictions

Guideline 1: Support the user to identify contradictory information and ideas that just do not make sense.

Analysts working with a visual analytics tool like VALCRI should be supported in detecting stories that do not make sense. The system should identify differences between observed patterns, e.g., in the modus operandi (MO) of criminals. Visualisations could be used to present these differences for easy inspection, to support the human in the loop. Contradictions warn the analysts that something is seriously wrong with the stories we are telling ourselves and the hypotheses we generate. More importantly, it tells us that we have weak anchors – important beliefs that are not fully supported. However, there is a trade-off because it might be difficult for users to reach a coherent conclusion if the system emphasises contradictions. In the end, we need to form hypotheses that can be followed up upon with further investigations.

Klein [Kle13] points out that contradictory evidence can be a powerful motivation for getting novel insights. There are different ways contradictory evidence can be made visible. Semi-automatic analysis methods can show patterns in the data. A human analyst is necessary to assess the significance of these patterns.

Gestalt theory is a powerful tool to highlight differences and similarities to users. In this regard the gestalt principles of proximity and similarity are relevant. The law of proximity (similarity) says that objects that are near (similar) to each other are seen as belonging together [Pin90, CS98].

This is relevant in the context of patterns, trends, relationships, and outliers detection. A volume crime-related task would be, e.g., to identify deviations from normal day-to-day activities which can be interpreted as indication of suspicious behaviour. Contradictions cause the analyst to think "That can't be right! That doesn't make sense!".

As part of the VAST Challenge 2014 [CGW14], the analysts were asked to assess the significance of deviations from normal working hours in a company. Four employees met regularly outside the premises of the company during working hours. This is contradictory evidence which has to be explained and can be an indication for something unusual. Such patterns can be made visible in a visualisation, e.g., by overlaying this information on a map, or possibly by animation. Sometimes there are reasonable explanations for deviations. The human user has to decide if a pattern is really unusual. When employees come to work at night, for example, it is not necessarily an unusual, suspicious behaviour, although it is a clear exception from the rule. There might be unexpected patterns that can be explained by additional information. The contradicting information can prompt further inquiry to find explanations that make sense.

G2 Clustering

Guideline 2: Enable the user to cluster entities individually.

The system should allow for easy clustering and hotspot generation. Visualisations can be used to achieve such clustering [ALS97]. Data may be clustered in different ways, depending on the criteria for the clustering procedure. Different granularities can result in a trade-off as they may induce clutter, which generally should be avoided (compare G11 Clutter).

It is sometimes difficult to identify a clustering procedure which helps the user to gain insights. In node-link-diagrams (networks), for example, the users may cluster persons with similar attributes, crimes with a similar MO. Note that it might sometimes be difficult to identify the criteria for what constitutes a similar MO to indicate that the same group of persons is responsible for a series of crimes.

Clustering is a possibility to show similarities in the data. There are various mathematical methods to cluster data. Such methods can support analysts in finding similarities, e.g., in MOs of criminal gangs. The results of the mathematical methods can be visualised on the screen so that analysts can see which data is similar at a glance. It should be possible for the users to influence the clustering process (e.g., to choose how many clusters should be presented), so that different views on the data are introduced.

Clustering can be enabled in various ways. Users could be given the possibility to let the system perform clustering algorithms or similar mathematical procedures. It is also possible to let the users cluster data manually in a visual form (e.g., in a social network).

G3 Connections

Guideline 3: Indicate connections in the data in an appropriate way and show new or other pieces of information and non-information.

This guideline is also based on Klein's models of sense-making [KPRP07, Kle13]. Klein suggests that facilitating the making of connections supports insight generation. The system should indicate connections between data in an appropriate way. Connections between people, for example, can be shown as social networks. Such networks can indicate connections between people, which are not obvious immediately. The system should enable the user to explore such networks.

Furthermore, the system should help the analyst in distinguishing between dots, non-dots, anti-dots, and other ambiguities. "Dots" represent information relevant to the problem; "Non-dots" show information that in fact contain irrelevant messages; "Anti-dots" are data points that are of different kind, contradicting, conflicting, such as situational information that may make the ideas seem less likely, though perhaps plausible. To make these new connections could spark up new ideas or open up new pathways to consider or investigate.

A trade-off in this guideline is that too many connections might be confusing and may lead to clutter, which should generally be avoided (compare G11 Clutter). Furthermore, there is an external restriction through the limits of screen sizes, which allow to show only a certain amount of connections.

Connections between people, for example, can be shown as social networks. There are different ways in which connections can be emphasised in the data. On the one hand, social networks or other network-like visualisations can make connections visible. The problem about social networks is that such visualisations may get too large for normal screens. If the network contains too many nodes and links, it is difficult to identify connections between single entities. Allowing the users to filter the data is a possibility to overcome this problem. This also enables the user to look at the data from different points of view. Another possibility to show connections is multiple views. Multiple views enable the user to see connections which cannot be shown in one single visualisation. It should be mentioned that multiple views can also show contradictions.

G4 Coincidences

Guideline 4: Enable the analyst to re-organise and present data in multiple ways that can reveal repetition or co-occurrence.

The visualisation and the interaction techniques should enable the analyst to see repeating events, patterns or co-occurring similar events in context. These are coincidences and associations one does not fully understand, or relationships one can not (yet) articulate – but they look like they are related in some way. This can cause one to break from initial stories or to break away from initial anchors, and to initiate alternate stories.

Klein [Kle13] as well as North and Shneiderman 2000 [Shn96] suggest that insights are sometimes coincidental. Multiple views, for example, can also show up coincidental connections. One possible evaluation criterion, therefore, is whether multiple views and linking and brushing exist or not.

G5 Deviations or unexpected ideas

Guideline 5: Enable the analyst to persistently see the ideas created by combining and grouping data.

The visualisation and interaction techniques should enable the analyst to be persistently exposed to other, unexpected and less expected ideas. This guideline is relevant for pattern detection, trends, relationships as well as outliers detection. It is based on Klein's Triple Path Model of Insight and addresses the situation of creative desperation, which needs to be overcome to reach an insight [Kle13].

Enabling the user to change the type of visualisation for the data can show different aspects and, therefore, increases the possibility to show unexpected information. This is a general guideline which is related to enable clustering (G2), showing connections (G3), and enable identification of coincidences (G4).

4.3.2 Human reasoning

G6 Facilitate the perception of curiosities

Guideline 6: Raise curiosity in the analyst for unusual single events that stands apart from the trend, such as an outlier.

Curiosities make the analyst wonder about something. This reaction is not yet the insight, but might trigger a path leading to insight. Curiosities are sparked by a single event, while coincidences are observed as a repetition of patterns [KPRP07, KJ11, Kle13]. Hence, it is suggested that the system should emphasise outliers or other unusual data, as well as coincidences (G4). Curious analysts ask themselves "what is going on here?" or follow up on thoughts, such as "that's funny", "that's strange", or "that's odd". However, curiosities might also manifest as a distraction. Curiosity might cause oversight of important patterns, e.g., when the analyst focuses too much on single events. A trade-off between supporting coincidences and curiosities is necessary.

G7 Provoke questions

Guideline 7: Provoke the analyst to ask questions.

The visual representation should support open exploration and provoke questions by system design. The intention is to create situations that prompt the analysts to ask questions such as, "That's interesting - why did that happen?" or "Why are there missing elements?" This can be done, e.g., by emphasising "Blackholes" or gaps where data is missing in a series. This guideline is crucial for tasks such as looking for patterns, trends, and relationships, as well as outlier detection. The challenge in the application is how to point out the *unknown unknowns* – the things we are not aware of not knowing.

This idea is based on Klein's Data-Frame Model ("Question a frame") [KPRP07], Klein's Triple Path Model of Insight [Kle13], Klein and Jarosz [KJ11] and Wong and Varga's "Asking questions" [WV12].

G8 Holistic sense-making processes

Guideline 8: Information should be structured in a coherent manner to support holistic sense-making processes.

Gestalt psychology indicates that sense-making is a holistic process where structure plays an important role. There is empirical evidence that everyone develops his or her own structure. It therefore makes sense to allow users to generate their own structures of knowledge whenever this is possible [PNS03]. The analysts should be able to, e.g., structure network representations (e.g., social networks) themselves if it makes sense in the context of their work. We emphasise that this needs to be an easy process. If structuring is too complicated or time-consuming it is unlikely that it will be used at all as a major characteristic of the intelligence environment is time-pressure.

This especially concerns the ability to freely organize different views in the user interface, but also adding personal information such as notes or annotations. This is also relevant for data presentation and reporting.

G9 Open exploration

Guideline 9: Allow for open exploration and the redefinition of the goal and methods.

Research in everyday reasoning indicates that so-called wicked problems are solved differently than clear-cut problems [DS03]. Wicked problems have no clear method or path to the solution and it is sometimes not clear how the solution might look like. It is typical for such problems that users explore the problem space and often redefine the problem while working on the solution. To support such problem solving processes, systems need to be open and allow the redefinition of the goal and the methods to reach a solution [Kle13]. Exploration should be supported in a way which does not lead users astray but helps to brainstorm on the one hand and on the other hand to focus at the end of the exploration process.

This applies to areas where it is necessary to be uninhibited in exploring the data creatively. From a system point of view it concerns data processing and the personal knowledge management.

G10 Facilitate the identification of assumptions

Guideline 10: Enable the analyst to identify and change assumptions that lock-in.

Klein [Kle13] discusses situations, where we become fixated on or trapped by our assumptions or habits. They are the basis of our understanding of the world and their events. Analysts should be encouraged to reflect on hegemonic beliefs and supported in thinking *outside of the box*.

This general guideline is especially relevant for data annotations in intelligence systems. User annotations and selections should be considered in the user interface and its system functionalities.

4.3.3 Visualisation principles

G11 Graph design

Guideline 11: Design of simple graphical representations.

This guideline concerns the visualisation of graphs. There is a systematic body of research on the design of simple graphs, which include aesthetic guidelines that can be applied in a straightforward way (e.g., avoid edge crossings) [Kos89]. The type of graph design needs to match the context. Literature also makes some suggestions which graph design suits a certain context. This kind of information supports the design process of simple graphical representations. Line graphs, for example, are usually interpreted as trends even if the data does not show a trend. Lines are better able to represent values than other visualisations.

A design guideline for node-link diagrams (G15) is described in subsection 4.3.4.

G12 Time

Guideline 12: Use different temporal representations for the visualisation of processes over time.

The traditional linear timeline is only one possibility to present temporal data. To describe recurring events, circular visualisations should be used, as they allow to identify recurring patterns easily [Joh96, AAD$^+$10]. There might be, for example, more crimes at specific times/days of the week, which is more evident in a circular visualisation than in a linear one. It is probably a good idea in many cases if a visualisation, that is most appropriate for the data, is recommended by the system.

Especially challenging is the design of several co-occurring patterns [KBB$^+$10]. Another challenge in this context is the fact that time has different and irregular granularities (years, months, weeks, days, hours, etc.). System design needs to take individual preferences into account, such as date formats which differ from country to country. As does the reading direction. Therefore, timestamps need to be visualised in an clear and unambiguous way.

A requirement in the context of crime intelligence, for example, is to determine similarities and highlight differences in MOs automatically. It is sometimes interesting if crime types are the same or if they are different, e.g., if they are becoming a bigger threat when they get more violent. Time as well as geographic information is of interest on different levels of detail (time of day, season variation, districts or larger regions, etc.) Semantic zoom should be implemented, so that more details are shown as the chart is enlarged in the area where the zoom is applied. This should be implemented not only for temporal, but also for spatial visualisations.

G13 Clutter

Guideline 13: Do not overwhelm the users with too many connections.

Visualisations, even static ones, easily overwhelm human perception when many data points are shown. Possible solutions for this problem could be filtering or highlighting mechanisms to reduce the visualisation space to connections of interest. Filtering mechanisms should be appropriate for the user's task. Peng at al. [PWR04] suggest that reordering of multi-dimensional data can already reduce clutter. In some situations an automated clutter reduction can be used but, here, the problem is to decide what constitutes clutter. This can sometimes be difficult because it depends on the users and their tasks how much information they need. A combination with hierarchical data visualisation is considered a good trade-off in this context.

Crime maps need to convey a lot of attributes in different granularities, such as the incident numbers per year (as well as per hour a day, etc.), type of the crime, boroughs or country boundaries. Spatial exaggeration is a critical factor. The data of interest needs to be exaggerated in order to make connections and contradictions visible. Here, telephone records and the routes of buses, for example, might be connected in some way and an exaggeration of antenna regions may reveal that the mobile phone data is following a certain bus route. This would be a valuable insight for the analyst.

G14 Interaction processes

Guideline 14: Provide interaction to support sense-making.

The theory of Distributed Cognition assumes that interaction with cognitive tools, such as the VALCRI system, plays an important role for sense-making processes [HHK00]. These interaction processes should be designed carefully and *less is sometimes more* is part of the design decision for interaction. This indicates that not every possible interaction technique should be implemented – indeed some interactions should rather be left out if they might mislead or distract the analyst. In most analysis contexts, e.g., crime intelligence, it is crucial to support bias mitigation techniques. When to offer filtering and which attributes should be supported in the filtering are examples where design decisions for interactions are necessary. These decisions need to be made based on a good foundation, such as empirical results from research.

4.3.4 Practical applications

G15 Design of node-link diagrams

Guideline 15: Design node-link diagrams in a way to support mental models.

Node-link diagrams can represent connections in the data. They should be designed in a way so that significant relationships in the data can be perceived and processed easily in order to develop a coherent mental model of the relationships in the data. The following design recommendations are suggested by, e.g., Huang et al.[HHE06] and Archambault and Purchase [AP16]:

- Reduce crossings of edges
- Cluster nodes which belong to the same group
- Highlight important nodes; put important nodes at the top or in the centre
- Preserve the mental map in a dynamically evolving node-link diagram to increase memorability

The design of node-link diagrams is relevant in several contexts where relationships are established, such as briefing/reporting, link/network charts, and visualisations for patterns, trends and outliers detection.

G16 Repeated search

Guideline 16: Provide a history of queried searches for quick reuse.

In principle, users appreciate the feature of saving search queries because they want to spend less time on repeating the same steps. They usually do not start with a fresh analysis every day and rather want to continue their work at the point where they left it the other day. Hence, a requirement is to restore a search result and being able to reuse the thinking steps that lead to it, i.e., the selection of keywords. This needs to be considered carefully in the system design. One possibility would be a search history.

Graphical histories facilitate analysis, communication and training [HMSA08]. However, such a feature comes with certain challenges, which necessitate a trade-off decision. The design of a search history is challenging because users can be overwhelmed easily by the amount of the accumulated information. The search for a saved query could result in extra work for the users and may take longer than to start a new one. Structured search results may be an approach to overcome the problems of a lot of accumulated results. However, another problematic aspect needs to be addressed at the same time. This feature could inhibit the users' motivation to make new searches. It therefore needs to an active decision and not as easygoing as possible, i.e., automatic restoring via default search queries is not recommended.

G17 Multiple views

Guideline 17: Show the data in multiple views to represent different aspects of the data.

Multiple views help users to assess the value of different solutions and might even help to overcome some forms of biases. This technique encourages the analyst to address the problem holistically and examine more than one type of the data presentation method [RFF+08]. Here, it is important that the transition between various data views is easy and fluid. The multiple linked views should further support cross filtering, so that connections can be made [NS03, KGS11]. Multiple linked views can also help to get a better picture of the events when different views can be compared next to each other. The restrictions due to screen size require a trade-off for this guideline because all views need to fit on the screen. Depending on the visualisation making the visualisations smaller might induce clutter, which should be avoided (G11).

In the case of CCA it can be useful if comparing crime rates in different years allowing to detect seasonal connections. Timeline plots and similar visualisations are suggested by Boba Santos [BS13].

4.4 Example of applied guidelines

VisuExplore, a system prototype for the analysis of medical data, proved to support the detection of connections (G3) and contradictions (G1) within different diagram types [PWR+11]. Different temporal representations are used for the visualisation of processes over time (G12). The prototype comprises event charts, line plots, bar charts and timelines, where diagrams share a horizontal time axis. The specific requirement of this project was to provide a simple visualisation. Encoding time by a common scale is perceptually effective and is supposed to be easy to learn. Furthermore, the tool is very flexible and provides interaction processes (G14). It is possible to change the diagram type, scale, add new and delete diagrams. For comparing different variables users can change the order of the diagrams to better see connections in the data. Moreover, one can zoom into the data and resize diagrams. The representation scales with the available screen size, showing more or less detail and thereby allowing the user to chose the level of detail, which supports holistic sense-making (G8).

However, there is a trade-off in how many diagrams can be compared at once. The idea is to solve this problem through interactivity. By dragging and dropping it is easy to change the view and to try out if something new pops up. With the VisuExplore tool analysts found a contradiction in the medical data of the patients' development when they compared weight and sugar level. They could see that the patients' weight was going up when the sugar level was going down. This is a contradiction because with growing weight the sugar level normally also goes up. This visualisation is simple and effective and works well for the range of offered activities.

In the course of the VALCRI project we also described criteria for the formative evaluation process, which can be used as evaluation criteria for visual analytics systems in general. To verify if the guidelines are followed one can ask the following questions:

- Can users interact with visualisations?
- Does the system support clustering?
- Does the system support the visualisation of and interaction with social network representations?
- Are there visualisations representing time?
- Are cyclical processes supported by specific visualisations?
- Does the system support multiple views and linking and brushing?
- Are there too many connections in the visualisation – is the visualisation cluttered?

4.5 Dissemination of guidelines

A detailed report of the guidelines was published in the proceedings of the International Summer School on Visual Computing [HPPW15]. The process of developing guidelines from empirical results as well as exemplary guidelines are published in the proceedings of the European Intelligence and Security Informatics Conference [PHPW15] and the Proceedings of the Human Factors and Ergonomics Society Annual Meeting [HPH$^+$nt], where I contributed the related work and the description of the guidelines.

CHAPTER 5

Empirical part B: Investigations on sense-making processes

The focus of our investigations lies on sense-making with visualisations on different levels of detail. At the beginning we evaluated simple systems, comparing static visualisations using a questionnaire with traditional measures for accuracy, error rates and preferences. We then investigated the process of transforming graphs from one visualisation to another using pen & paper, allowing more creative data collection. To investigate sense-making strategies in visual analytics we used the thinking aloud method with parts of the VALCRI prototype as well as an integrated system prototype.

An overview of the visualisations and methods used in the investigations:

- Investigation 1: Map, Gantt chart, matrix and line diagram studied with a survey (task: Identify location-based meetings)
- Investigation 2 & 3: Node-link diagram vs. matrix
  - Investigation 2: Transformation from node-link to matrix diagrams and vice versa studied with a pen & paper questionnaire (task: transfer mental model)
  - Investigation 3: Dual system designed for realistic tasks studied with observation and thinking aloud (task: assess temporal evolution)
- Investigation 4: Multi-component prototype (integrated system prototype) studied with observation and thinking aloud (task: describe insights from crime analysis)
- Investigation 5: Choropleth map vs. centroid-dot map studied with observation and thinking aloud (task: identify higher autocorrelation in line-ups)

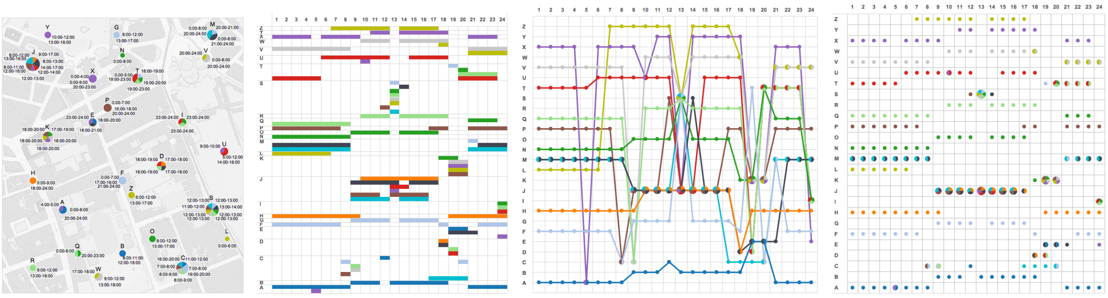
The goal in the first and third investigation was to find suitable representations for specific tasks in the intelligence analysis domain. Several visualisations showing the same data – information on actors over time – get evaluated. We study meetings of suspects (investigation 1) and network developments of co-offenders (investigation 3). The first investigation evaluates Gantt charts, bar charts, line charts and a map with straightforward lookup tasks. We also compare a node-link based multigraph and an enhanced matrix in a more realistic system evaluation setting in investigation 3. The tasks were realistic and open-ended, which was ensured by designing them with criminal intelligence analysts. The study design was open too in a way that users could switch freely between the available tools and were not forced to use both for every task. Using complex tasks enabled us to investigate participants' sense-making processes. This is a link to the second investigation, which was concerned about how graphs are represented in ones' mind (mental model) and which strategies are used to transform graphs from a node-link diagram to a matrix and vice versa.

In investigation 4 we evaluated the system prototype of VALCRI – a decision support system for criminal intelligence analysis - with regard to sense-making strategies. The goal was to find successful sense-making strategies that lead to insights when using a complex tool for visual analytics. We were interested if they are similar to strategies that could be identified when using simple tasks or tools as investigated in investigation 3.

We also conducted a more general investigation with maps concerning patterns of spatial autocorrelation – the measure of correlation of a variable with itself through space, which is relevant not only in the intelligence domain but any domain using geographic visualisations. In investigation 5 we evaluate two map types, a traditional choropleth map and a newly designed centroid-dot map, with spatial line-up tests to see which is better suited for assessing spatial autocorrelation. A high spatial autocorrelation in a map exists when similar values are located near to each other. This can be calculated by different means but it is interesting if sense-making strategies are biased due to the geographic interpretation.

In the following I will describe the parts of the studies, which are relevant for the results and my thesis. This includes the description of the visualisations, the dataset and tasks and the outcome of the study. The outcome of a user study can often times inform system design, although the generalisation of results can be challenging. In the reviewing process we were often asked to answer questions such as *How do these results affect system design?* or *What implications do these results have?* In an effort to collect specific design recommendations from each contextual result I created guidelines in the spirit of part A of this thesis, chapter 4, to provide advice on the application of the results. The design recommendations yield a new result from the analysis for the thesis and are not included in the publications. Finally, I define my responsibilities in each investigation and published paper.

5.1 Investigation 1: Location-based meetings



(a) (b) (c) (d)

Figure 5.1: Four visualisations were used to depict locations of 12 persons: a) a map, b) a Gantt chart, c) a matrix and d) a line diagram. To counteract occlusion a), c) and d) use pie-charts for depicting the meeting information. A meeting is defined as two or more persons being at the same location at the same time.

In this study, we wanted to know what kinds of visualisation support the sense-making processes of intelligence analysts (research question 1) and if it is more helpful to adapt visualisations that are already used (research question 2), like timelines or geographical visualisations or if new visualisations can be beneficial. We conducted two surveys in a within-subject design with the goal to identify visualisations that give a good overview of location based information in combination with temporal information. We collected qualitative and quantitative data from 24 participants in an online questionnaire, based on which we were able to derive six specific design recommendations (see Table 5.1).

Table 5.1: Design recommendations based on investigation I1.

	Recommendation	Description
1.1	Avoid using maps for the representation of temporal data	We evaluated a static (Two-dimensional (2D)) map containing time information as labels of locations.
1.2	Use maps in combination with other visualisations	It may be useful to combine a map with one of the other visualisations, for example, in a coordinated multiple view setting.
1.3	Use auxiliary lines carefully	We suggest to use auxiliary lines with caution, for example, only on a certain subset of entities.

Continued on next page

Table 5.1 – *Continued from previous page*

Recommendation		Description
1.4	Avoid white space between entities of the same group	It may be beneficial to provide possibilities to rearrange the order of entities in the visualisation to ensure that entities belonging to the same group are displayed next to each other.
1.5	Show duration of meetings explicitly	Showing the time explicitly improved the identification of (recurring) group meetings.
1.6	Consider the influence of visual properties	The qualitative content analysis revealed an influence of certain properties of the visualisation on which groups were perceived as salient or not.

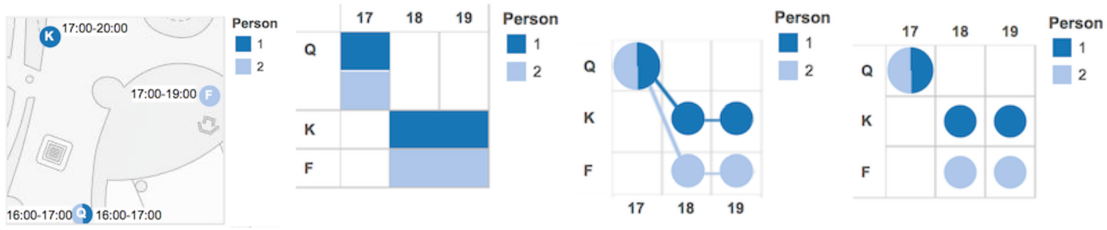
5.1.1 Motivation

We were interested in the visualisation of geographic information of social networks over time. To show where people are at different times we developed four representations: a map visualisation Figure 5.2a, which is very intuitive but has well-known drawbacks, such as overlapping data points or information overflow when temporal information about movement or meetings is included. In comparison to this traditional geo representation we show locations in an abstract way in a structured Gantt chart Figure 5.2b with continuous time-bars as well as in a matrix representation Figure 5.2d with colour coded dots per time step and pie charts for group meetings. As a fourth representation we added lines to connect actors and compare this line chart to the traditional matrix Figure 5.2c.

With this study we address the problem that the representation of movement data on maps leads to visual clutter quickly (e.g., occluded data at the same position). Furthermore, the analysis of movement data over time is often a big challenge with traditional visualisation approaches [AA08]. An interesting information in the context of Crime Intelligence Analysis is the position of persons at the same points in time to identify potential meetings. Since the results of an extensive literature research show no solution to this problem, we designed alternative visualisation techniques for this problem. We consider linear diagrams in addition to the traditional map representation. Lines and bars in linear diagrams previously showed similar performance [RSC15], which is why we developed a new visualisation technique using pie charts to show overlaps.

5.1.2 Description of the visualisations

Figure 5.2 gives an overview of the four visualisation techniques that we considered in our study using a minimal demonstration dataset. Each visualisation shows who meets whom at which time. They represent the locations Q, K and F, two actors, the blue Person 1 and light blue Person 2 and their social network (meeting activity) over three time points from 5-7pm.



(a) Map (b) Gantt (c) Line (d) Matrix

Figure 5.2: Person 1 and Person 2 meet at location Q. A Map (a), Gantt (b), Line (c) and Matrix (d) representation of two actors at 5, 6 and 7 p.m. [KHWP16]

In the map locations and time spans are given as labels. Persons are encoded in colour and are visualised with a pie chart to overcome the problem of overlaying points when persons are at the same location. The pie chart in location Q is composed of light blue (Person 1) and dark blue (Person 2). The time frame of a person being in a specific location is given in the label of the location for each person.

The axes of the Gantt chart are time and location. Persons are encoded with colour. Persons in the same location are drawn beneath each other.

The axes of the matrix chart are time and location. Similar to the Gantt chart, persons are encoded with colour. Locations are drawn as a pie chart. The size of the pie chart increases the more persons are at the same location.

In this line chart the locations of persons are connected by a line. To overcome the problem of overlaying points when persons are at the same location, pie charts are used again in the same manner as in the map and the matrix.

5.1.3 Tasks

The participants were given four questions per visualisations:

- Question 1: Who meets for more than one hour?
- Question 2: Who meets at different places?
- Question 3: Which group is most/second most/third most striking? Please describe what makes them striking.
- Question 4: Please describe any patterns or anything interesting.

5.1.4 Approach

We conducted an extensive online questionnaire to analyse different visualisation techniques - a Gantt chart, a matrix, and a line chart - as alternatives to the commonly used map visualisation. The ultimate goal was to identify which of these techniques give a good overview of location based information in combination with time information.

The questionnaire was split in two parts due to the heavy workload – the second part was available after a one-week break. One questionnaire lasted about one hour. For this study we generated a simplified version of the tracking and movement data from the Mini Challenge 2 of the VAST Challenge 2014 [CGW14]. The assumption is that analysts focus on groups of suspicious people only. Hence, we designed a smaller dataset with 12 persons, 26 locations and two different complexities in time, 24h and 48h. With four different days we ensure that the task performance is not bound to a single dataset.

Both questionnaires were grouped in

- General/demographic questions: In addition to their age and gender, the participants were asked to rate their familiarity with visualisation techniques
- Visualisation questions: The participants were asked to identify person groups who were at the same position longer than one hour, as well as to identify identical groups who were on different positions at the same time. Furthermore, they were asked which groups were interesting or if they notice unusual movements.
- Preferences: In the end we asked the participants which visualisation techniques they liked most.

The two questionnaires were sent out via email to Computer Science students and results are based on 24 responses for the first questionnaire and 18 responses for the second questionnaire, respectively.

The survey included closed questions using checkboxes accompanied by open-ended questions to offer the possibility to explain decisions in a free form text field, and, finally, rank-ordered questions to rank preferences of the different representations. All tasks had to be solved with all visualisations, which were shown in a random order to avoid biased results.

The dataset was kept moderately small to ensure that the visualisations fit on an average screen without having to scroll. The shown data was the same in each visualisation. Learning effects were considered in the design of the task such that answers were not easy to memorise even after answering the task several times. To facilitate this, we conducted pre-tests to see whether participants could remember answers from one visualisation to the next. The four visualisations representing the first test case is given in Figure 5.3.

Dataset

We generated two simplified variants for our case study on the basis of the VAST Challenge Dataset 2014 [CGW14]. That was necessary to not overload the user with too much data and keep the questionnaire within the range of one hour. We chose two different timespans (24h and 48h) to test different sizes of the dataset and 26 locations (A-Z) for our abstract dataset, making it realistic enough for the tasks and to counteract any learning effect. We ensured that each person is at least meeting someone once in each test case to keep the number of relevant persons at the same level. 12 persons were colour-coded using a colour-scheme for qualitative data from ColorBrewer [HB11].

An overview of the four test cases is given in Figure 5.4.

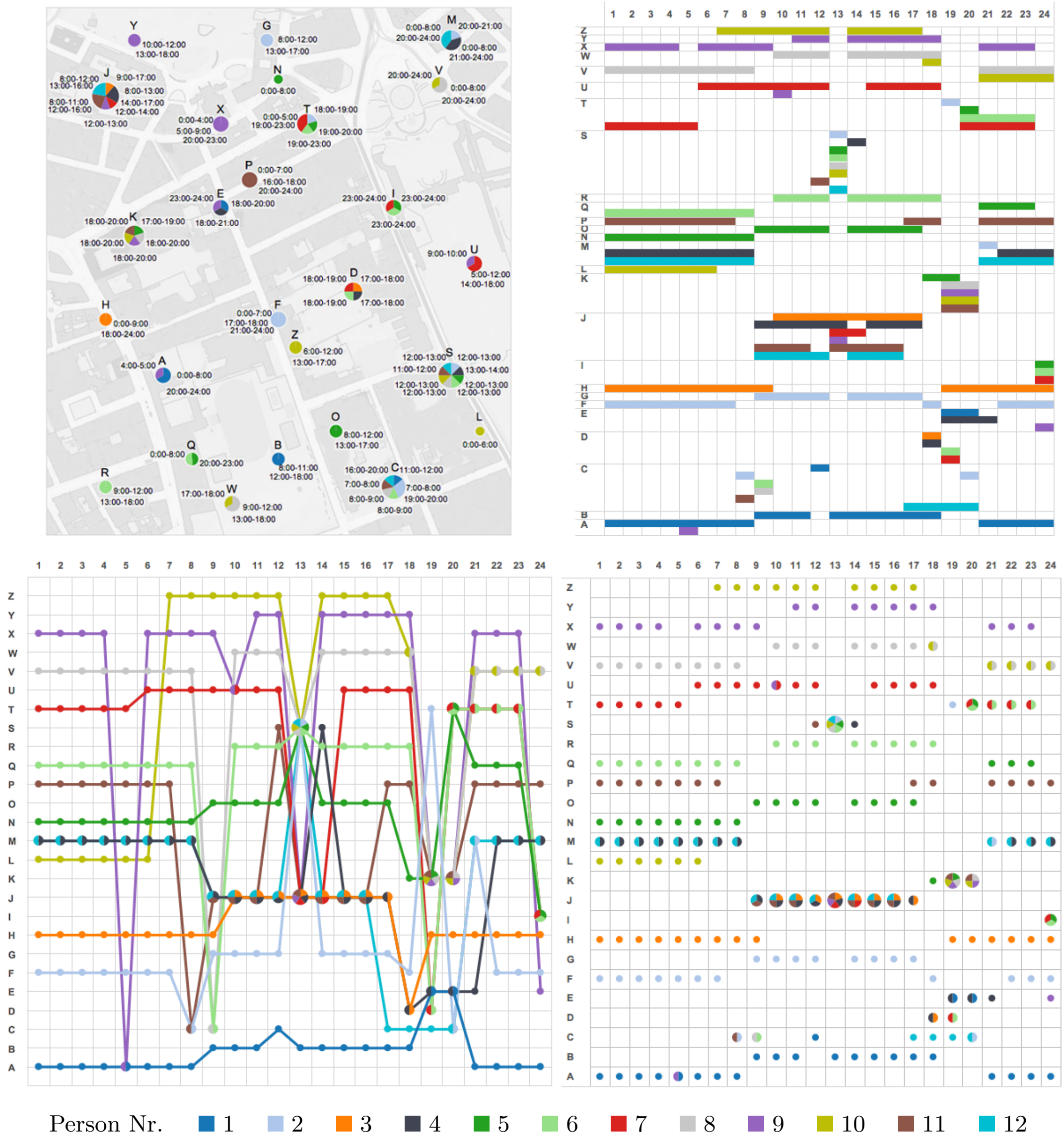


Figure 5.3: Day 1 represented in a map, a Gantt chart, a line chart and a matrix. The visualisations were given in random order – a learning effect from using the same data within the four visualisations could not be observed.

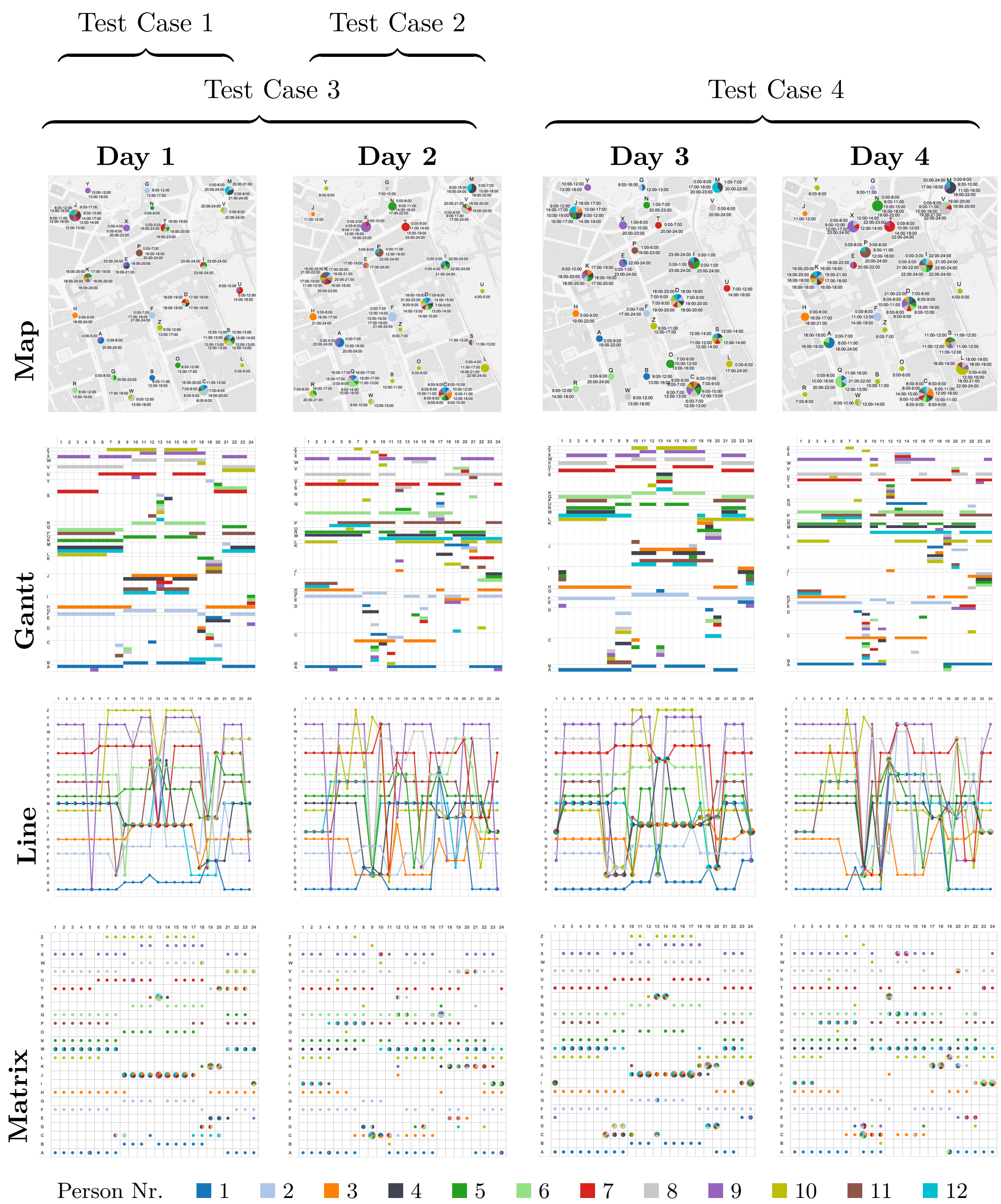


Figure 5.4: Overview of the time-dependent location visualisations and the different test cases used for the surveys. There was no effect of the data size. Participants preferred the matrix visualisation, which yielded the lowest error rate.

5.1.5 Results

The 24 participants were 70% male and 30% female between 23 and 35 years. None reported any colour vision deficiencies.

We targeted participants who are used to work with different kinds of visualisations and recruited computer science students in an advanced master study to participate in our study. Asking about their familiarity with visualisations 45.8% reported very familiar in a five-point Likert scale ranging from not at all familiar and slightly (12.5%) to highly familiar (12.5%). The remaining 29.2% stated moderate familiarity.

The results showed us that the matrix and line chart are good alternatives for the representation of location based data. Participants preferred the matrix visualisation (58%) to the line (33%) and the Gantt chart (8%), and the map was liked the least for the given tasks (0%). In the interview they reported that they could develop a structured way to answer the tasks with the matrix visualisation. The results from the tasks show that the matrix yielded the fewest errors and the highest acceptance rate. Participants ranked the line chart as second best, with which they made errors especially in busy areas where more people meet.

5.1.6 Conclusion

The map visualisation as a traditional way to show position data seems to be not efficient for displaying movement data in combination with time information to identify groups. Furthermore, the pie charts that were used to cluster persons who were at the same position on the map appear to be less-than-optimal because the people-per-pie limit is reached quickly. A pie chart can optimally represent meetings of two persons, but as soon as more persons are at one place at different times it is difficult to extract this information. However, the spatial information can lead to new insights, for example, that people stay in the same area.

5.1.7 Comments on dissemination

My responsibilities in this investigations include the implementation of the visualisations and the design of the datasets and execution of the survey. The qualitative analysis and the dissemination of results was done in co-operation with the first author of the paper *"Who, Where, When and with Whom? Evaluation of Group Meeting Visualisations"* [KHWP16], which was published in the 9th International Conference on Diagrammatic Representation and Inference in the Diagram Layouts track.

5.2 Investigation 2: How users transform node-link diagrams to matrices and vice versa

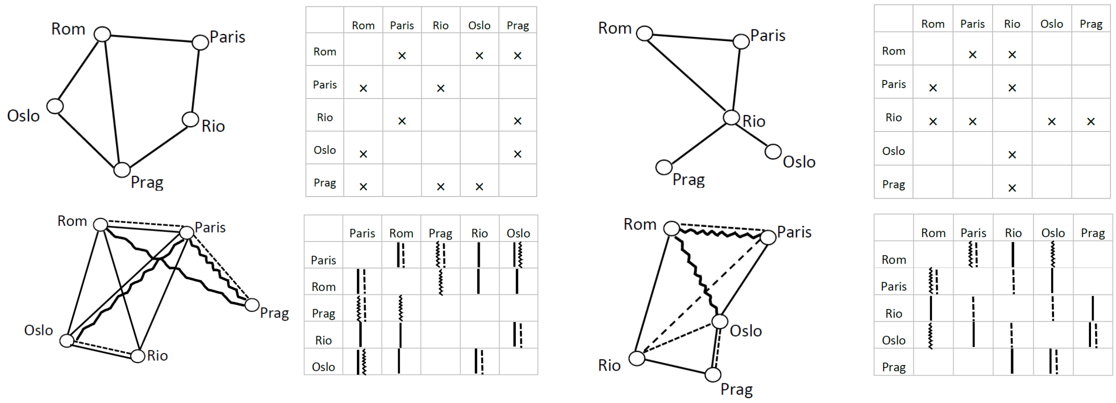


Figure 5.5: Four graphs, two with a circular structure and two with a star-like structure, are each represented in a node-link diagram and a matrix.

We conducted a user study with the aim to better understand how users transform a node-link diagram to an adjacency matrix and vice versa. We are on the one hand interested in the efficiency of the transformation, and on the other hand, the interpretation of the graph and how adjacency can be depicted. As research questions we defined the performance and interpretation of the transformation in terms of correctness, completion time and sketching of the graph in the respective visualisations. We chose different connection types and complexities for the test cases to see how they influence the performance. The results also indicate how easy it is to switch from one representation to the other in a multiple view visualisation.

Table 5.2: Design recommendations based on investigation I2.

	Recommendation	Description
2.1	Use clockwise or counter-clockwise arrangement of nodes	We suggest a clockwise or counter-clockwise arrangement of nodes if the meaning of the node-link diagram is not defined.
2.2	Use × symbol to represent one connection type	A × is an intuitive symbol used for a connection between nodes in a matrix, especially when the connection is simple.

Continued on next page

Table 5.2 – *Continued from previous page*

Recommendation		Description
2.3	Use the same representation for connection types in both visualisations	Switching from on visualisation to the other is supported by using the same representation for links, e.g., using the same line types which represent the different connections between nodes in a node-link diagram in the corresponding matrix and vice versa.
2.4	Avoid semantic meaning of the order of connections	Avoid that the order of the connection types represents any semantic meaning in case this information is not already visible in other ways. If the order of connections, for example, represents the timing of events (such as the left line showing the first year, middle line illustrating the second year, and right line representing the third year) it may be useful to use another visual property, like, e.g., colour (red for the left line, blue for the middle line, and green for right line).
2.5	Represent the reading direction of connections along the left side	The results revealed a preference to indicate the reading direction of the connections between nodes along the left side of the matrix, e.g., as arrow.
2.6	Use intra-cell representations for showing second-degree neighbours in matrices	This form of representation was preferred by the participants of our study.

5.2.1 Motivation

The current literature suggests that it is beneficial to combine different views that complement each other, such as the node-link diagram and a matrix representation. The considerable amount of research about their respective strengths and weaknesses is discussed in the theoretical background, subsection 2.2.3. In essence their upsides can be leveraged by combining the representations but users have to read them in different ways, suggesting a difference in the cognitive processes and sense-making strategies.

5.2.2 Approach

We used the pen & paper method and asked participants to draw node-link diagrams based on given matrices and vice versa. We handed out the transformation tasks in three rounds and finally, collected the participants' preferences regarding the type of visualisation. To identify strategies how users convert these visualisations we randomly

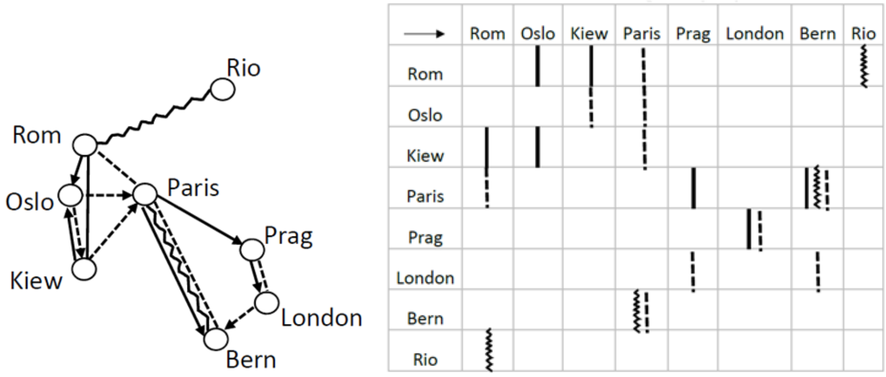


Figure 5.6: Node-link and matrix representation of the directed graph [KPDH18].

assigned the participants into two groups and let half of the participants draw a matrix first and the other half start with the node-link diagram.

The first two rounds included single connection examples which lasted 47 minutes on average. In the third round we handed out two tasks about design solutions for extended adjacency matrices representing second degree connections, i.e., neighbours of neighbours, which took the participants only around 14 minutes to complete. Finally, we asked the participants if they found the conversion from matrices to node-link diagram or from node-link diagram to matrices more laborious.

5.2.3 Dataset

We used two test cases, each containing four graphs with 5-10 nodes and 5-8 links between them. The context of the data were cities and possible transportation modes. In the first graph we only used a simple graph with one transportation mode to verify if the transformation between the visualisations is understood in general. The second and third graph showed between one to three transportation modes, e.g., bus, train, plane, for two complexities: a smaller graph with five nodes and a bigger graph with ten nodes. Figure 5.5 shows the first two graphs of the two test cases, which contain the five cities Oslo, Paris, Prague, Rio, and Rome (in alphabetical order) and for the first graph single connections (top row) and for the second graph multiple connections (bottom row).

5.2.4 Description of the visualisations

We used parallel lines for links in both visualisations. The different transportation modes are represented by different line styles: a straight continuous line, a dotted line and a zick zack line. We also evaluated directed connections, compare Figure 5.6.

The positions of the nodes play a huge role in the node-link diagram. They can either emphasise the structure of the graph in a clear way by balancing the white space and shortened lines, reduced edge crossings, or, obscure it by producing edge crossings and cluttered areas in a chaotic manner. We follow the guidelines for graph aesthetics in the node-link diagram as discussed in the related work.

We use two datasets: one with a circular- and one with a star-like structure. The structure of the data has an influence on the representations. The rows and columns of the matrix can be sorted to highlight groupings, as the positioning of nodes in the node-link diagram can emphasise group structures.

5.2.5 Results

The results were analysed in two steps. First, we assessed the performance looking at the correctness of the sketches and the completion times. In a second step, we analysed the sketches iteratively using qualitative content analysis [Sch12] to understand how node-link diagrams and matrices were drawn and, specifically, how multiple connection types were represented. A key outcome is that transforming node-link diagrams to matrices is cognitively less demanding.

Style of the transferred matrices

All participants depicted a symmetric matrix. Most interestingly, the resulting sketches of the participants depend decisively on the *order* in which either the matrix or the node-link diagram have been used, compare Figure 5.7.

Numbers were used, for example, in the matrices to represent the connection types between nodes instead of different line types. In return, there were no differences in converting matrices to node-link diagrams, regardless of any experience the participants had with converting matrices to node-link diagrams or vice versa. This leads to the conclusion that the design of node-link diagrams is straight-forward, whereas the representation of matrices includes several options, such as numbers vs. symbols.

Style of the transferred node-link diagrams

In contrast to converting a node-link diagram to a matrix, it seems that graph drawing aesthetics gain in importance (e.g., minimisation of edge crossings) if the number of nodes is increased. This aligns with the result of participants' preferring the matrix representation because of its clear arrangement in the raster layout. This is an advantage over node-link diagrams which suffer from edge crossings and can get confusing easily.

The structure of the graph was reflected in the position of nodes in the node-link diagrams. Adjacent nodes were drawn in a circle when the data showed a circular structure, compare Figure 5.8a and a main connector was put in the centre when the data showed a star-like structure, compare Figure 5.8b. A disadvantage of the pen & paper method is that participants can not reposition a drawn node easily. However, for the analysis this can be seen as an advantage since the positions suggest how the drawing evolved, which helped during the analysis of the drawings. We assume that participants started either with the top node or a centre node.

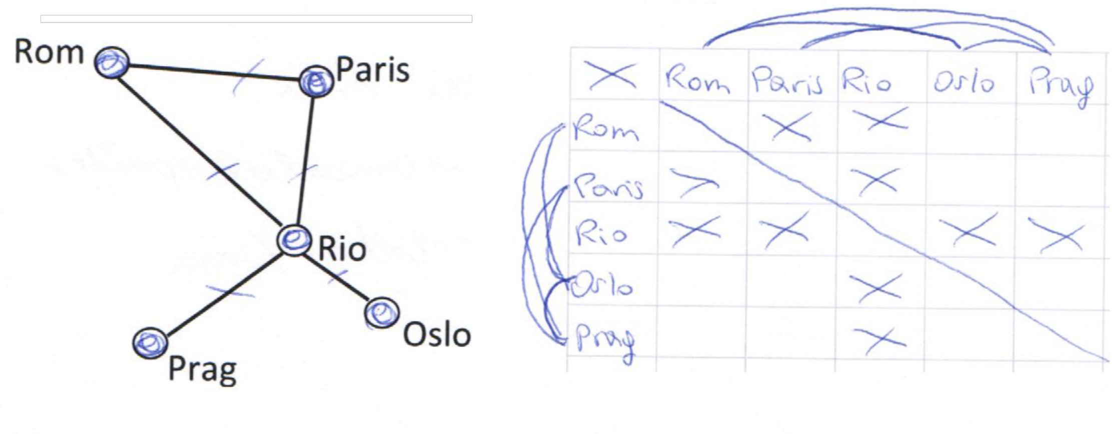


Figure 5.7: Example of the transformation from a node-link diagram to a matrix representation [KPDH18].

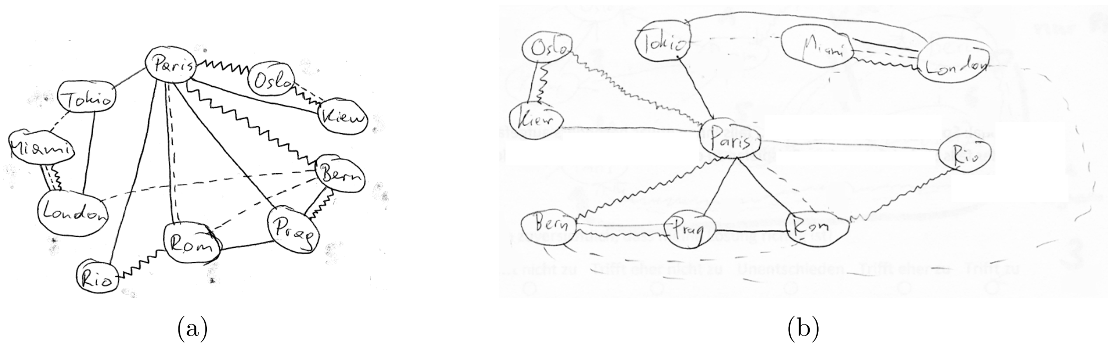


Figure 5.8: (a) Circular structure, (b) star-like structure [KPDH18].

5.2.6 Conclusion

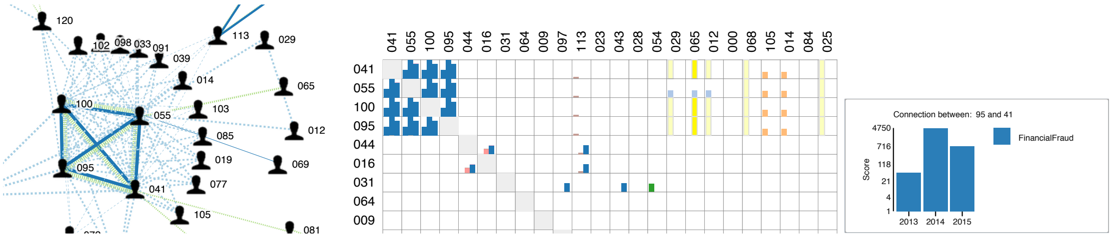
The familiarity with node-link diagrams has an influence how well they could be designed to show different connections. Multiple parallel edges are not as familiar as single edges. Our results confirm previous results on the importance of the structure of node-link diagrams. We did not give a meaning to the position of the nodes but used different structures, in contrast to Ballweg et al. [BPWv17], who used a strictly hierarchical layout in the arrangement of nodes in a node-link diagram. Providing a meaning, such as grouping or hierarchy, seems to influence the importance of edges. The results of Ballweg et al. indicate that edges are less important for the assessment of similarity in node-link diagrams.

A clear result is that converting node-link diagrams to matrices was cognitively less demanding than drawing a node-link diagram on the basis of a matrix. This is evident from faster completion times and a higher correctness rate while drawing a matrix on the basis of a node-link diagram. We further observed that matrices allowed more creativity in depicting the connections than the node-link diagrams. We conclude with a number of design recommendations, which are especially relevant for hybrid visualisations that combine a node-link diagram with a matrix. The design recommendations are summarised in Table 5.2.

5.2.7 Comments on dissemination

I participated in the design and execution of the study and collaborated in the analysis of the results. I further contributed the related work for publishing. The paper *"How Users Transform Node-Link Diagrams to Matrices and Vice Versa"* [KPDH18] appeared in the proceedings of the Diagrammatic Representation and Inference Conference 2018.

5.3 Investigation 3: Social network evolution



(a) (b) (c)

Figure 5.9: a) Node-link diagram with multiple parallel links representing the criminal activity over time. b) Matrix representation using intra-cell bar charts to show the criminal activity over time. c) Bar chart of criminal activity over time show in a pop-up view in both, the node-link diagram and the matrix.

In the third investigation we evaluated two visualisations, a matrix and a node-link diagram with regard to sense-making. We examined the capabilities of the visualisations for weighted co-offender relationships and for interpreting the superimposed temporal evolution per visualisation. We also investigated the feasibility of combining both visualisations. As a result we describe sense-making strategies of participants observed during realistic tasks. In turn, this allows us to suggest the design of a new system. Finally, this investigation led to the following design recommendations:

Table 5.3: Design recommendations based on investigation I3.

Recommendation		Description
3.1	Show connections of a social network	Node-link diagrams yield a good understanding of connections in the data; physical lines can be used to show extended connections in adjacency matrices.
3.2	Show temporal evolution in matrix representations.	The matrix representation proved to be efficient in look-up tasks (*Connections*) as well as *Comparison* tasks on temporal evolution.
3.3	Show data in different views	For the *Verification* of results, two representations such as a matrix and a node-link representation of social networks can be used in combination. It may be useful to combine both views, for example, in a multiple view setting. Strategies such as *Comparison* and *Contradictions* can be supported by multiple linked views.

Continued on next page

Table 5.3 – *Continued from previous page*

Recommendation		Description
3.4	Support *Trend Analysis*	Bar charts can be used to represent temporal development integrated into a node-link or matrix diagram.
3.5	Support *Storytelling*	Showing up gaps and contradictory information can provoke questions which might be explained through the users' imagination and capability to build narratives that make sense.
3.6	Support *Elimination*	Reducing the search space can be supported by interaction possibilities, such as highlighting or filtering.

5.3.1 Motivation

The analysis of evolving networks improves the situational awareness in strategic analyses. An overview of large networks of more than one hundred nodes enables the exploration and discovery of specific nodes of interest.

In the field of intelligence analysis social networks are used to evaluate problematic fields that might emerge, i.e., crime series that might become a bigger problem. Also, applied mitigation strategies are monitored by analysing the evolution of the criminal activity of such networks. Visualising such network developments is not straightforward because there are many attributes of the actors that are of interest for the analyst and it is not clear how the interaction in the tool is best designed to ease exploration. Experienced analysts know their tools very well and are very fast in the steps they take as time is mostly a crucial factor during their work. Especially when working with visual analytics tools it is essential to know which representation suits which task in terms of time and utility.

The analysts we worked with in the project experience a lack of visualisation tools but have specific ideas of how an intelligent system could support them. Therefore, we formulated the research question of how to visualise weighted networks over time. We sought to lay out the terrain of the criminal networks by making the constraints and boundaries visible in one view. Secondly, we wanted to semantically map the relevant relationships to appropriate visual geometries and represent the evolution of those relationships over time. Thus, appropriate visualisations need to represent co-offender *relationships, temporal development* and *weights* for the events.

5.3.2 Approach

In this study we worked with two different target groups in two separate studies. First we showed the visualisations to police analysts ($N = 6$) and observed how experienced analysts make use of them. Second, we conducted a controlled thinking aloud study with students ($N = 31$) to get a sense of how they work with different types of visualisations.

In both evaluations we observed how the participants worked with the visualisations and how they worked on realistic tasks. A task based methodology and the thinking aloud protocol were used in an open, flexible way with the experts and a strict, rigorous routine was followed with the students. For the analysis we used a mixed approach of quantitative and qualitative methods using audio recordings and screen captures of the experiment sessions.

We learned a lot about the requirements and tools police forces currently have at hand during the early stage of the VALCRI project. In volume crime analysis a focus is set by the respective agency, in our case a police force. The police force priority adds an aspect to the traditional network characteristics, relationships, attributes and time, namely, an artificial weight, that might be changed as the priority of the analysis changes. The priority of a police force usually is not altered more than three times a year. Working within the UK National Intelligence Model [Nat08], those priorities are set within a Shared and Publicly Agreed Priority forum involving community partnerships. Priorities must be SMART, i.e., Specific, Measurable, Achievable, Relevant and Timed, allowing for true accountability [Bog05]. Topological network measures can be deceiving if used in isolation. They are not able to reflect individual differences in the offender behaviour, which might be crucial in identifying critical offenders and strategic interventions. Hence, to fulfil the requirements of volume crime analysis an additional encoding for social network evolutions is needed. Therefore, we take police force priorities into account that stem from their internal intelligence cycle and prioritise the analysis by assigning weights to different crime types.

Once the priorities have been determined, it is necessary to target offenders who are responsible for committing crimes in that arena. New offences need to be assessed accordingly. To establish the significance of a crime, a current priority might be, for example, to determine which offenders are responsible for causing *harm* in the community. The definition of harm relies on the individual police force, e.g., "Drug Supply" can be prioritised differently in different regions. Nevertheless, police forces typically have to target too many individual offenders. Therefore, it is important to nominate those who are causing the most harm. This process relies upon the provision of intelligence from a variety of sources and the analysis of trained professionals to provide a judgement.

5.3.3 Description of the visualisations

A network of co-offenders, occurring if actors commit a crime together, is depicted in a node-link diagram and a matrix. Typically the past three to five years are considered for current policing decisions. Therefore, we decided to show the crimes of three years aggregated per year to show the development over the years. We considered the interaction and exploration framework "overview first, filter and details on demand" in the design of the visualisations [DS09]. The requirements of volume crime analysis should be covered in a single overview of the network over time. We determined animation to be inappropriate for evolutionary crime networks because they usually contain hundreds of nodes and, as in our case, weighted links add another dimension of complexity to the visualisation.

The analysts would much likely miss changes while following animated data of that magnitude.

A line in the node-link diagram represents the existence of a relation. We chose to use colour and style of the line for the temporal encoding of the time at which the crimes were committed between pairs, compare Figure 5.10. The encoding of time via segments or multiple parallel lines depends on the reading direction, i.e., left to right or bottom to top. A horizontal encoding of time in segments of the line is not unambiguous for symmetric relations in a social network. We decided on multiple parallel lines where each line represents a time interval, encoded by colour as the layout algorithm does not ensure that all lines are interpretable unambiguously (vertical lines occur). We use a colour-blind safe, 3-class-paired colour scheme recommended by colorbrewer.org [HB03]. To support the ageing metaphor, the line style changes as well from continuous to dashed lines for older relations. The recency can be discerned by fading colours, i.e., lighter colours represent older connections. The line thickness depicts harmfulness of the relation, i.e., more frequent and higher prioritised co-offences are represented by thicker line widths. The focus in our dataset lies on *Financial Fraud* and *Drug Supply*, which were assigned the highest weights.

Hence, we encode the temporal information in the node-link diagram as stacked coloured lines, like *multiple threads* [KAW$^+$14], with an additional differentiation in the line style. Only the most recent relation is shown via a continuous line, the earlier years are dotted lines with decreasing dash lengths for recency.

The matrix shows the adjacent relations of the network and the evolution between two actors via intra-cell bar charts. Additionally, indirect neighbours of 2nd degree are indicated by yellow bars, so that extended relationships can be investigated. Harmful developments, e.g., an increase in the criminal activity, is made evident through the height of the bar for each time period in the cells, compare Figure 5.11. The tabular format of a matrix provides a structured overview of the dataset, compared to the node-link diagram that follows a force-directed graph drawing algorithm. These algorithms ensure more or less equal lengths of the edges and minimal edge crossings, which is perceived as aesthetically pleasing. The full set of possible relations in the matrix can overwhelm novice users but experienced analysts are enabled to use an efficient working style. Problematic suspects can be represented in both visualisations through the layout. The most active co-offenders are centred in the middle in the node-link diagram, and in the top left corner in the matrix, compare the top four offenders in Figure 5.10 and Figure 5.11.

Therefore, both visualisations, the matrix and the node-link diagram, emphasise potential harmful developments and allow to investigate the whole network at the same time. Showing the same information in different ways leaves room for interpretations by the human in the loop.

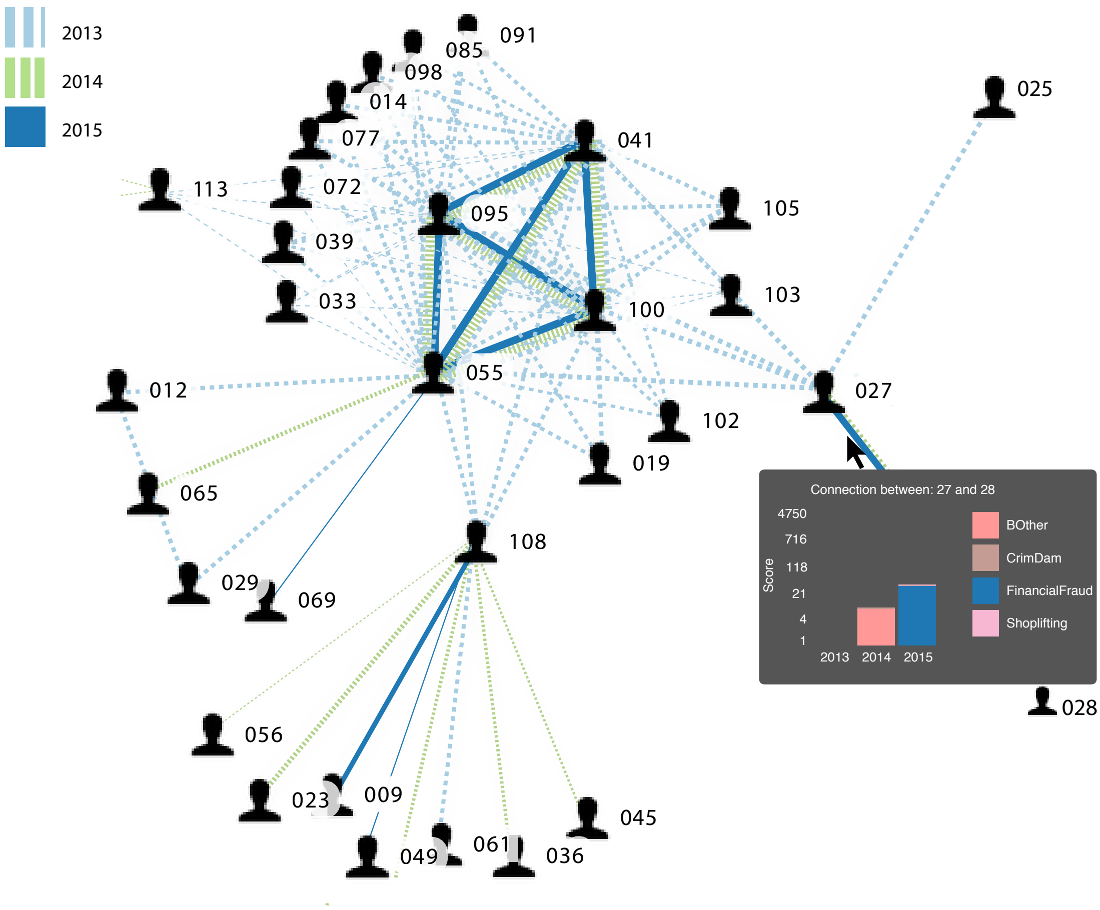


Figure 5.10: Excerpt of the node-link diagram showing detailed information of a crime relation in a pop-up window. Parallel lines between nodes represent the aggregated co-offences over three years encoded by colour and line style. Older relations appear faded out by using lighter colour hues and dashed lines. The width of a line represents the criminal activity: Person 27 and 28 (bottom right corner) committed more crime in 2015 than in 2014.

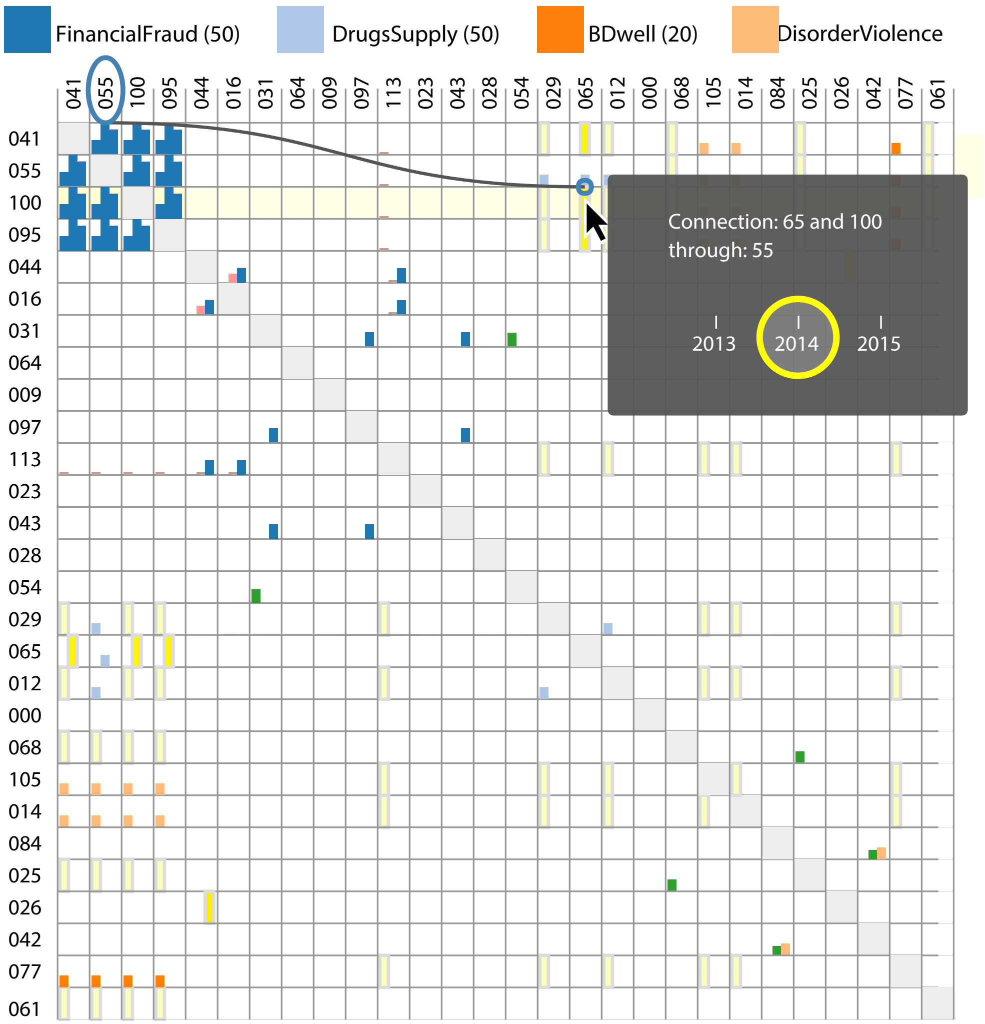


Figure 5.11: Excerpt from the matrix representation: intra-cell bar charts represent the co-offences over time; the height representing the criminal activity and the horizontal axis the years like a timeline. Yellow bars indicate 2nd degree relationships, i.e., indirect neighbours. Details of such a middleman relationship (100 - 55 - 65) are shown in a pop-up view when hovering over a yellow bar.

Dataset and participants

A realistic co-offender network from real police data was used. The dataset ran through a de-personalisation process to ensure anonymity. We chose a network with 121 actors from a set of hundreds of networks that resulted from a three-step algorithm to create co-offender networks under a prioritisation strategy [ABW08]. The first step removed offenders acting alone, the second step generated networks per co-offender assigning weights per crime type to apply the prioritisation strategy, and the third step aggregated the networks of overlapping offenders. The network was chosen due to its size. A requirement for the controlled study was to find a network that can be analysed within the available time frame of one hour.

For the first study, six senior police analysts spanning three different police forces from the UK and Belgium were recruited. All participants reported normal colour vision and none had prior experience using our visualisations. The participants were informed about the procedure and after signing a confidentiality agreement they were trained on the tool. The sessions lasted 42 minutes on average.

For the second study, 31 participants were recruited from two universities. 18 were male and 13 were female, with their age ranging from 24 to 34 years. 21 of the participants were Bachelor or Master students from Vienna University of Technology, the remaining 10 participants were PhD students from Middlesex University, London. The experiment session lasted one hour in average; the follow-up interview about five minutes.

5.3.4 Semi-controlled study with experts

We showed the visualisations to six police analysts and observed how they worked with them. We were interested if they could be used for realistic tasks of their domain, such as capturing harmful developments (R1) and problematic suspects (R2), or monitoring the effectiveness of mitigation strategies (R3). According to these requirements we developed three tasks (Table 5.4) in close co-operation with police analysts to ensure realism and generalisability. We trained the police analysts in the use of the visualisations and assisted them when questions arose during the tasks, which they worked on one at a time.

The study aimed to collect immediate feedback on the design concepts. We took notes and recorded the audio and screen capture of the sessions. After the task were solved, participants filled out a questionnaire with six open-end questions and a comment field to collect their feedback on the system.

Table 5.4: Overview of tasks for the police analysts addressing requirements 1-3.

Requirement	Task
R1	Identify harmful development (individual & group level)
R2	Identify connectors of subgroups under consideration of the domain information
R3	Identify if mitigation strategies of the past had any effect on the network

Tasks

In task 1 participants had to identify who seems to be causing growth of harm for the community through activity and crime type.

Task 2 focuses on the structural development of nodes and node similarities to increase understanding of local dynamics.

In task 3 we aim to identify if the inclusion of police force priorities helps in tackling the top networks problems, e.g., in the judgement of the overall network trend (i.e., if it is evolving to the better or not). This is usually a result of either active offender management or the problem not being perceived as a problem any more. In both cases this should be provided through indicators going down.

After this exploration part of the evaluation we asked five open-ended questions about applicability and usefulness of the visualisations. The experts were also asked for further improvements. The questions were:

- How helpful were the two visualisations for the different tasks and why?
- What challenges did they impose on your efforts?
- Are these types of visualisations useful for your work and why?
- What are relevant scenarios where they could be applied?
- What are possible enhancements in functionality?

5.3.5 Controlled study with novices

With the second target group we conducted a rigorous thinking aloud study of seven exploratory tasks, where participants were prompted to give specific answers. We conducted a full content analysis [Sch12] to see how the system was interacted with and to enable meaningful coding. In a mixed-methods approach we use thinking aloud protocols, observation notes, screen & audio recordings as well as retrospective cueing as a basis for a qualitative content analysis.

At the beginning of the study participants got a training task to get familiar with the visualisations. After that they got seven tasks which were recorded via screen captures including audio recordings. Furthermore, the experimenter noted the answers to each task, which were enforced after five to ten minutes at the latest. Finally, the experimenter conducted a short interview to collect preferences and feedback.

The experiment design allowed all participants to use both representations at all times and the time participants spent with a representation and how often they switched between them was tracked. This data was used for a qualitative content analysis. We performed a content analysis of thirteen thinking aloud protocols and derived nine cognitive strategies. We further categorised statements about problems, the task description and reasons for switching as meta-strategies, which could not be identified in a clear way as cognitive strategies. Around half of the protocols contained such meta-reports on how they try to solve a task and why they change to the other view.

Table 5.5: Overview of tasks used with the second target group ($N = 31$), T0 was used for the training, T1-T7 for the controlled study. The tasks address four research questions of the study.

Research question	Task	Description
	T0	Identify offender Nr.8 and its co-offenders, as well as their criminal activities over the shown time period.
Weights	T1	Identify co-offenders whose criminal activity increases over the shown time period.
Time	T2	Identify a group (3 or more) whose criminal activity increases or crime type worsens over the shown time period.
Relationships	T3	Identify connectors (actors that connect two clusters) between groups and their possible successors.
Relationships	T4	Identify a remarkably well connected offender that had many 2^{nd} degree connections in the last year.
Weights	T5	Identify two problematic crime types that cause most problems in the network.
Time	T6	What seems to be the overall tendency of the network in terms of crime activity and harm to community?
Visualisation	T7	What else did you notice in the network, are there any patterns or other interesting observations?

Tasks

The participants were given eight tasks of which the first was used for training (T0). The subsequent tasks addressed the research questions about weights (T1 & T5), temporal change (T2 & T6), relationships (T3 & T4) and the visualisations themselves (T7), compare Table 5.5. A mapping of the task to the research questions is given in Table 5.6. This overview shows that in some tasks both techniques yield plausible results and, therefore, both are valuable visualisations in the context of social network evolution.

5.3.6 Results from the quantitative analysis

The large sample size ($N = 31$) allows to statistically analyse the differences among groups via the analysis of variance (ANOVA). We analysed whether there is a relationship between the time spent on a task (T) and the visualisation (V) used: node-link diagram or matrix.

The results from the ANOVA analysis (compare Table 5.7) show a significant main effect of the visualisation [1]. Post-hoc tests using Bonferroni correction revealed differences between the matrix and the node-link diagram in tasks T2, T3, T5 and T7 ($p < .04$ in each case). There is also a significant interaction between visualisation and task in the

[1] F(1;336)=34,849; p<.0001, η_p^2=.053

Table 5.6: We quantitatively analyse the duration of the used visualisation and the quality of the results per task to show which visualisation caused more plausible answers per research question. The symbol * represents equally plausible results for node-link (NL) diagram and matrix (M).

	Relations	Weights	Time	NL vs. M
T1			*	
T2		M		
T3	NL			
T4	*			
T5		NL		
T6			M	
T7				*

times participants needed. Decomposing the interaction effect by visualisation confirms that differences are statistically significant between tasks T3×T1, tasks T3×T2, tasks T3×T4, as well as tasks T3×T5 for each visualisation (each $p < .021$). The participants preferred the node-link diagram and used it more than the matrix. The distribution on a task base is shown in Figure 5.12a.

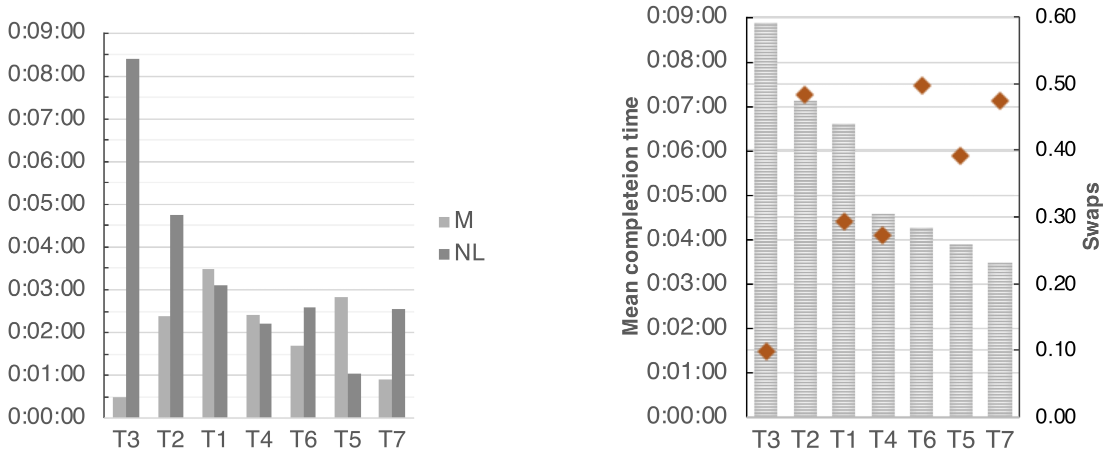
The main effect of the task was also significant [2]. Participants spent significantly more time with task T3 than with task T1 ($p < .007$). Tasks T2 and T3 were each significantly longer than tasks T4-T7, respectively ($p < .04$ in each case). We also tested whether the visualisation usage time T correlates with the number of swaps S. This is not the case ($\rho(T, S) = 0.006$).

We also analysed how often swaps between visualisations occurred to identify user preferences and which tasks are best solved with which visualisation. Participants swapped the visualisations least in the most time consuming task T3 and the most in the second most time consuming task T2, compare Figure 5.12b. The data of the randomly shown visualisation at the beginning, swaps and completion times per task, duration and preferred visualisation from the interview is given for all participants in Figure 5.13. The analysis of the visualisation usage is discussed in more detail below.

5.3.7 Results from the qualitative analysis

The qualitative analysis includes swapping behaviour between the two visualisations and preferred visualisation collected from interviews with 31 participants and the content analysis of thinking aloud protocols of 13 participants. The analysis of *emergent themes* led to ten sense-making strategies during tool use and five motivation themes. The data analysis approach was based on: (i) Familiarisation with the data, (ii) Coding, (iii) Identifying emergent themes, and (iv) Reviewing themes [WB02]. We observed five different motivations for user strategies: 1) to get an overview, 2) to gain new knowledge,

[2] $F(6;336)=11{,}320$; $p < .0001$, $\eta_p^2=.098$



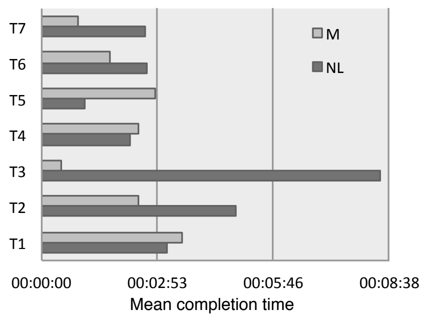
(a) (b)

Figure 5.12: a) Mean completion time per task in descending order of the total completion time. Overall, the NL diagram was used 60% of the time and the matrix 40%. Panel (a) shows the distribution per visualisation technique, and (b) the mean swaps per minute. Participants swapped least in task T3, with the longest completion time [DHGP+18].

Table 5.7: Summary table of the repeated measure ANOVA results (times) and means of completion times per task.

Source	**df**	**F**	η_p^2	**p**
V	1	34,849	.053	$< .0001*$
T	6	11,320	.098	$< .0001*$
$V \times T$	6	32,239	.296	$< .0001*$

V (visualisation), T (task), η_p^2 (partial eta squared). Significant differences in completion times are indicated by *.



3) to elaborate, 4) to improve certainty and 5) to create an overarching explanation. In our study we observed that they occur in both a chaotic and cyclic manner.

We used tasks relevant for intelligence analysis. Nevertheless, we think the strategies can be generalised to other domains. In general, the strategies we identified represent very complex behaviour, going beyond simple lookup tasks and the identification of data points.

On the one hand, participants adopted strategies to get an overview of the data, such as trend analysis or the analysis of *Increase in criminal severity*. On the other hand, they gained new knowledge by using strategies like *Pattern recognition*, *Relationships*

Start Vis	G	M	G	M	G	G	M	G	G	M	M	G	M	G	M	G	G	G	M	G	G	G	M	M	M	G	G	G	M	M	G
T1	0	1	0	0	0	2	4	4	1	0	2	0	2	1	2	2	5	1	4	3	1	3	1	1	0	2	1	3	1	7	3
T2	0	0	0	0	0	2	0	0	2	2	0	2	0	3	6	2	1	1	1	0	2	2	2	2	1	13	9	11	17	8	19
T3	0	0	0	0	0	0	0	0	1	1	2	0	0	0	0	1	0	0	2	0	1	2	4	0	0	2	0	0	1	6	4
T4	0	0	1	0	0	0	0	0	0	0	0	2	0	3	0	2	1	1	0	4	1	1	0	5	1	1	5	2	0	0	2
T5	0	0	0	0	1	0	0	0	1	1	0	2	0	0	1	1	2	0	2	4	2	3	1	4	4	2	1	5	1	7	3
T6	0	0	0	2	1	0	0	0	0	1	1	1	6	0	0	1	1	1	1	0	1	3	7	3	4	0	3	2	7	6	4
T7	0	0	1	0	2	0	0	1	0	1	1	1	0	2	0	1	1	7	1	1	6	1	0	1	9	4	6	3	1	1	2
Sum	**0**	**1**	**2**	**2**	**4**	**4**	**4**	**5**	**5**	**6**	**6**	**8**	**8**	**9**	**9**	**10**	**11**	**11**	**11**	**12**	**14**	**15**	**15**	**16**	**19**	**24**	**25**	**26**	**28**	**35**	**37**
Duration	49	48	48	15	53	43	61	35	57	25	49	48	45	24	38	42	45	45	43	40	44	37	36	48	42	51	35	40	44	50	48
Pref. Vis	G	G	G	G	G	T	M	G	T	T	T	G	T	G	G	G	M	M	T	G	G	G	G	G	T	T	G	G	G	T	M

Figure 5.13: Results used for the quantitative analysis in ascending order of swaps, $N = 31$: swap numbers per task (darker backgrounds represent higher values), visualisation participants started with (the visualisation, node-link diagram G or matrix M, was randomly chosen for task 1), duration of the whole session and user preference collected from the interview. The swapping behaviour does not correlate with the duration. The preferred visualisation was the node-link diagram, but many reported that it depended on the task (T).

and *Profiling*. As analysts interpreted the data, they elaborated their understanding and created new insights by *Comparing* indicator sets, *Laddering* from one information to the other and *Summarising* findings. To improve certainty, participants used *Elimination* and *Verification* when they moved from low uncertainty to high certainty. For an overarching explanation, participants adopted the strategy of *Storytelling*. These strategies took place in a chaotic and cyclic manner depending on the information available and the tasks required to gain cognitive traction. Although none of these strategies take place in a linear manner we observe an order of sense-making processes over time. Interactions between these sense-making strategies are shown in Figure 5.14. They take a fluid interaction based on the analysis.

We analysed the data of thirteen thinking aloud protocols systematically to investigate the participants' sense-making strategies in more detail. We identified the frequency of the involved strategies and whether they were applied in many different contexts or, rather, if they were only adopted for specific kinds of tasks. For this kind of analysis we made use of a count-coding system.

Coding depends on the data and the goal of the analysis. Our goal was to identify naturalistic sense-making strategies to inform visualisation design. We developed a coding scheme in two iterations to develop appropriate categories that fit the data. In the first iteration we used the results from the preliminary analysis and the Triple Path Model as a basis and created missing categories during the coding of the whole dataset of one participant. We then discussed the results to ensure that the categories are a) comprehensive, b) unambiguous and mutually exclusive and c) complete to a degree to describe sense-making during fulfilling the tasks of this study. This combination of a top-down and bottom-up data analysis process helped to identify users' sense-making strategies. The result of this discussion allowed to reduce from sixteen sense-making &

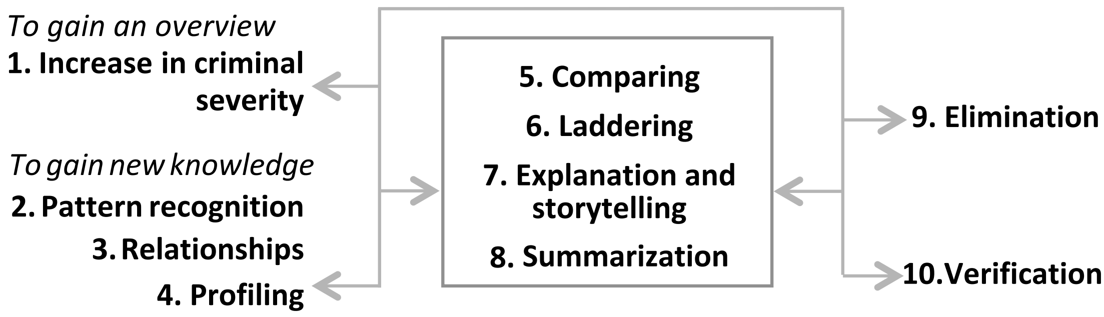


Figure 5.14: Interaction of strategies observed during the thinking aloud sessions. Participants started the analysis by gaining an overview of the data and by gaining new knowledge. New knowledge was created using strategies such as *Pattern recognition*, *Relationships* and *Profiling*. Depending on their understanding at the time they used *Comparison* and *Storytelling* to elaborate their understanding. Then participants used strategies such as *Elimination* and *Verification* when they wanted to move from low uncertainty to high certainty [DHSP+17].

meta strategies to the nine most relevant ones.

In the second iteration one experimenter coded the entire protocols of thirteen participants. The second experimenter coded one task per participant at random using the reduced coding scheme. This resulted in nine broad strategies featuring a high inter-coder agreement that emerged from the analysis (Cohen's Kappa $\kappa = .82$, compare Table 5.8).

In the following we describe the codes derived thereby in detail to capture the strategies participants used to gain insights. After that we relate the usage behaviour to the original assumptions on a task basis.

5.3.8 The observed sense-making strategies

We used Klein's Triple Path Model of Insight for the top down development of the coding scheme because it fits the setting and tasks. The main path of the model is the connection path, which we observed most often in our study as well. Other strategies, such as *Comparison*, *Coincidental Aha's*, *Creative desperation* and *Contradiction*, could be observed less frequently. They also occurred much less in Klein's research but, as Klein argues, are existentially different and, therefore, have the right to form their own pathways.

Participants used strategies such as looking for *Connections* and *Trends*, which are native to visualisations in general as visualisations provide a way to gain an overview of data and find patterns. Nevertheless, the applied strategies per visualisation show

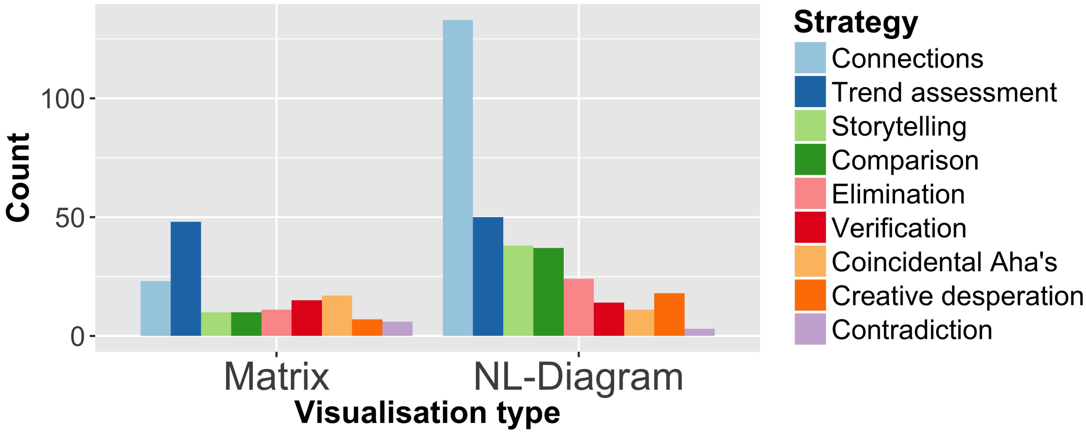


Figure 5.15: We identified nine sense-making strategies, which are in descending order of frequency: *Connections, Trends, Storytelling, Comparison, Elimination, Verification, Coincidental Aha's, Creative desperation and Contradiction.* Comparing their employment per visualisation type shows the predominant use of the node-link (NL) diagram, especially for looking at *connections* [DHGP$^+$18].

up tendencies of how the visualisations were used. An overview of the strategies per visualisation is given in Figure 5.15.

Table 5.8: The sense-making categories used to evaluate the second study, sorted in descending order of their frequency in the transcripts ($N = 13$).

Code	Description	Coding frequency	Agreement rate
Connections	Relations in the network	156	79%
Trend assessment	Temporal evolution of a person	98	75%
Storytelling	Explain behaviour of a person, give meaning to it	48	100%
Comparison	Comparison of persons or attributes	47	50%
Elimination	Elimination of elements from search space	35	77%
Verification	Verify with the other view	29	75%
Coincidental *Aha's*!	New, sudden ideas	28	100%
Creative desperation	Impasse situation	25	n.a.
Contradiction	Something is odd	9	n.a.

Connections

Due to the nature of our tasks *(connections of co-offenders)*, participants often looked at relations between actors and connections within the crime types that two offenders committed together. Direct and indirect (2nd degree) relationships were analysed, for example, a statement that was coded as *connection* is *"This is the man in the middle connecting these networks"*.

Trend assessment

Participants looked for multiple lines between co-offenders, in the node-link diagram, and the bar charts in the cells of the matrix, which represented the years 2013 to 2015. An ascending bar chart, for example, showed that the criminal activity is on the rise. In the multiple lines of the node-link diagram the colours get darker as the line style changes from dashed to a continuous line. Statements about the temporal encoding were coded accordingly, for example, *"Here I look at the lines to see whether they were active all the time"* or, *"I can see in 2013 they have not committed a crime, but it is increasing, if you look at 2014 and 2015"*.

Storytelling

Participants constructed a story to explain the behaviour of criminals, crimes, time intervals and relationships they observed within the data by using their experience or imagination. By adding new information they filled the gaps to create stories that *make sense* in the context of the available data. In that way they elaborated explanations to create bridges to a new understanding. Participants inspected an actor of interest, for example, and started looking at the number of 1st and 2nd degree connections to get an idea whether the offender works with few or many criminals. They further looked at crime types to get an idea of what kind of crimes the actor might commit in the future, if the criminal activity was high or low in the past and how recently crimes were committed. Time intervals and other associations to the neighbour's criminal activity were studied to see if offender pairs were rather active in the past or just started their crime career recently. The factual interpretation by the participants was an enrichment that likely stemmed from training, previous experiences, but also their creativity. Some participants were using this strategy extensively with great imagination, whereas others did not use storytelling at all and stuck to the presented facts. Speculative statements were coded as storytelling as well, such as *"They are committing crimes over all these years, so I guess they must be very good friends or successful because they didn't get caught"*.

Comparison

Comparison was used to identify similarities or differences between actors of the network (offenders) depending on the task. One participant who previously identified a set of actors, for example, consequently compared the group with another set to assess their activity and see if they operate in similar ways. Statements about comparison were coded when the behaviour and verbalisation was obvious, e.g., *"Here we have Financial Fraud ... here they did something else..."*.

Elimination

Filtering the search space by eliminating data that was considered as irrelevant generated a new understanding of a situation as part of the solution of a task. First the participant identified possible answers and then eliminated those that do not fit the requirements. In that way participants eliminated mismatches from the search space quickly and did not have to considered them any more. The search pattern was often specified through elimination, such as, *"Here I have to search for blue lines, they were active in one year only"*.

Verification to conclude

Participants consulted both representations for verification when they came to the end of a task. To infer how long an offender was active, for example, participants looked up multiple lines in the node-link diagram (line style and colour) and verified their findings by switching to the matrix and its bar charts. They stated, e.g., *"I am going to the matrix as it can show the crime times in the visualisation. The matrix confirms my discovery"*. Participants also re-validated their findings within one visualisation by looking again more closely. Examples: *"So let me check again", "Did I forget something?"*.

Creative desperation

Experiencing an impasse is also part of the Triple Path Model developed by Klein [Kle13]. It describes the situation of not knowing what to do next and the feeling of being stuck. We could clearly observe such situations where a strategy change needed to happen. It typically took the participants some time to escape the impasse by changing the direction and following a new idea. Statements that were coded for this creative desperation are, e.g., *"I have no clue", "nothing sticks out here", "too little information", "too complex", "I don't find anybody"*.

Coincidental Aha's

New, sudden ideas indicated a spontaneous comprehension of a problem and the emergence of a solution or an insight. In the literature this is often referred to as Aha-moments. Coincidences are coded as such if it was not clear to the observer *why* an action was taken, or where a new idea came from, i.e., previous statements in the verbal protocols do not explain the insight. When a new insight occurs and the origin is untraceable, the statement is coded as coincidental or even if "Aha" is used for the spontaneous insight, e.g., *"Aha ...", "I think this could be...", "Now here I found that..."*.

Contradiction

In contrast to the *Coincidental Aha*, here contradictory information could be observed as the reason for an insight. However, the experience of "Something's odd" could be observed least often. It also forms a category in the Triple Path Model, which describes that new insights can be triggered by contradicting information. Obvious mismatches in what is hypothesised or thought of being true, however, were observed rarely. The few examples are, e.g., when participants expressed their wonders and new realisations, such as, *"Mhhh, this is different here... [wondering voice]"* or *"No, this can't be right"*, *"There aren't as many here anymore [surprised voice]"*.

5.3.9 Usage of the visualisations

Understanding relationships played a role in several tasks. We assumed that relationships would be detected more easily using the node-link diagram. Indeed, the node-link diagram was used more often (37 times) than the matrix (14 times) to find well connected actors (and possible successors in task T3). This required a detailed analysis of the offences that two persons were convicted for. Most frequently actor 18 was named as a prominent connector, who could be substituted by actor 8 (10 times). Another plausible answer was actor 44 with actor 18 being a possible successor (9 times).

Especially in task T4 we considered the node-link diagram to be superior as this type of representation is supposed to depict structure better. However, for this task the matrix was used more often (39 times matrix, graph 30 times). We interpret this effect to be stemming from the special encoding of the 2nd degree relations in the matrix, which they remembered from the training. One participant recalled *"The position of the yellow bars again represents the year"* and formulated a pattern she could look for. Another participant concluded *"this means I have to look for dark yellow bars in the rightmost part of the cells"* using the matrix to look for an answer. Participants using the node-link diagram reasoned that they had to look for a neighbour of very well connected actors who, in addition, have dark, continuous lines because of the latest year (which was asked for). Here, they first explored the large central group, but had difficulties to decide on one answer and eventually switched to the matrix. Problems became apparent in identifying distinct connections due to overlapping edges. Another problem with the node-link diagram was to discriminate between the years. Although participants started to look for dark, continuous lines for the most recent year, for example, two clearly forgot about the pattern they were looking for and continued looking at all relations (disregarding the colour and line style).

More participants used the matrix to identify relations with growing trend (task T1), which confirms the assumption that the evolution of co-offenders is easier to inspect in the tabular overview. However, some switched to the node-link diagram in between to confirm their findings. There were also participants who started with the node-link diagram right away and did not bother to change to the matrix at all. Participants who solely used the node-link diagram looked for a pattern of parallel lines that become thicker and many focused on the continuous blue lines, i.e., on showing a relation in the latest year. There were seven co-offender pairs that could be named to provide a valid answer. We asked for an unspecified number, so participants tended to search the whole dataset. For many the main central group constituted an ambiguous case as it does not show a linear upward trend (rather an increase followed by a decrease). Especially this group was explored often in both the matrix and the node-link diagram. In sum, both visualisations were used 43 times.

In general, we assumed that the trend is depicted in both visualisations equally well. The participants' usage in judging the network's trend (task T6) validated this as it was distributed more or less equally (39/41 times). The number of crimes decreased over the years in the depicted evolution. 18 participants perceived the decreasing trend correctly. Only two participant saw an increase in the crime evolution, one suggested a steady trend. The remaining ten participants were undecided. Two participants said that they could only guess because it is too hard with either visualisation.

Participants looked for *"fat lines"* and if there were more of them in the node-link diagram. Interestingly, only one person could make a sound judgement when looking for that pattern in the node-link diagram. Six participants did not correctly perceive the lightly coloured (older) links and concluded that there were more crimes committed now in the latest year. They also investigated the strongly connected group in the centre but after spending some time on this group, they switched to the matrix to analyse it further. Two persons then changed their mind about their answer. Consequently one person, for example, stated *"I realise that I got the wrong impression with the graph [node-link diagram]"*. Another person got more insecure about the answer after taking into account both visualisations, stating that *"the crime seems not to become more or less but rather shift in the type of crime"*.

Six participants additionally analysed the change in the crime type. One participant identified a shift from financial crimes to more violent crimes and inferred an increased trend in terms of the seriousness of the crime. The criminal activity could be analysed in detail using the bar charts in both visualisations. Here, we assumed that this was easier in the matrix because the bar charts are immediately visible in the overview. Co-offenders who performed a shift in their crime type (task T2) were looked up in both visualisations (using the node-link diagram 67 times, and almost as often using the matrix 62 times). As people tried to match pairs found in the node-link diagram to the matrix and vice versa for reassurance, this task led to the most frequent switching between the two visualisations. Here, participants found it difficult to memorise identifiers in order to find offenders in the other visualisation, necessitating frequent switching. This resulted in short visits in one visualisation (1-5 seconds) and a high swapping number ($\tilde{0}.5$ per minute; AVG= 0.35), compare Figure 5.12b. In the node-link diagram, criminal activity had to be investigated by hovering and, therefore, participants liked the matrix better as it provides an overview over the crime type evolution. Reasons for switching were *"finding groups is easier in the graph [node-link diagram]"* and *"I can see the crime types better in the matrix"*. Searching for an ongoing criminal activity over the three years in the node-link diagram was done by looking for multiple lines. One participant, for example, started with the matrix, but was unsure if the identified co-offenders she found would build a group and, therefore, switched to the node-link diagram to cross-check the connections.

Looking for dominant crime types (task T5) is presumably easier to answer with the matrix and our results confirmed this. The matrix was used 39 times to solve this task and participants only consulted the node-link diagram half of the time (27 times). Every participant ranked "Financial Fraud" as the most dominant answer, which was a correct choice (rank 1). This was a clear case as both frequency and weights were highest for this type of crime. "Disorder violence", the second most dominant crime, was only identified by 19%. 58% rather chose "Minor Assault" (rank 4) as more dominant, which appears more frequently but has a lower assigned weight than both "Disorder Violence" (rank 2) and "Burglary Other" (rank 5). Five participants chose "Burglary Other" as more dominant overall. The worst assessment was done by the one participant who only used the node-link diagram, missing nine more dominant crime types by reporting "Criminal Damage" (rank 11). The high weight of "Drugs Supply" was not perceived as nobody reported this type of crime as dominant although its weighted score reached rank 3. Hence, the frequency of crimes had a bigger impact than the weights, since having the highest weight (the same as "Financial Fraud") did not compensate for the fact that it occurred six times less often than "Minor Assault". Another reason for not seeing "Drugs Supply" as a dominant problem is the decreasing trend over the years, up to no occurrences in the last time frame. Using the matrix participants acknowledged that this type of crime was causing problems, but not any more, which exemplifies that the matrix gave a good overview over the network development.

5.3.10 Strategy usage per visualisation

Overall the most frequent strategy was *Looking for connections*. Participants mainly applied this strategy in the node-link diagram. *Trends* were assessed in both visualisations at the same rate, as well as the *Verification* and *Contradiction* strategies, which were used about equally often. *Storytelling, Comparison, Elimination,* but also *Creative desperation* occurred more often in the node-link diagram. *Contradictions*, on the other hand, occurred twice as much in the matrix than in the node-link diagram.

Looking for connections was used in every task but dominantly in the relational tasks (tasks T3 and T4). Participants could only apply this strategy in single cases using the matrix and predominantly applied it in the node-link diagram. *Trends* were more often assessed in the matrix in tasks asking for temporal developments (tasks T1 and T6). However, in tasks regarding weighted crime activities, *Trends* were assessed as well and here the usage was the other way around, i.e., the node-link diagram was used more often (compare RQ2 in Figure 5.16).

Storytelling was the third most frequent strategy employed overall and much more often used in the node-link diagram. There it notably occurred in all tasks, but least in those regarding time, whereas in the matrix *Storytelling* was used most frequently for the temporal tasks (compare RQ3 in Figure 5.16).

Overall participants *verified* results only once more in the matrix than in the node-link diagram. In the node-link diagram weights were mostly verified and in the matrix verification was mainly applied for temporal aspects (tasks T1 & T6).

Participants *compared* results more often in the node-link diagram than in the matrix, especially in the tasks regarding relationships and weights. In the matrix answers to time related tasks got compared most often.

Participants *eliminated* information to a similar extent in the matrix and in the node-link diagram. On a task basis *Elimination* was also used as often in both views for time related tasks (tasks T1 & T6). In all other tasks it was dominantly used in the node-link diagram.

Creative desperation happened in all tasks in the node-link diagram, where it was observed more often overall. Primarily when looking for relations and weights participants expressed impasse situations. In the matrix participants could avoid them in the relational tasks (likely because they barely used the matrix for those tasks). Reaching an impasse situation happened more often with regard to weights than to time.

Coincidental Aha-moments occurred more often in the matrix in every task. In the node-link diagram they could be observed mostly with regard to weights and time, but in the matrix they occurred in relational tasks as well.

Participants had realisations because of *contradictory* information in the matrix only in time related tasks and when talking about the visualisation capabilities (task T7). Interestingly, no contradictions occurred in the node-link diagram for temporal tasks, i.e., the encoding in the lines, compare Figure 5.16.

5.3.11 Conclusion

We analysed sense-making strategies during the use of two visualisations and compared their application on the basis of the visualisations and tasks. The tasks can be categorised into 4 groups: time, relationships, weights and visualisations capabilities themselves. Two strategies concern insight generation: *Coincidental Aha's* are unexpected insights, and *contradictions* often led to insights, but also to creative desperation, which on the other hand did not directly lead to insight but to a change of strategy.

We could show that matrices are superior for representing temporal development. In general, our results do not agree with previous results from literature, which suggests that matrices are best used for dense networks and effective overviews, and node-link diagrams shall be used to show structure and hierarchies. Our results suggest that the matrix is also useful for structural tasks and that the additional encoding of extended relations of 2nd degrees is beneficial. We further show that the combination of both visualisations is beneficial in certain tasks, especially for verification purposes. When asking for 2nd degree relationships the matrix was preferred, possibly due to the overlapping of edges and the lack of a uniform layout in the node-link diagram.

We assumed that the identification of crime weights would be easier with the matrix as it shows crime types at a glance, whereas participants have to hover over relations to see them in the node-link diagram and have to memorise this information while investigating different links. Even those participants who were keen to use the node-link diagram claimed that the matrix provides the better overview in this use case. The matrix provided a better overall assessment of temporal developments of criminal activity,

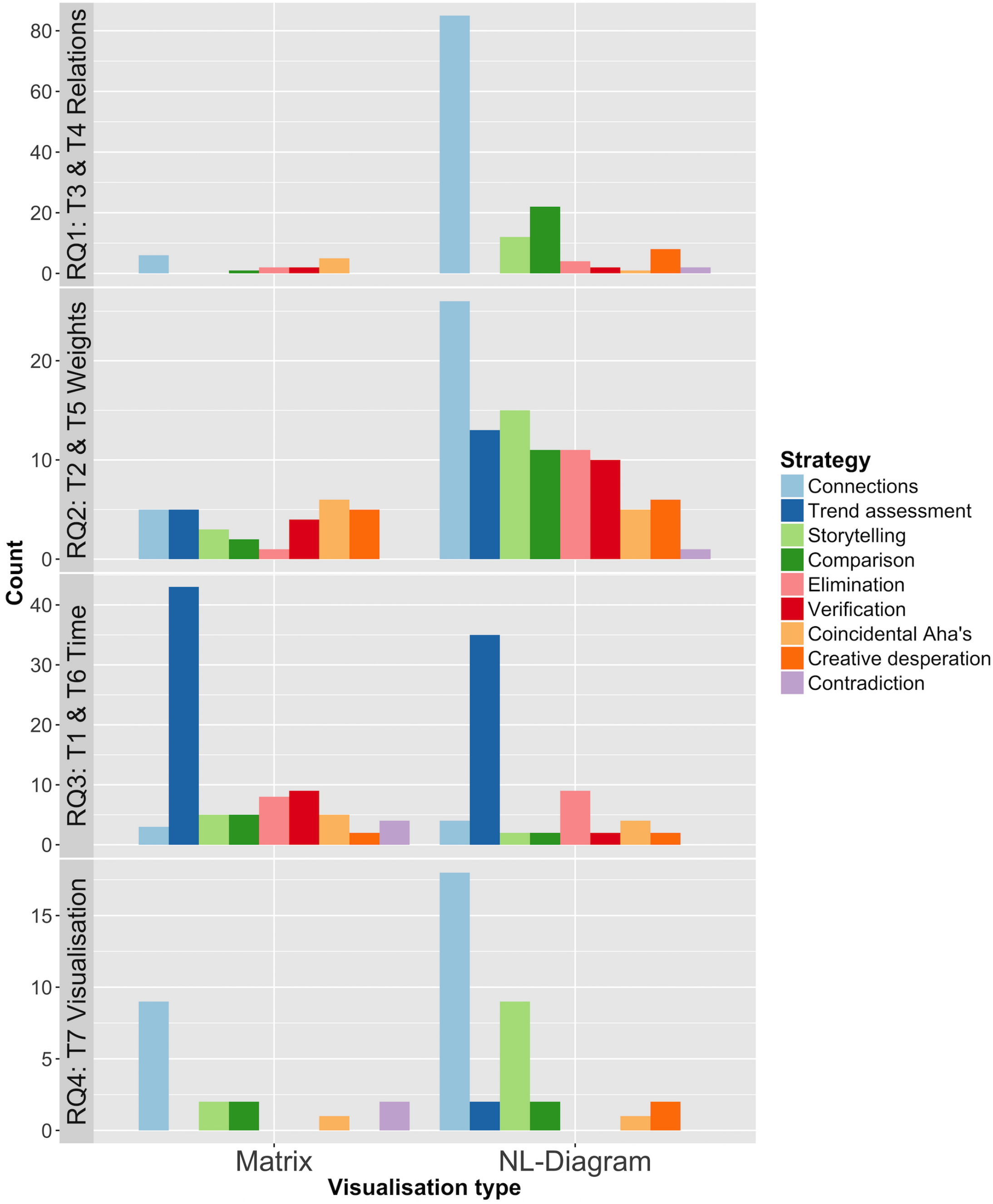


Figure 5.16: The frequency of the nine sense-making strategies, which we observed in the experimental setup shown per research question RQ1-4, for both the matrix and node-link (NL) diagram. We used two tasks for each research question. RQ1 Relations gets addressed by tasks T3 & T4, RQ2 Weights by tasks T2 & T5, RQ3 Time by tasks T1 & T6 and RQ4 Visualisation by task T7.

but the usage in this task was shared equally between both visualisations. It was possible to retrieve an overall impression of development in the network through the node-link diagram, however, this sometimes misled participants to concentrate on the thicker and continuous lines and to underestimate the previous time periods due to the dashed line style.

A key result is that although participants had no prior knowledge of the two visualisations, they were able to assess the views quickly and decide which was better fitting to a task. The results show clear tendencies in four out of seven tasks. In tasks regarding individual relationships and the overall trend the advantages of both visualisations were leveraged and there were no clear favourites. Finally, individual results show that the subjective preference was hindering the user in swapping views and preferring the node-link diagram led to a different insight in one task than using either only the matrix or both views. This strongly argues for providing both visualisation techniques and training the intelligence analysts in using several visualisations for advanced insight generation.

To conclude, this investigation contributes in the following ways:

- Foremost, a novel visual analytics approach adapted to the requirements of criminal intelligence gets introduced. We represent a weighted network over 3 years within two established techniques, a node-link and a matrix based visualisation.
- Furthermore, we describe a set of sense-making strategies to generate insights while using two different visualisations for the same tasks, which can be used for further studies with similarly structured data.
- Finally, we find that using indicators for 2$^{\text{nd}}$ degree neighbours in an adjacency matrix proved to be more precise than the analysis of this relationship in the structure-inherent node-link diagram.

5.3.12 Comments on dissemination

The two visualisations were designed in co-operation with A E Solutions (BI) Ltd based on the requirement analysis within the VALCRI project and state-of-the-art literature. My contribution was in planning and conducting the studies in co-operation with the co-authors.

Three papers were published on this investigation:

- First we describe the technical point of view of strategic intelligence analysis and the focused target group in study 1 at the European Intelligence and Security Informatics Conference 2016 [SHK$^+$16]. Here we contributed mainly in the evaluation and discussion of the results.
- Second, we wrote a short article with the viewpoint on sense-making strategies *"How Analysts Think: Sense-making Strategies in the Analysis of Temporal Evolution and Criminal Network Structures and Activities"* [DHSP$^+$17], which appeared in the proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting in

2017. In this first evaluation of the second study of the investigation we collaborated with Middlesex University.

- Third, we followed-up with an extended qualitative analysis in 2018, which was done in collaboration with a diploma student at Vienna University of Technology. We wrote a journal article about the results of this in-depth analysis, entitled *Sense-making Strategies in Explorative Intelligence Analysis of Network Evolutions* [DHGP+18] for a special issue on human centred visual analytics in the Journal of Behaviour & Information Technology. Here my contribution as the main author lie in planning and structuring the analysis, as well as coordinating the discussion of the results and the input of the co-authors.

5.4 Investigation 4: How analysts think and form hypotheses with visual analytics

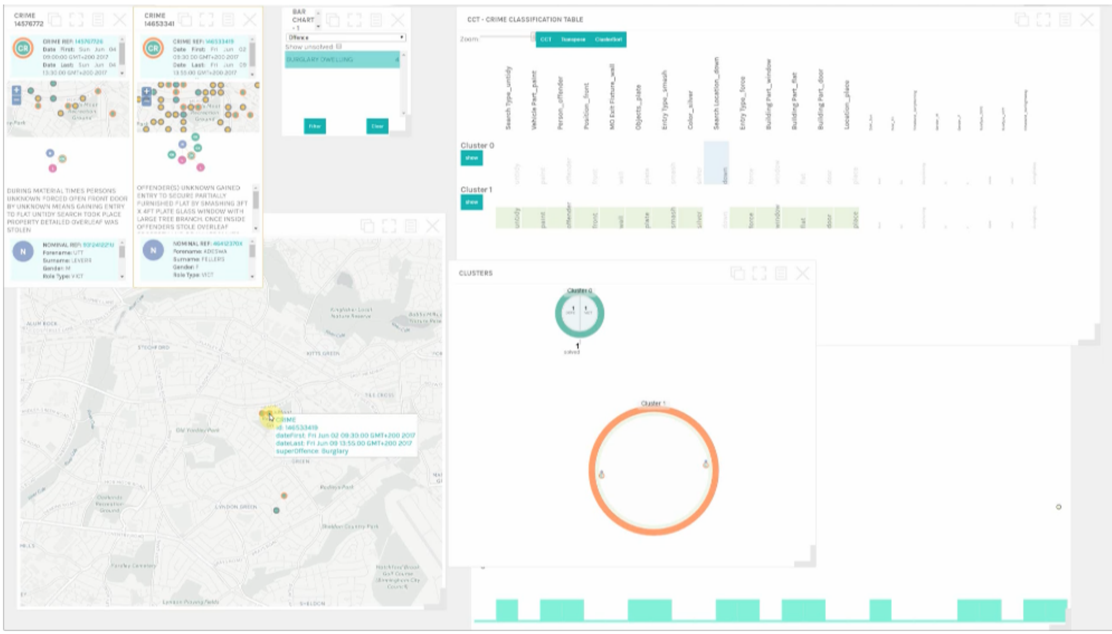


Figure 5.17: Example of a large canvas as provided in VALCRI. Several views may overlap each other as they can be manually repositioned and resized (a full-screen option is included as well).

The goal of this study is, similar to the previous investigation, to better understand and report sense-making aspects of analysts during realistic tasks. Here, we make use of a complex visual analytics system which was designed to support and augment the intelligence analysis process: VALCRI. The system uses tiles to offer various visualisations that can be arranged flexibly on a large canvas. The formative evaluation process of the VALCRI project ensured the usefulness by evaluating prototypes of the system incrementally with police analysts to be able to adapt the prototype along the way. In the course of this process we evaluated the integrated prototype with regard to sense-making in June 2017, which led to suggestions for improvements as well as general design recommendations. In addition, we again analysed the sense-making processes of 18 graduate students at Vienna University of Technology.

Table 5.9: Design recommendations based on investigation I4.

Recommendation		Description
4.1	Offer simple visualisations	Offer simple visualisations such as maps, timelines and bar charts. These well known visualisations were preferred to more complex visualisations in our study. We observed many combinations of the simple visualisations with the more specific ones like, e.g., Statistical Process Chart (SPC), Crime Classification Table (CCT) or Space Similarity Selector (S3).
4.2	Support *Verification*	The verification of results promotes high-level insights. By assessing the quality of insights we observed differences between high-level insights, including inferences rather than just single facts, which we refer to as low-level insights.
4.3	Motivate *Trend Analysis*	Participants who focused on analysing trends tended to report better insights.
4.4	Motivate complementing views	The combination of an interactive timeline and a SPC, for example, may be used in multiple views.
4.5	Emphasise training of the tool use	Our results showed that participants who focused on fewer strategies gained more insights. Knowing the tool and what to expect from a visualisation might help to focus on certain strategies. Enable focused work within a visualisation.

5.4.1 Motivation

Intelligence analysis suggests that over time similar MO of crimes can be observed repeatedly in certain areas. Hence, by purposely fusing multiple data sources to enable immediate detection of known crime patterns, visual analytics systems can facilitate criminal intelligence analysis and thereby influence decision-making in policing.

The main challenge is that the sense-making processes of analysts during the use of visual analytics systems are not fully understood, which impedes the implementation of appropriate support mechanisms. Therefore, we are interested in the following research questions:

- RQ1: How do users generate insights with visualisations for intelligence analysis and which visualisation tools do they use to achieve this?

- RQ2: Which sense-making strategies do participants use most?
- RQ3: Are sense-making strategies related to the number and/or quality of insights?

VALCRI encourages the user's intuition based upon the available data, her analysis goal at the time and her experience. The underlying idea is that requiring minimal effort results in low commitment at the early stages and, thus, supports the creative, playful, generative, chaotic and tentative nature of human reasoning. The system also enables formal assessments by allowing the analyst to test her initial claims, by smoothly transitioning into formal methods. This is necessary to support claims made through evidence.

5.4.2 Description of the prototype

The prototype consists of three main parts. First, a search allows to filter the extensive anonymised police database to a manageable set of entities that can be explored during analysis. Secondly, a clustering algorithm extracts possible connections within the whole dataset and gives insight on cluster characteristics. Finally, a SPC reveals outliers over time and supports the temporal analysis of the database. This visualisation points to noteworthy time ranges when crime rates differ from the usual crime behaviour, which can be used again as an input to filter the results via a search function.

VALCRI starts the analysis with a search, which consists of a text field for a keyword search and a date range selector, shown in Figure 5.18. After running the search with the keyword *school* three linked visualisations are used to represent the vast set of results: a timeline, a map and categorical terms in a bar chart. They are linked, which means that each visualisation acts as a potential filter on its own.

The time view shows the period of time that the crimes on the result set were committed in a timeline – an area timeline can be used to select a subset of that time frame. The map shows crime event locations aggregated with an overlaying grid for result sets exceeding 300 events, and the bar chart shows the distribution of different characteristics, e.g., the crime type (default), or the categorical term of the town, compare Figure 5.18. It further allows splitting results between custom attribute overlays, for example, for solved or unsolved crimes.

The analyst can start several searches in parallel to compare different result groups. They can be distinguished by a yellow border that shows connected search results. Closing the respective search bar closes all corresponding views.

Each visualisation acts as a filter for the result group, e.g., selecting an offence type in the bar chart reduces the result set to the respective offences. When the resulting number falls below 300 each entity location is shown as a single dot in the map. Choosing a time frame in the area timeline applies a gradient colour-coding on the dots in the map to show temporal uncertainty, compare Figure 5.19.

Intelligence specific tools are the CCT, the SPC and the S3 with an additional cluster view.

The SPC shows the mean crime rate of a selected time period and its standard deviation, facilitating the detection of anomalies in the data, such as very large or sudden

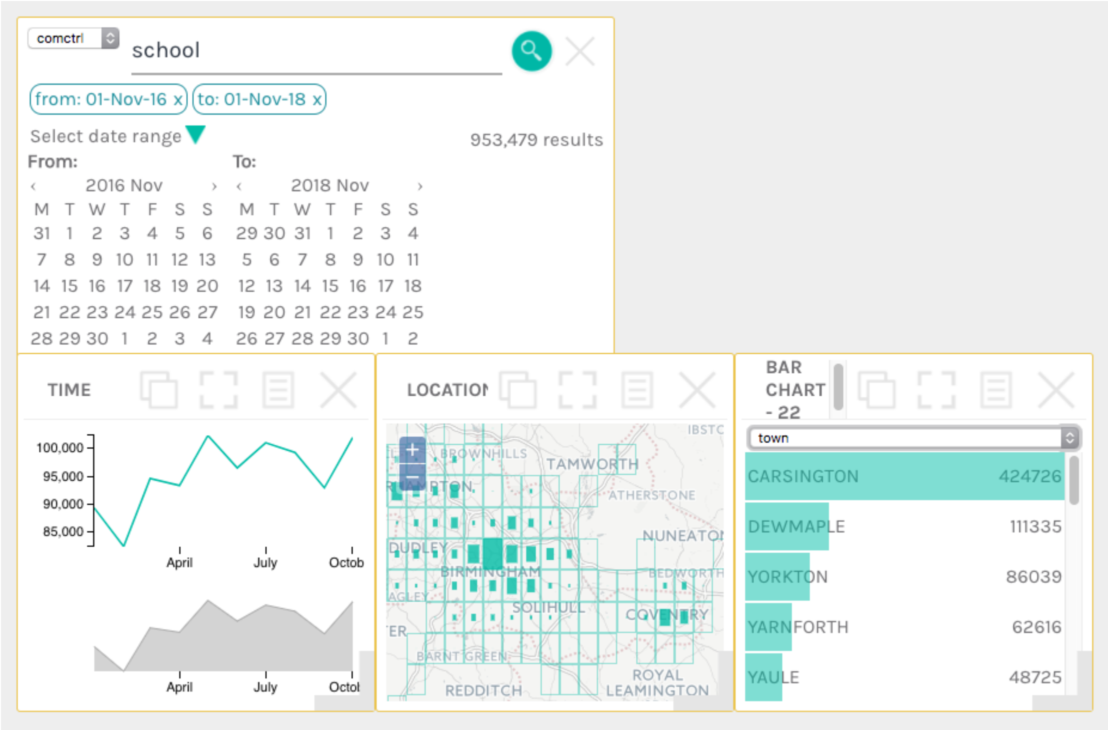


Figure 5.18: Screenshot of the VALCRI prototype, showing the components an analyst starts with: a search bar with keyword search and a date range selector. The results are shown in three basic visualisations: a timeline, a map and a bar chart. The bar chart may represent different categorical terms, such as offence type, day of the week or time in the day – here, the towns where the crimes happened are shown.

shifts. Intelligence analysts would interpret various types of deviation from the mean over consecutive points typically as a new trend, which is encoded as turquoise points in the SPC. Outliers are highlighted with red crosses (see Figure 5.20).

The CCT summarises all the information from the MO description of crime reports, which consistently contain the time band and a role type, i.e., victim or defendant, showing all features in columns and crimes per row. Analysts may then make use of a clustering tool and visualise them via the S3. S3 groups crimes into clusters based on their similarities allowing to set parameters for the k-means algorithm, such as the number of clusters. Analysts looking for crime series are advised to look for small clusters containing six to ten crimes in order to perform victimology analysis (determining age, gender, etc.). Each crime is represented by a dot, where the distance to others represents its similarity - closely grouped dots show a strong similarity of their crime reports. In Figure 5.21 we can see overlapping dots in cluster 3, which means that there are very similar crimes.

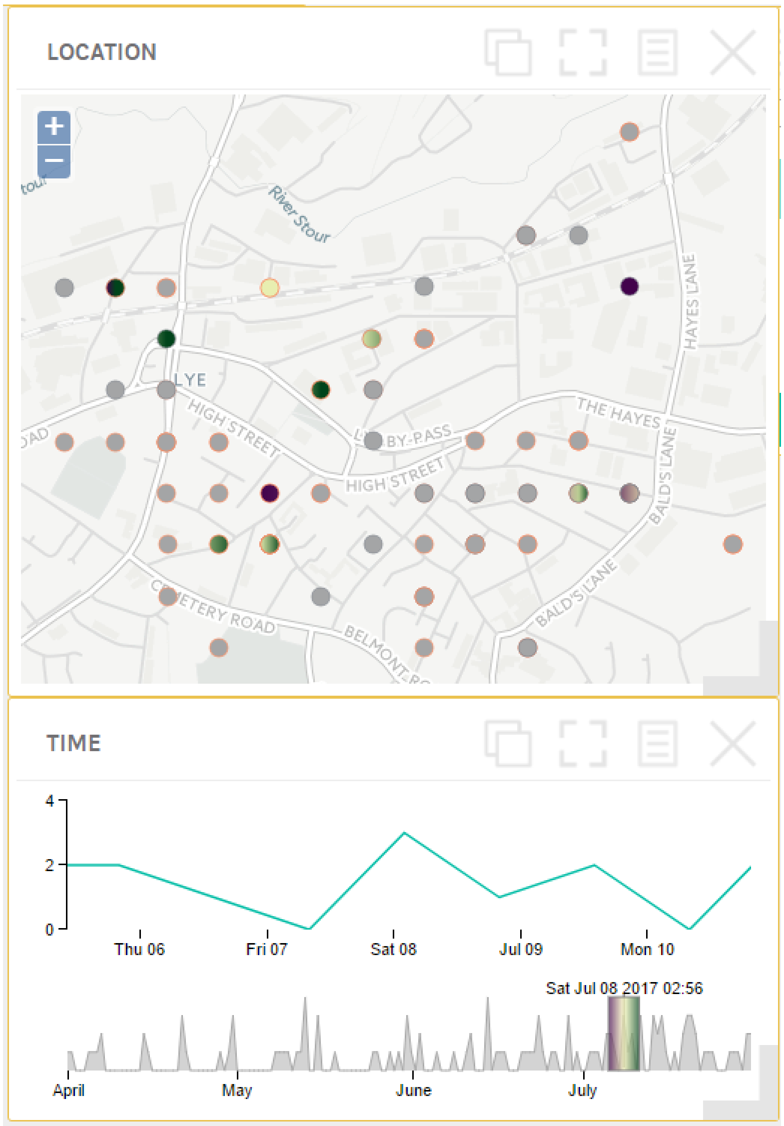


Figure 5.19: Single crimes are shown in the map if the result set is smaller than 300 events. Colour coded dots represent the time the crimes happened in relation to the time selected in the timeline (before and after the selected timestamp).

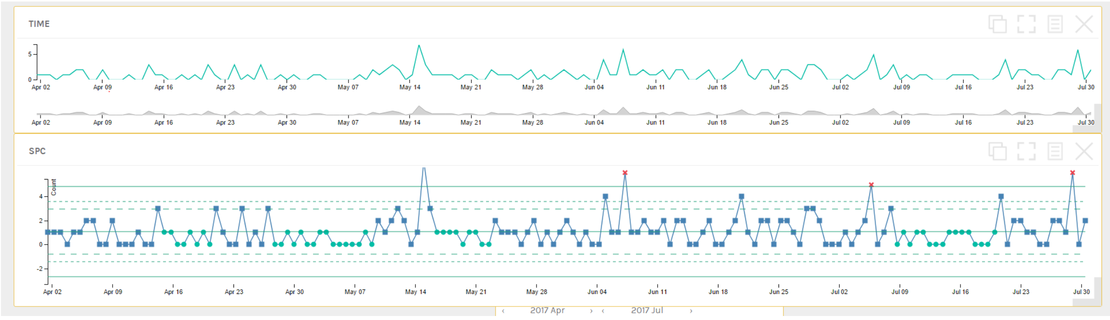


Figure 5.20: Crime rates over time in the basic timeline (top panel) and an SPC (bottom panel): the date is mapped on the x-axis, the counts are mapped on the y-axis. Mean and standard deviation are shown as horizontal lines in the SPC.

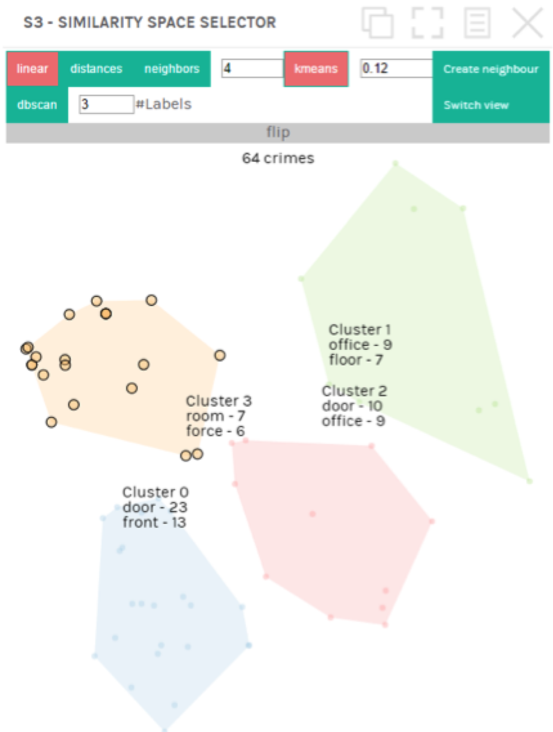


Figure 5.21: The S3 visualisation allows to specify clusters based on keywords from the MO description. Crimes are represented by dots and the distance between dots represents similarity. Dots that are overlapping, therefore, represent events with matching textual descriptions, such as occur in cluster 3 where "room" and "force" are the main common keywords.

5.4.3 Approach

We analyse the interaction with ten visualisations within a prototype of VALCRI:

1. Search bar
2. Timeline
3. Map
4. Bar chart
5. Entity card (Crime report)
6. List of crime reports
7. Statistical Process Chart (SPC)
8. Space Similarity Selector (S3)
9. Clusters
10. Crime Classification Table (CCT)

We analyse the use of a multi-component visualisation with regard to insight and sense-making by extracting strategies employed during the work with the system. Hence, we observe how the participants make sense of information provided by the visual analytics tool. We expect participants to use different components to answer different kinds of questions and follow up on investigative hypotheses during an analytical task.

To address these questions we conducted an extensive experimental thinking aloud recruiting eighteen computer science students (5 female, 13 male) with a Bachelor degree in computer science. We trained them on basic policing principles, as experts are not available at this scale. No colour vision impairments were reported. We trained the participants about strategic analysis goals and introduced the system in a walk-through session and a video tutorial explaining the context and aims of the police work. A user manual was provided for studying the system prior to the study session. Furthermore, we started the study with a hands-on training task to establish a feeling for the tool and to answer any questions.

Tasks: Making sense of crime in specific areas

The tasks were developed in cooperation with end users from police forces. They reflect realistic application scenarios and are adapted to the dataset.

- Training task: Search for knife crimes in Birmingham during February 2017 and investigate the information in the given visualisations.

- Analysis task: Burglary dwelling in the town of Tormington - Analyse offences that occurred between 1 April and 31 July 2017 and prepare a short report that summarises your findings.

The task description of the analysis task further includes dates relevant to the analysis (bank and school holidays) as well as the goals to be able to improve surveillance, allocate the appropriate resources and apprehend the suspect(s). The report should include insights, an assessment of relevant crime data by time, location, and MO. We think that the task implicitly defines the insights that should be the outcome. This makes it more obvious how insights can be measured. In our case, an insight is a recommendation accompanied by one or two screenshots showing the visualisations that helped the participants to get that insight. This screenshot was in most cases accompanied by text explaining the insight and/or the recommendation.

5.4.4 Results

The analysis of sense-making strategies during the task reveals differences in the insight generation process and the frequency and duration of applied strategies, which addresses research question 1 and 2.

We adapted the coding scheme from the previous study to align the level of the strategies and get clear-cut definitions for mutually exclusive strategies. In the extensive investigation of an explorative visual analysis task it is not feasible to observe lower level activities such as *Comparison, Laddering and Summarising.* With *Comparison* we experienced that it was used within one visualisation as well as beyond one tool by memorising the information to compare it with the information given by another tool. We were able to detect this due to the experimental setting of study 3, where participants could only see one visualisation at a time, but switch whenever they wanted. Here we use a multi-component prototype with a focus on higher level activities such as *Pattern analysis, Trend analysis* or *Storytelling.* We generalised the main category of the network analysis in investigation 3, *Connections*, to *Patterns* for the analysis of spatiotemporal-thematic event data in the present investigation. A lesson learned from the previous investigation was that mutually exclusive categories enable straightforward coding and evaluation processes. As it was sometimes hard to assign codes unambiguously we now decided to add hybrid codes for these cases. This resulted in the two combined codes *Pattern with Profiling* and *Elimination with Trend.* We think the resulting set of codes is generalisable to other domains using multivariate event data.

A detailed description of the sense-making strategies including coding criteria and exemplary statements is provided in Table 5.10.

Table 5.10: The coding scheme for the structured analysis of thinking aloud protocols including criteria and exemplary quotes.

Category	Criteria	Example
Pattern: Looking for similarities across several (groups of) actors between crime types, criminals time intervals, direct and indirect relationships, etc.	– Considering various characteristics of several, at least two, individuals or groups of actors or time spans – Mentioning patterns (i.e., similarities) across dimensions	**P4:** "So what I am interested in is how they got in, window or door [looking at clusters in CCT]." **P11:** "I can create a neighbour [map from list] and compare the locations [of clusters]."
Profiling: Characterising crimes or criminals based on features such as links (identify gangs), crime types (what type of crimes is most critical), or time intervals (when and why actors are active).	– Inspecting specific individuals or groups of actors or crime – Considering various aspects about the suspects, going into detail - may contain trend assessment	**P16:** "I have a crime with a victim, I will have a look at [the victim], maybe he was even mugged twice." **P18:** "I will read the description [of a crime report]"
Pattern with Profiling: Combination of Pattern and Profiling.	– Comparing crime reports with each other – Activities where it is hard to separate *Pattern* from *Profiling*.	**P18:** [putting several crime reports next to each other]
Trend: Looking for trends in the data. Attempting to determine increasing or decreasing criminal activity over time.	– Inspecting crime rate changes in timeline – Comparing development across time	**P2:** "In this time period it [crime rate] goes up."
Elimination: Generating new understanding by eliminating data considered as not relevant.	– Filtering out data that does not fit or is not interesting	**P10:** "I will select burglary dwelling in the bar chart [which filters all crime events by this crime type]."

Continued on next page

Table 5.10 – *Continued from previous page*

Category	Criteria	Example
Elimination with Trend: Reducing the search space with respect to time.	– Reducing possibilities that do not fit a temporal requirement	**P10:** "I will look at 16th July."
Explanation and Storytelling: Constructing a story by explaining the MO and relationships that were observed within the data by using one's experience or imagination, thus, adding information that subjectively make sense.	– After profiling (likely) – Added information that is not visibly obtained from the data – "Made-up" or subjectively enhanced information depending on the previous life experience of the analysts	**P1:** "Christmas days are less affected, I assume criminals have a family too." **P2:** "People come home from the Easter holidays, kids go back to school and parents work and then we have all the dwellings."
Verification: Consulting both representations for verification purposes.	– Looking up information in another visualisation – Validate an assumption	**P1:** "I suppose I can see this here [in the other visualisation] as well."
Creative Desperation: Not knowing what to do next and the feeling of being stuck at an impasse.	– Expressing that one is stuck and doesn't know how to continue	**P4:** "Now I am not sure of what best to look for."
Coincidental Aha's: If it is not clear to the observer why an action was taken, or where a new idea came from.	– Seemingly coincidental insights – No clear line of thought.	**P3:** "Ah! That's a good cluster." **P13:** "Ah, that's interesting."
Contradiction: Obvious mismatch of what was hypothesized or thought of being true.	Noticing that previous thoughts or assumptions were not right.	**P3:** "I don't see a clear pattern as I would expect from..."

Table 5.11: Frequencies of codes by experimenter 1 and agreement rate with experimenter 2.

Code	Frequency	Agreement
Pattern	286	93%
Trend	158	90%
Profiling	121	86%
Elimination	116	85%
Storytelling	75	100%
Elimination incl. Trend	57	100%
Creative Desperation	33	100%
Contradiction	26	100%
Verification	25	75%
Pattern incl. Profiling	15	67%
Coincidental Aha	12	50%

Observed sense-making strategies

Table 5.11 shows how often the observed strategies were used by the analysts in descending order. Many strategies are used much less than the top five. The coding scheme is partially based on the *Triple Path Model of Insight* [Kle13], which describes scientific discoveries and inventions. We assume that there is a significant difference between this process and everyday problem solving processes, which results in the difference between the assumptions of the Triple Path Model and our results. We think that this is due to the fact that Klein's research investigates insight generation on a general level, whereas we look at insight generation at a much finer level of granularity. *Contradiction* plays an important role in Klein's model but is, for example, difficult to observe because it does not occur that often in everyday situations and participants are not sensitised towards detecting contradictions. The identification of more *Contradictions*, however, would be favourable in the hypotheses generation process because the assumption is that contradictions lead to insights.

The analysis of insights, which got documented by the user during the tool use, indicates how visualisations are related to the insight generation process. To answers research question 3 we count the number of insights and contrast their quality to the applied strategies.

We identified *Looking for Patterns* as the predominant strategy in our evaluation, coded nearly twice as much as the second most frequent strategy. It includes comparing characteristics of crime events, such as location or offence type, and often describes manual clustering. Hence, participants were mostly looking for similarities in crimes *(Patterns)* and secondly, were looking for *Trends*. The task description contains fairly much temporal information, such as bank and school holidays, and all participants utilized these time periods during their analysis. In contrast, the other specifications (burglary type and town) were sometimes disregarded. The offence type was disregarded in eight

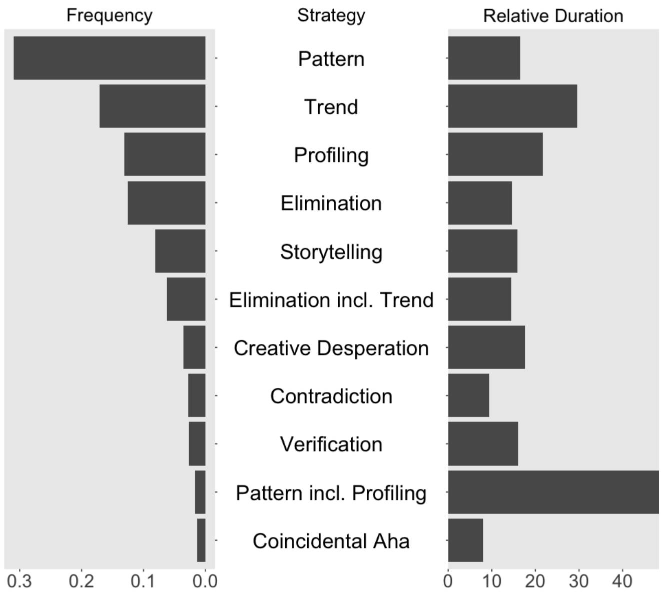


Figure 5.22: Frequency of applied sense-making strategies and their relative duration (sec), ordered by frequency [DHDP+19].

cases and the fictional town in five cases.

There is a difference in how long the different strategies are used. We measured how long the strategies are employed during every occurrence. *Patterns with Profiling* took the longest with approximately 48 seconds, thus, reading crime reports was the most time consuming strategy. Participants searched longer for *Trends* than, e.g., for *Patterns* or *Verification*. *Profiling* and *Pattern with Profiling* took longer than just looking for *Patterns*. Our results further confirm that the *Aha moment* is a quick process and it turned out to be a shorter moment than the one for *Contradiction*, when a participant realises that something is odd (compare Figure 5.22).

Strategies and number of insights

We analysed sense-making strategies in relation to the number of reported insights. The number of employed strategies is moderately associated with the number of reported insights ($r = -.441$). Participants who reported more insights in general used a smaller set of different strategies. There is also a small correlation between how many insights were gained and how often participants changed strategies ($r = -.294$).

To follow up on research question 3, *Are sense-making strategies related to the number of insights?*, we grouped participants by the number of reported insights into three equally sized groups. Participants reporting the fewest insights used the *Profiling* strategy the most. A Kruskal-Wallis test revealed a significant difference in the use of *Profiling* ($\chi^2 = 6.106, p - value = .0473$) across three different groups (Group 1, $n = 6$, 1-3 insights, Group 2: $n = 6$, 4-6 insights, Group 3: $n = 6$, 7-11 insights). Post-hoc pairwise-

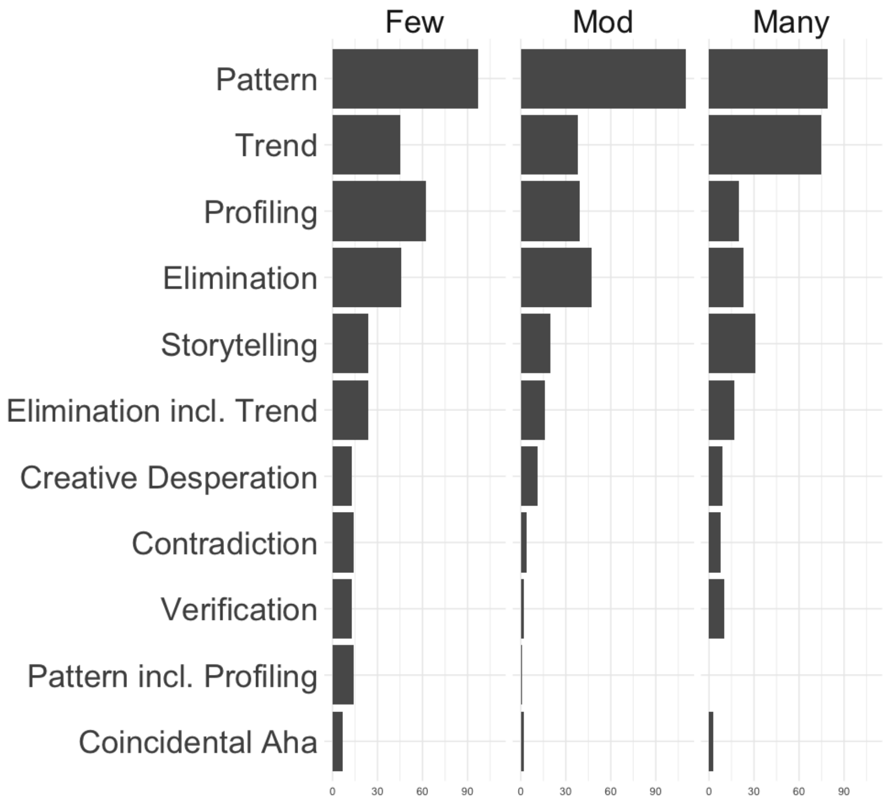


Figure 5.23: Number of insights - more diverse strategies led to fewer insights. *Trends* were analysed most often when many insights got reported [DHDP+19].

comparison of groups using a Bonferroni adjustment revealed that *Profiling* significantly led to fewer insights ($p < .05$) with a large effect size ($r = 0.58$). More insights were reported when *Trends* and *Patterns* were used in a similar fashion (compare Figure 5.23).

A distinct feature of Group 1 (fewest insights) is the usage of *Verification*, *Creative Desperation*, *Coincidental Aha's*, *Contradiction* and *Pattern with Profiling*, which have little to no occurrences in the other groups, compare Table 5.12. This indicates that all these strategies, not only *Creative Desperation*, are related to struggle and a lack of a systematic approach. The *Aha-moment*, for example, sometimes seems to indicate a contradiction that participants were not completely aware of.

Reporting insights

We asked participants to report their insights in a tool of their choice (MS Word/PowerPoint) using screenshots to capture the state of the tool when an insight was gained and add annotations for explanations. We analysed the screenshots to see which visualisations are used to convey an insight and the text descriptions to assess insight quality. We also counted the number of insights. It was up to the individual to choose relevant insights for

Table 5.12: Medians and interquartile ranges of strategy counts per number of insights (few, moderate, many) for all strategies.

Nr of insights	**Few** ($n = 6$) **Mdn (P25 to 75)**	**Mod** ($n = 6$) **Mdn (P25 to 75)**	**Many** ($n = 6$) **Mdn (P25 to 75)**	**p-value**
Pattern	17 (14.50 to 18.75)	19.5 (16 to 20)	11 (7.25 to 16.25)	.2524286
Trend	5.5 (5 to 9)	6 (5.25 to 7.5)	13.5 (10.75 to 14.75)	.06715
Profiling	8.5 (8 to 13.5)	5.5 (5 to 6,75)	2.5 (1.25 to 5.25)	**.04737***
Elimination	7 (5 to 11.25)	6.5 (5.25 to 10)	5 (1.25 to 5.75)	.271
Storytelling	3.5 (2.25 to 4.75)	3 (1.25 to 4.75)	4 (3.25 to 7.25)	.5424
Elimination incl. Trend	4 (0.5 to 6.75)	2.5 (0.25 to 4.75)	3 (0.5 to 4.75)	.7156
Creative Desperation	1.5 (1 to 2.75)	2 (1 to 3)	0.5 (0 to 1.75)	.5708
Contradiction	2 (1 to 3.75)	1 (0.25 to 1)	1.5 (0.25 to 2)	.2504
Verification	2 (1 to 3)	0 (0 to 0.75)	1 (0.25 to 3.25)	.1222
Pattern incl. Profiling	0 (0 to 0.25)	0 (0 to 0)	0 (0 to 0)	.2834
Coincidental Aha	0.5 (0 to 2.5)	0 (0 to 0)	0.5 (0 to 1)	.4406

* significant at $p < .05$; p-value obtained using the Kruskal-Wallis H test.

the report and no minimum requirements were asked (even though the task description suggested time, location and MO to be included). In total 89 insights were reported using 1.7 visualisations per insight on average. Most participants (two-thirds) reported more than three insights. One insight was the minimum, while eleven the maximum. The report quality is moderately correlated with the number of insights ($r = .343$), i.e., when more insights are reported they tend to be of higher quality.

We evaluated the quality of the insight reports, assessing the depth of an insight depending on the policing recommendation. Low-level insights are statements about single facts, such as crime rates at certain dates shown via screenshots. Moderate reports are explanations of a hypothesis and high-level insights are conclusion that can be drawn from discovered problematic crime scenes. A high-level insight report includes at least one policing recommendation.

Example 1: Participant P6 reported high-level insights, such as: *"This display indicates that burglars often choose the rear entrance. Surveillance should therefore focus on that, maybe by choosing patrol routes where officers have clear sight to rear entrances of houses."* She used five different sense-making strategies with the smallest switching behaviour (32 times) and reported eight insights, of which six included a policing recommendation.

Example 2: Participant P16, on the other hand, used eleven different sense-making strategies, switching strategies more often than any other participant (98 times, compare Figure 5.24), and reported three low-level insights. They contained facts on similar crimes using only screenshots of crime reports aligned next to each other for comparison. Hence, *profiling* was applied most, which did not lead to any policing recommendation.

Strategies and quality of insights

Using fewer strategies has a small correlation with higher insight quality ($r = -.209$) and there is a moderate association between how good the quality of the insights were and how often strategies were changed in total ($r = -.365$). Similar to the situation in the case of the number of insights, the quality of insights increases the less people

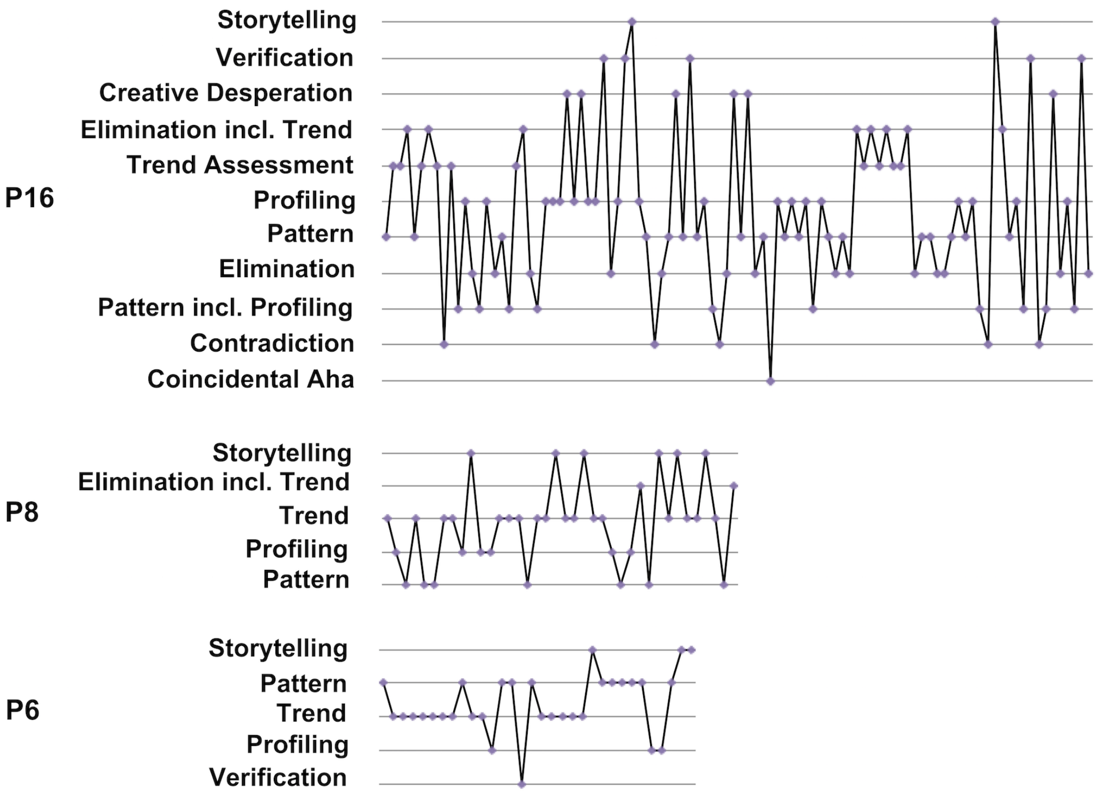


Figure 5.24: The number of applied sense-making strategies varied in the approximately same time used for the task. We observe a better insight quality for participants sticking with each strategy longer, thus, using less strategies in total (P16 reported few, low-level insights, P8 many, moderate, and P6 many, high-level insights).

jumped between different sense-making strategies. *Trends* were assessed more often in the better performing groups (moderate and high-level insights), than in then low-level group. *Patterns and profiling*, on the contrary, were most often used in the low-level insight group. Participants with high-level insights *verified* the most. Figure 5.25 shows the distribution of the mean usage of strategies normalised per strategy.

A Kruskal-Wallis test shows a significant difference in the use of *Pattern* ($\chi^2 = 6.7427$, $p = .034$), *Verification* ($\chi^2 = 6.016$, $p = .049$), and *Profiling* ($\chi^2 = 9.711$, $p = .007$) across the three quality groups (Group C: $n = 5$, high-level insights, Group B: $n = 6$, moderate insights, Group A: $n = 7$, low-level insights). Pairwise comparisons using the Wilcoxon rank sum test showed that *Profiling* led to significantly more low-level insights compared to the number of high-level insights ($p < .05$) with a large effect size ($r = .73$). *Verification* led to significantly more high-level insights ($p < .05$) with a moderate effect size ($r = .57$). The difference of *Pattern* leading to more low-level insights was not significant compared to the other groups. The remaining strategies were used more

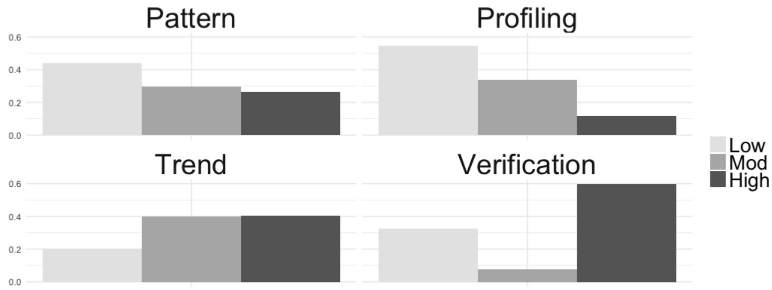


Figure 5.25: Looking at the insight quality, *Pattern* and *Profiling* show a clear tendency to lead to low-level insights. Higher-level insights were gained with the *Trend* and *Verification* strategies [DHDP+19].

equally between the groups, compare Table 5.13.

Table 5.13: Medians and interquartile ranges of strategy counts for quality of insights (low-level, moderate, high-level) for all strategies.

Insight quality	**Low** ($n = 7$)	**Mod** ($n = 6$)	**High** ($n = 5$)	**p-value**
2-4	**Mdn (P25 to 75)**	**Mdn (P25 to 75)**	**Mdn (P25 to 75)**	
Pattern	19 (18.5 to 23)	13 (11 to 18)	11 (10 to 16)	**.03434**
Trend	5 (5 to 6)	11.5 (7 to 14.5)	10 (8 to 14)	.05184
Profiling	8 (7.5 to 11.5)	5.5 (4.25 to 8.25)	2 (1 to 3)	**.007823****
Elimination	7 (5 to 11)	3.5 (2 to 5.75)	6 (5 to 7)	.2404
Storytelling	4 (2.5 to 4.5)	3.5 (2 to 7.5)	4 (3 to 5)	.9349
Elimination with Trend	5 (0 to 6)	4 (2.5 to 4.75)	0 (0 to 2)	.5026
Creative Desperation	1 (0.5. to 2.5)	2 (0.25 to 3)	1 (1 to 2)	.9383
Contradiction	1 (1 to 2)	1 (0.25 to 1.75)	1 (0 to 2)	.8481
Verification	1 (1 to 4)	0 (0 to 0.75)	3 (1 to 4)	**.04938**
Pattern with Profiling	0 (0 to 2)	0 (0 to 0)	0 (0 to 0)	.07098
Coincidental Aha	0 (0 to 1.5)	0 (0 to 0)	0 (0 to 1)	.32

* significant at $p < .05$, ** $p < .01$; p-value obtained using the Kruskal-Wallis H test.

Tool use for insight reporting

The three visualisations from the search result group, time, map and bar chart, were used most often for reporting insights. They often were useful on their own, but also in combination with another view (see Table 5.14). Another basic visualisation, a list, however, was used just once in combination with the map. From the intelligence specific tools the CCT from the MO notes was used the most. The most frequently used combination was the SPC with the time view, which complement each other by additional information over the same period of time. The SPC additionally offers the interaction possibility of selecting time stamp, which allows to investigate the events on a detailed level.

Combinations of a bar chart with another bar chart were used to show crime statistics together with a bar chart with time to emphasise the peaks in the timeline. The most frequent combinations were a mix of a basic and one of the specific views. When more

Table 5.14: Frequency of tools used to report relevant insights in a single or combined view.

Tool use	Single	Combined
Timeline	10	22
Statistical Process Chart (SPC)	3	13
Map (Location)	11	17
Search bar	1	3
Bar chart	6	17
Crime Classification Table (CCT)	10	7
Space Similarity Selector (S3)	2	7
Clusters	-	7
Crime Reports	-	15
List	-	1

than two views were combined they were used more individually, e.g., to show three crime reports next to each other, or to compare several maps. Interestingly, the participant reporting most insights only used the conventional visualisation tools and heavily exploited the bar chart to gain insights. Tools representing the temporal development (Time and SPC) were used more frequently when high-level insights could be gained, which lead to specific recommendations. Also, the CCT was used twice as often in the high-level insight reports as in the low-level insight report group, compare Figure 5.26.

5.4.5 Conclusion

We derived eleven sense-making strategies from the observation of an exploratory analysis task in the area of criminal intelligence analysis. The evaluation of these strategies indicates that concentrating on fewer strategies leads to more hypotheses and better insights. Looking for connections in the data, i.e., *Patterns*, was the main strategy. It was, however, less beneficial for insight generation than we assumed beforehand. The tools seem to support looking for connection *Patterns* better than looking for *Trends*. Participants who focused on analysing *Trends* tended to report better insights. *Coincidental Aha's* occurred rarely and we find little difference within the groups that produced more or less insights than others.

However, there is a significant difference in the *Profiling* and *Verification* strategy suggesting that *Profiling* in our case does not support insight generation well. Better analysis is done via *Verification*, i.e., participants who validated more often also generated more hypotheses. However, participants who do not have many hypotheses in the first place cannot *verify* them. They rather change to another strategy quite arbitrarily, which is reflected in the number of strategies used. Our results indicate that verification should be supported by visual analytics systems to a greater extent. From the literature on cognitive biases we also know that *Verification* can help to overcome biases [NVH+16]. Participants who concentrated on the *Profiling* strategy, on the other hand, tended to

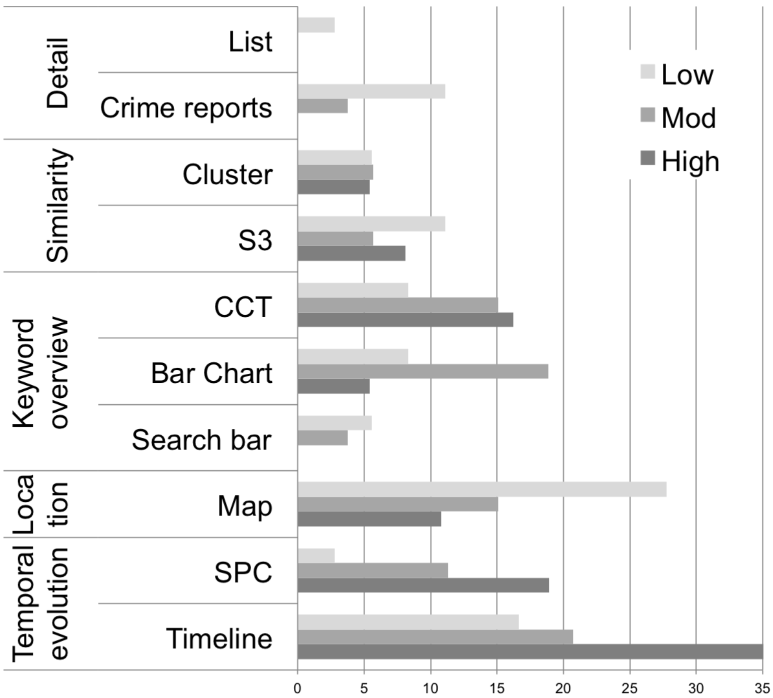


Figure 5.26: Tools used for reporting insights grouped by quality: low-level, moderate, and high-level insights [DHDP+19].

focus on details and could not create more general recommendations. The quality of their insights was, therefore, not as good.

In general, we think that more insights are better than less. In the context of intelligence analysis, however, a quantity of insights is not always desirable if immediate action is required. Therefore, we not only compared strategies to the number of insights but also considered the quality of the reports. Nevertheless, the subsets match well and there is some indication that the participants who reported more had better insights too.

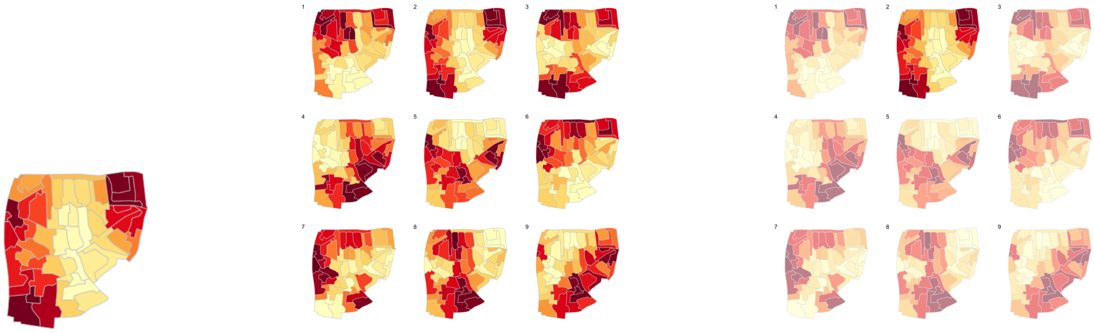
The problem of how to define insights, especially the granularity of insights, has been discussed in the literature extensively. As of yet, there is generally accepted solution for this problem. We approached this problem in a pragmatic way because the granularity was predefined by the tasks created by the domain experts. We think that this bottom-up approach for defining what constitutes an insight might be viable to overcome this problem. Domain experts usually have a fairly precise idea what makes up a relevant insight in their domain. In this way, insights can at least be defined for specific domains.

5.4.6 Comments on dissemination

This work was conducted in co-operation with project partners from Middlesex University London and a manuscript entitled *Sensemaking Strategies vs. Quality of Insights: Inves-*

tigating Analysis Processes Using Multiple Visualizations [DHDP+19] is, at the moment of writing, submitted for publication. I was responsible for conducting the study, data analysis and publishing. In the paper I mainly contributed to the results, but also to the related work, system and study description (Sections 2.3. Criminal Intelligence Analysis, 3. System Description, 4. Study).

5.5 Investigation 5: Geographic map line-ups



(a) (b) (c)

Figure 5.27: Example of a spatial line-up task. A real map a) is hidden in a number of random b) decoys generated to show a lower spatial autocorrelation. c) The task is to identify the map with the higher autocorrelation in the *line-up* of such similar maps.

To analyse sense-making strategies further we were interested if certain sense-making strategies lead to better decisions, and, consequently better results. Therefore, we chose tasks with clear answers for the final study of this thesis, the graphical line-up test [WCHB10]. This test is a practical way of effecting more reliable interpretations. Graphical line-ups, as depicted in Figure 5.27, can be considered as visual equivalents of test statistics. Based on the work of Beecham et al. [BDM+17] we were interested in how people judge the autocorrelation inherent in geographic maps, which is also relevant for the domain of criminal intelligence. The results provide a set of general as well as specific guidelines for choropleth and dot-centroid maps.

Table 5.15: Design recommendations based on investigation I5.

Recommendation		Description
5.1	Use choropleth maps to show spatial autocorrelation	Generally choropleth maps performed better and were preferred in most of the cases. They were especially successful in high values of autocorrelation and irregular geographies.
5.2	Train participants in spending more time on a line-up task	The impartial observer should take the time to study each plot.

Continued on next page

Table 5.15 – *Continued from previous page*

Recommendation		Description
5.3	Train participants on the importance of transitions	Looking for connections and clusters was more intuitive than assessing transitions. Taking a closer look at transitions would benefit in this task as the number of clusters can have a negative impact. Participants favoured plot 6 over 2 in the choropleth line-up example of Figure 5.27 because of the smaller number of clusters, while the correct plot 2 shows a smoother transition.
5.4	Use dot-centroid maps carefully for lower autocorrelated data and more regular geographies	Choropleth maps were preferred by the users, but some lower autocorrelated line-ups could be better assessed with the dot-centroid map. A combination of both or being able to switch the visualisation technique might prove beneficial.

5.5.1 Motivation

> Everything is related to everything else, but near things are more related than distant things.
>
> Tobler's First Law of Geography [Tob70, p.236]

If statistical graphics are to be used in data analysis and reporting, there needs to be reassurance that the statistical effect implied by a graphic can be perceived reliably. The possibility of a mismatch between statistical effect and its visual perception is especially relevant to geovisualisation. Whilst maps convey information around the location and extent of phenomena that may be difficult to imagine using non-visual techniques, they may also lead to artefacts that are incidental to the statistical structure under investigation and that may even induce interpretation of false structure.

In the graphical line-up test a plot of real data is hidden amongst a set of decoys generated under a null hypothesis. If an impartial observer, an individual who has not previously seen the plot, is able to correctly identify the real data from the decoys, then this would lend confidence to the claim that a statistical effect exists. In null hypothesis significance testing this would suggest that the observed data is not consistent with the specified null hypothesis.

Graphical line-up tests are straight-forward to implement and conceptually appealing. They offer much potential to geo-spatial analysis [BDM$^+$17]. However, they do not fully negate concerns around reliability of perception. Recent empirical studies have demonstrated that perception of statistical effects varies systematically with the intensity of effect [BG10], with visualisation design [CYFR14] and in the case of geovisualisation, with the geometric properties of the regions being studied [BDM+17]. Whilst there is evidence to suggest that these variations in perception are sufficiently systematic to be modelled (e.g. [CYFR14]), the evidence is less compelling for representations of geo-spatial data in choropleth maps (e.g. [BDM+17]).

Beecham et al. [BDM+17] speculated about potential explanations – why it is that, after modelling for variation due to statistical intensity and geometric irregularity, there is still much variation in participants performance on assessing the statistical structure encoded in maps. Elsewhere, Hofmann [HFMC12] and later VanderPlas and Hofmann [VH16] investigated whether or not the ability to make correct judgements in (non geo-spatial) line-up tests varies as a function of the individuals' perceptual capability and reasoning or some other demographic characteristics. Whilst both studies found variation in individual ability to interpret line-up tests, this variation was not consistent with demographics or visual abilities.

This study attempts to address the problem from a different perspective. Through structured qualitative analysis of thinking aloud protocols, we attempt to expose the *sense-making processes* through which judgements are made during line-ups displaying geo-spatial data. If we are aware of the users' sense-making processes while they solve problems with visualisations, recommendations can be provided for designers on how to design graphics and develop systems accordingly.

Easy line-up tasks are those that suggest a solution immediately. This is due to pre-attentive processes which are well-known to enable fast pattern recognition. We are interested in the not so clear cases, where other strategies have to be advised, and designed demanding line-up tests where participants fail as often as they find the true answer. This was accomplished by two rounds of pre-tests. On the one hand we wanted to know if humans come to the same conclusions disregarding the truthfulness of their answers and if a gut feeling still exists when they can not explain their assessments. In addition, we developed difficult line-up tasks to be able to assess which sense-making strategies are successful and which are not. An additional question is whether differences in maps can be detected by using similar strategies as in our previous studies.

5.5.2 Approach

When presenting data in maps, analysts are often concerned with the role of space, or spatial association, in phenomena. One question might be, for example, to what extent high crime rates are concentrated in certain neighbourhoods of a city and low crime rates in others. *Spatial autocorrelation* is a concept used widely for describing such tendencies [OU10] which helps to understand the degree to which one object is similar to other nearby objects. We use the formal statistic of *Moran's I* for quantifying the amount or *intensity* of spatial autocorrelation [Mor50].

A test statistic for spatial autocorrelation is typically derived by comparing an observed intensity of *Moran's I* against a distribution that would be expected under

Complete Spatial Randomness (CSR) or some sensible prior knowledge [BDM+17]. Whilst theoretically convenient, the assumption of CSR has been much criticised since it is a state that rarely exists in practice. Within the more informal framework of graphical inference, however, it is possible to perform tests of spatial autocorrelation in line-ups with decoys that assume autocorrelation based on some sensible prior knowledge [BDM+17].

Previously, Beecham et al. [BDM+17] measured the precision with which differences in spatial autocorrelation can be perceived in choropleth maps through a large crowd-sourced experiment. The authors investigated how this precision varies with increasing geometric irregularity and increasing intensity of statistical effect (see Figure 5.28). They found that as the intensity of autocorrelation structure increases, the difference in statistical effect necessary to correctly discriminate that structure decreases. Additionally, they found that as geometric irregularity increases, so too does the difference in statistical effect necessary to correctly discriminate that structure. Beecham et al. [BDM+17] also found much variation in ability to discriminate structures that could not be explained easily through the experiment conditions. This unexplained variation may relate to physical artefacts introduced in choropleth maps that could not be controlled systematically. Or it may relate to differences in qualitative heuristics – *strategies* – applied by participants when making judgements.

This research aims to expose and characterise these *strategies* that are employed when making judgements in *map* line-up tests.

We developed a series of map line-up tasks with line-ups consisting of nine data graphics: one plot of *real* data hidden amongst a set of eight decoy plots. The study conditions were generated by adapting the resources from Beecham et al. [BDM+17] to generate 3×3 line-up plots.

We conducted a randomised controlled study ($N = 19$) where the following conditions were varied:

- Geovisualisation design: choropleth map | centroid-dot map
- Geometric irregularity: artificial grid | real geography, regularly shaped | real geography, irregularly shaped
- Graphic size: small | large
- Statistical intensity: low | high

We analysed how the ability to perform line-up tasks, i.e., to correctly identify the real from the decoys, varies under the above conditions. We also paid attention to participants' perceived confidence in making line-up judgements under different conditions and their preferences amongst the different conditions.

We use the same geometries as Beecham et al. [BDM+17] in two sizes, a small and a large dataset, and two variants of the same data, a traditional filled areas map (choropleth) and a centroid-dot representation. Additionally, we design line-up tests with two intensities of baseline statistical effect: *Moran's I* of 0.8 (high) and 0.3 (low). A summary of the conditions tested is displayed in Table 5.16.

Table 5.16: Study conditions.

2	*geovis types* (choropleth, centroid-dot)	×
3	*geometric irregularity* (grid, regular, irregular)	×
2	*map size* (small ≈ 50 units, large ≈ 100 units)	×
2	*statistical effect* (low 0.3, high 0.8 – *Moran's I*)	×
19	participants	=
456	tests overall	
24	unique test conditions	

5.5.3 Description of the visualisations

One hypothesis for the large variation in perception identified in Beecham et al. [BDM+17] relates to geovisualisation design. In a choropleth map, the entirety of each spatial units is given a single colour, which can result in salience bias favouring larger regions and other artefacts that are due to statistical effect. We therefore also test a centroid-dot alternative, where a coloured dot represents the autocorrelation value in the centre of an area, as opposed to the whole area being colour-filled. As an example compare first with second row in Figure 5.28.

We vary geometric irregularity in the same way Beecham et al. did: a grid, regular real map and irregular real map. We use the same geographic maps, but add a further set of three maps with approximately the same levels of geometric irregularity but twice the number of spatial units (from approximately 50 to approximately 100 units).

Introducing the centroid-dot maps brings some additional challenges: with a white

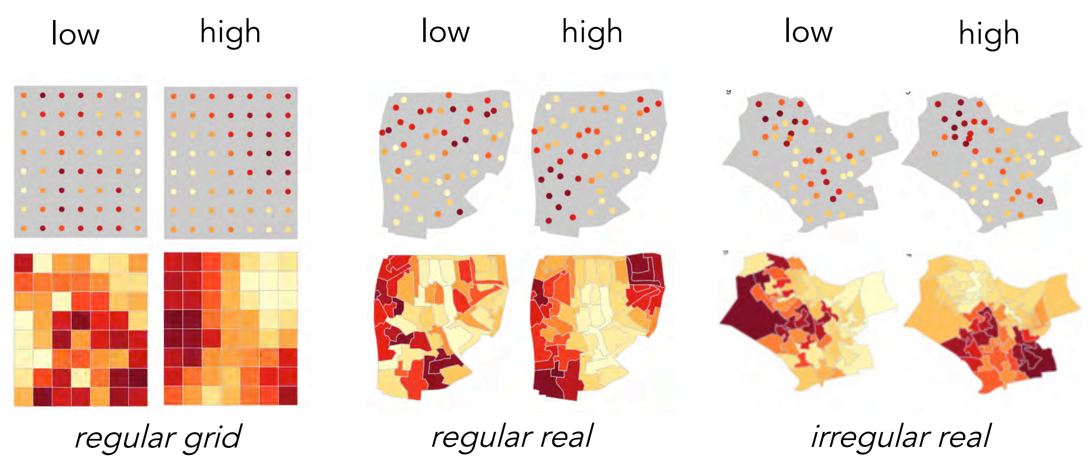


Figure 5.28: The difficulty of line-up tests is assumed to increase with increasing irregularity of spatial units, i.e., grids and regular geometries are easier to assess than more complex, irregular geometries. The lower autocorrelation condition is assumed to be easier than the high condition for each geometry [DHBDP19].

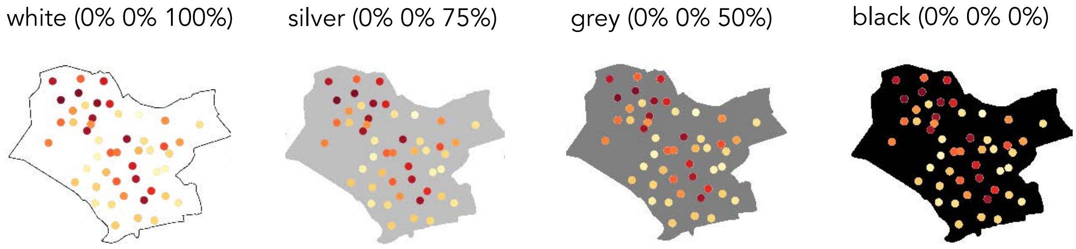


Figure 5.29: In the centroid-dots map the background colour influences the perception of coloured dots. Both light yellow and dark red should be perceptually on an equal level. A black background emphasises light dots, whereas a white background the dark ones. A 50% grey proved to be more dominant than a silver background [DHBDP19].

background, dark dots gain greater saliency, whilst the contrary is true of a black background. A light grey background appeared to minimise these artefacts (compare Figure 5.29).

Since we wanted to expose the various strategies employed when making line-up judgements we generated challenging line-up tasks. We therefore selected *difference* values for decoy plots based on the thresholds published in [BDM$^+$17]. These thresholds loosely represent the minimum difference necessary to correctly judge between two choropleth maps 75% of the time – a quantity referred to as Just Noticeable Difference (JND) – and take into account the modelled influence of geometric irregularity and intensity of statistical effect. By choosing difference values in this way, we hope to control the influence of geometry and intensity of any statistical effect, i.e., we *expect* no difference in performance due to these factors.

5.5.4 Design and tasks

We used a within-subject ($N = 19$), counter-balanced design, where every participant performed 24 line-up tasks in random order. Each line-up was composed of nine images: eight *decoy plots* and one correct *target*, randomly positioned in a 3×3 array. The *target* was made different from the decoys by increasing its autocorrelation value in line with the thresholds published in Beecham et al. [BDM+17] – one-tenth higher than the decoys. The JND value depends on the geometric irregularity (grid | regular | irregular) and baseline intensity of *Moran's I* (low | high). All eight decoys contain approximately the same autocorrelation level. Note that each decoy is unique – even if it contains the same spatial autocorrelation value. The decoys were generated using a permutation approach published in [BDM+17]. Through permutation of the regions it is possible to generate several maps for one autocorrelation value. Hence, recognising the same data could not happen and memorising results was not possible.

For each map line-up we asked the following three questions:

1. Which is the plot with the highest spatial autocorrelation?
2. How confident are you in your choice?
3. Are there possible alternatives?

We specifically decided to develop challenging tasks that forced participants to reflect explicitly about the problem and possible solutions. In this way, we were able to evaluate the employed strategies better than with simple tasks that can be solved at a glance. In easier scenarios, participants are less able to verbalise how they reached a solution because the reasoning process is fast and unconscious.

Dataset

We arrived at threshold values for our conditions based on the experience of two pre-tests. Important considerations included the time taken to complete the experiment (restricted to 60 minutes) and generating stimuli of sufficient levels of difficulty to trigger slow sense-making processes. We, therefore, expose *strategies*, rather than testing for pre-attentive perceptive abilities. In addition, we wanted to assess which sense-making strategies are successful and which are not. This is only possible if we can observe a certain number of failures.

Difference in the decoys of the line-ups

The JND thresholds by Beecham et al [BDM+17] were generated under a setting that was different to ours. Rather than a full map line-up, participants had to compare two images at a time in an established staircase procedure where the difficulty level changed based on participant performance. The aim was to encourage learning and improve performance to the extent that the JND level represents the minimum perceptible difference between the two stimuli. Since only two stimuli are used, the intuitive explanation of JND –

the difference necessary to correctly discriminate 75% of the time – cannot be easily transferred to sour study since the 75% figure must also include some chance-guessing.

We started by constructing line-up tests using the minimum JND threshold values published by Beecham et al [BDM+17] and completed two pre-tests with two and three participants in each. After the first round the feedback was that line-ups were very challenging and exhausting – participants sensed that their performance was poor and the level of challenge resulted in a lack of motivation or engagement. Participants managed to correctly identify the target from the decoys in approximately a third of occasions. This difference from the expected success rate given Beecham et al.'s JND thresholds is likely due to departures in procedure – most probably that the line-up protocol minimises chance-guessing and that no staircase procedure was used. We hypothesise that a 50:50 success rate would suggest tests that are sufficiently challenging to expose user strategies and maintain the motivation of the participants. Primarily, it is necessary to provide a sufficient number of correct and incorrect tests in order to analyse the circumstances under which correct and erroneous judgements are made.

In a second round we increased the difference to the median JND thresholds used by Beecham et al [BDM+17] and found that the target was too easy to identify, with the effect that almost all answers were correct. Based on this observation we chose the value midway between the minimum and the mean JND's per geometry (mean(JND) − (mean(JND) − min(JND))/2) and Moran's I, which yielded the anticipated 50:50 performance (see below for details).

5.5.5 Participants and procedure

We conducted a study with 19 computer science students with a Bachelor's degree or higher, from which eleven were male and eight female. Participants were between 23 and 31 years old with fair knowledge of geovisualisation (average 3.15 on a five-point Likert scale). We asked *"How familiar are you with map visualisations?"* with ranges from extremely, moderately, somewhat, slightly to not at all familiar. One participant reported a mild red-green colour perception deficiency. Due to incomplete data we had to exclude one participant from the analysis. Participants were trained on both map types with different data than in the experiment.

Participants had 10-15 minutes time for training prior to the experiment. They were introduced to spatial autocorrelation by using the same material as Beecham et al.[BDM+17] (see Figure 5.30). To explain the concept of spatial autocorrelation we used a neutral context, i.e., we used examples from real estate and terrain data. Secondly, we tested their understanding by providing six example line-up, one per map type and geography (2×3), and let them learn by giving feedback on their answers. We asked participants to complete line-up tasks as depicted in Figure 5.27 by identifying the plot with the *greater* level of spatial autocorrelation. They were asked to think aloud in these examples, so they practised the method at the same time.

In the trial we first collected demographic information, followed by the line-up tasks and finally preferences on the map types. In the actual experiment, no feedback on participant performance was given. Additionally, we deliberately provided no context with

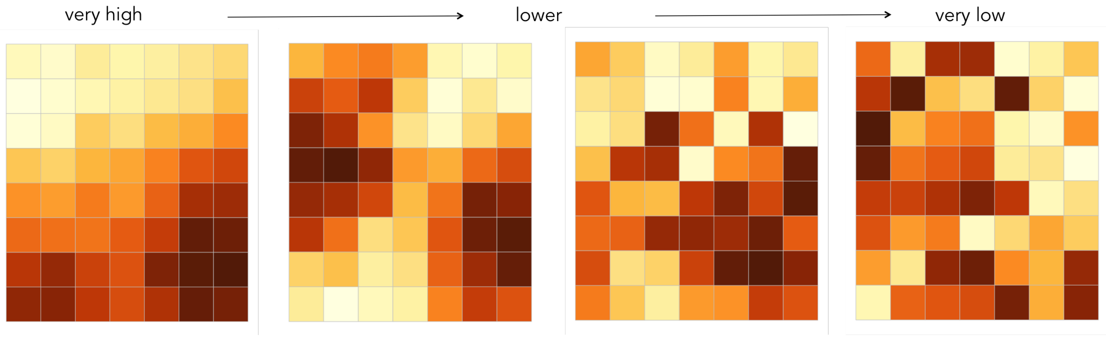


Figure 5.30: Example of decreasing spatial autocorrelation from very high to very low. Image reproduced with permission from Beecham et al. [BDM+17].

respect to the phenomena and spatial processes under investigation. Special attention was paid to participants providing explanations behind their judgements that included storytelling. Experiment sessions lasted between 50 and 60 minutes.

5.5.6 Results from the quantitative analysis

The participant performance was as anticipated – overall around 50% of line-ups were answered correctly, corresponding to 213 line-ups , i.e., the target could be identified from the decoys, and 219 incorrectly identified the wrong target. For each test condition we considered the number of participants that performed the line-up correctly. Figure 5.31 displays this information as well as a frequency plot of participants' self-reported confidence in their answers for that condition on a 5-point Likert scale (1=not at all confident; 5=very confident).

The test condition with the highest success rate was the small centroid-dot map with a regular geometry and low level of baseline autocorrelation. Comparing success levels between geovisualisation type, we found, opposed to our expectation, that the choropleth maps were associated with higher success rates than the centroid-dot maps (Cohen's d effect size 0.64).

An even larger effect was observed between the high and low baseline autocorrelation cases ($d = 1.64$). Our participants were more successful in low autocorrelation conditions than in high ones. There is no obvious difference in success rates between map size (small and large) ($d = 0.02$).

On average, participants needed 42 minutes to complete the line-ups. There is a small correlation between used time and performance ($\rho(T, P) = .27$). The greater the time spent studying the line-ups, the better the performance. We have eight participants with a good success rate of more than 50%. The remaining 10 answered less than half of the test cases correctly. The individual performances differ significantly. The best performance is 18 out of 24 line-up tasks answered correctly (75%), the worst performance is 5 out 24 (20,8%). We can neither observe a clear increase nor decrease in performance over time: whilst participant performance did not improve over time, neither

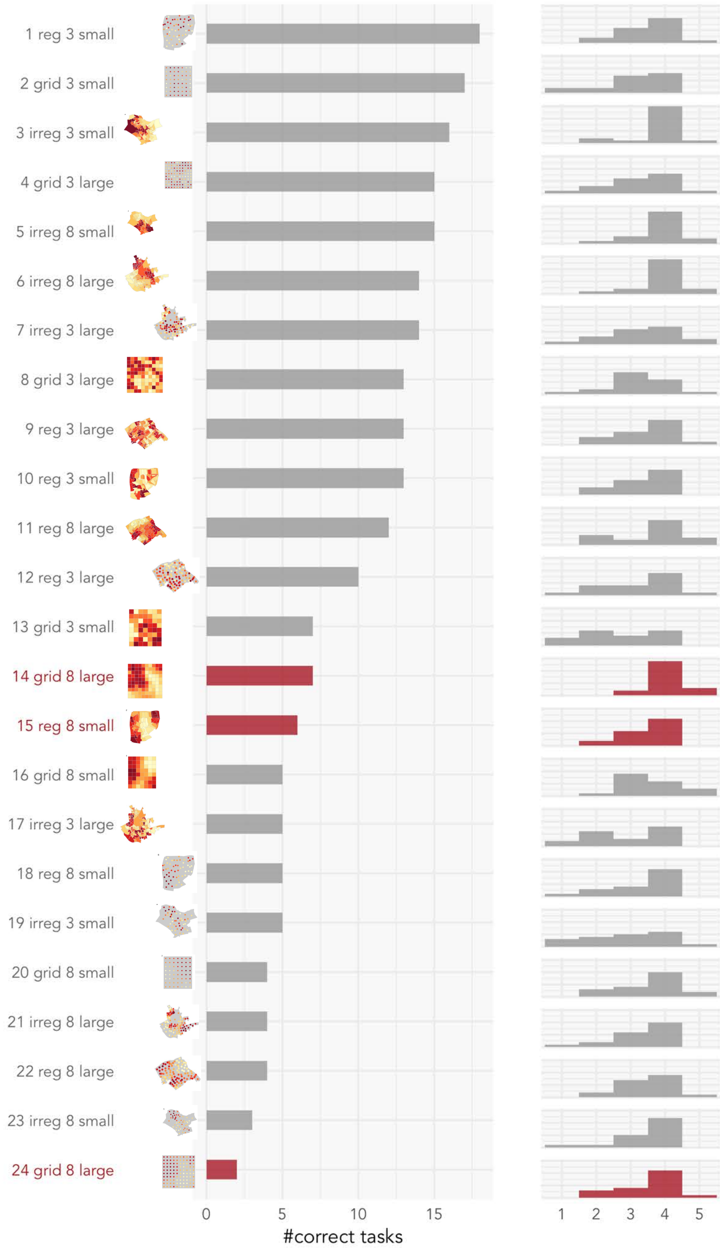


Figure 5.31: The conditions 1-24 are ranked by participant performance, i.e., the number of correct solutions. They are labelled by a description of geovisualisation type, Moran's I, size and a thumbnail of the real plot to show the map type. The reported confidence on a Likert scale from 1 (not at all) to 5 (very confident) shows that participants were overconfident in the incorrect cases. Red lines show conditions that are discussed in the strategy examples, compare Figure 5.27, Figure 5.35 and Figure 5.34, respectively [DHBDP19].

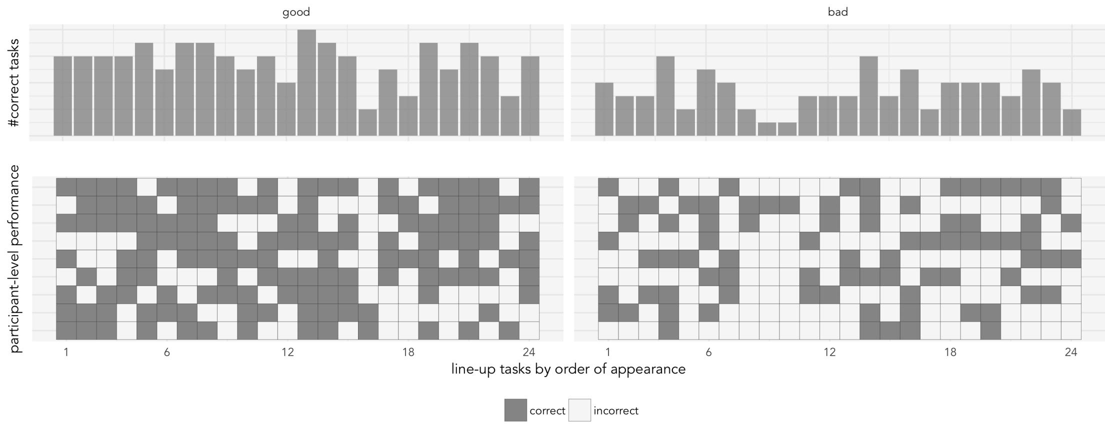


Figure 5.32: Participant performance over time, split into *good* participants (those with a success rate of > 50%) and *bad* participants (with a success rate < 50%). The left-most column represents the first test that participants performed; the right-most column the final (24th) test. The bottom graphic displays individual, participant-level performance – each row represents a participant and each column identifies their nth test. The top graphic displays counts of these columns – the number of correct identifications for the nth test that participants performed. Note that the conditions shown to participants were randomised – therefore we do not expect patterns of performance over time to be a function of relative difficulty.

did it deteriorate towards the end of the experiment, although verbal protocols include statements regarding fatigue in the last third of the experiment (compare Figure 5.32).

Participant preference

Participants expressed a strong preference for choropleth maps over the centroid-dot maps. This was true of all geometries (grid 29:9, regular 32:6, irregular 32:6). Only a small number of participants preferred the centroid-dot maps and in a small number of specific cases: in the grid geometry five participants expressed a preference for the centroid-dot maps.

Overall the centroid-dots irritated the participants due to overlaps and differently sized areas between the dots. Nevertheless, when reporting this frustration, participants reflected on the importance of inter-zone distances. In one dot case a participant pitied a *lonesome* dot, as it *"stands all alone by itself"*, reporting that it would especially catch their attention. Our design seemingly did not overcome the saliency problem with large areas and may have exacerbated it.

Participant confidence

Asked directly about their confidence, participants were more confident with the choropleth maps (mean 3.49) than with the centroid-dot maps (mean 3.33). The top four

Table 5.17: The sense-making codes show a high inter-coder agreement (Cohen's Kappa $\kappa = .976$, $N = 19$).

Code	Frequency	Agreement rate
Cluster	499	99.6%
Transition	355	98.7%
Outlier	354	99.1%
Colour	162	97.6%
Figure	70	85.1%
Gut feeling	58	93.9%

mean confidence values were reported in choropleth map conditions. The success rate and self-reported confidence of each condition is shown in Figure 5.31. A chi-square test of independence showed no significant linear association between the confidence and success rate, $\chi^2(1) = 6.334, p = .175, (V = .118)$. This is confirmed by visually scanning Figure 5.31. An interesting observation, hidden by the summary statistic here is that high success rates are generally associated with high levels of confidence, but also that the lowest levels of confidence do not appear at the bottom of the graphic – where the success rate is lowest.

Participants were asked to provide an optional second choice for the target. On average only 6 of the 19 participants named a possible alternative for the line-up test. There is a small negative correlation between number of alternatives and mean confidence ($\rho(A, C) = -.190$), i.e., the higher the confidence the fewer the number of alternatives. This association is confirmed by the fact that the condition with the fewest alternatives offered was associated with a high success rate ($> 75\%$), whereas that with the most alternatives offered (12 participants in total specified an alternative target) had a success rate of only 50%.

5.5.7 Qualitative analysis of strategies

We conducted a thinking aloud to study to investigate the processes underlying the perception and interpretation of autocorrelation in maps and how people try to solve such tasks. For the interpretation of the thinking aloud data we used qualitative content analysis [Sch12, May03]. One of the main goals was to identify the main cognitive strategies people use in such a context.

Qualitative content analysis can either adopt a bottom-up or a top-down approach. The top-down approach requires investigators to develop a frame of categories before the study, the bottom-up approach consists of a repeated processing of the material with the goal to structure the material in a way to derive the categories from the material itself. We adopted a bottom-up approach but were inspired by codes identified in previous research. After transcription of the thinking aloud protocols we developed a coding scheme, similarly to the previous studies. The codes for sense-making strategies were based on the content and consequently discussed and reviewed with the co-authors.

The final six codes were used by both researchers and their high inter-coder agreement (Cohen's Kappa $\kappa = .976$, compare Table 5.17) shows that the codes are unambiguous and fitting for the data.

We, therefore, identified six strategies: *Searching for clusters/identifying connections*, *Analysing transitions, Identifying outliers, Emphasising colour, Recognising figures (Storytelling)* and *Gut feeling*. The most important strategies are *Searching for clusters, Analysing transitions, Identifying outliers* and *Emphasising colour*. The other two strategies were less important in this context.

Contrary to our expectation, we could not find any relationship between the strategy used and the performance of the participants. We specifically assumed that the strategy analysing transition would be more successful than the others because it is tightly related to the task of finding autocorrelation. This was not the case. In the following we describe the results of the performance and the strategies with exemplary participants' statements.

Strategy usage and performance

Better participants switched less frequently and applied fewer strategies than participants with a success rate of less than 50%. Their employed strategies are shown in the top row of Figure 5.33. Comparing them to the worse performing group, *transition* could be better employed by the good performers, as could be *outliers* – but a clear pattern does not emerge. The strategies of the worse performance group is shown at the bottom of Figure 5.33.

The individual performances differ to a high degree, from the best performance with 18 out of 24 line-up correct cases (75%) to the worst performance of 5 out of 24 (20,8%). The best participant took more time (55 minutes) than the average (42 minutes) for the tasks and looked at every map in close detail. She reported that the centroid-dot condition was harder as was the higher autocorrelated condition compared to the lower, a good self-assessment as five out of the six wrong answers showed low autocorrelations. The best participant used 81 strategies (with an average of about 83).

Most often *Clusters* was used with *Outliers* being the second most popular choice of strategy. These strategies were the only ones that could be applied successfully, i.e., leading to more correct answers than wrong ones. The participant with the least correct answers reported a red-green colour-vision impairment but afterwards reported that she had no problems discerning the colours. She noted that lighter colours seemed to stick out less than the dark ones. Although she completed the line-ups in average time, much more strategies (118) than the average (83) were used. We observed quick switching between strategies and insecurity about whether a strategy was useful for the task. Interestingly, three out of the five correct line-ups were the supposedly more difficult ones, the low autocorrelated centroid-dot regular real and the high autocorrelated irregular real small and large choropleth map.

The strategies will be described in detail in the following.

Figure 5.33: Strategies used by the better participants group (top left), the best participant (top right), the worse participants (bottom left) and the worst participant (bottom right).

Searching for clusters

The grouping of units into clusters of the same colour has a big influence on the decision about autocorrelation.

Examples: *"There is one big cluster in the middle", "Here we have two clusters on the sides", "I take the one with the fewer clusters", "I will choose the one with the big centred cluster above this one which also looks nice, with the left to right separation"*.

Identifying clusters was the most dominant strategy employed, including mentioning the number, size and the position of the cluster. If the form of a cluster was mentioned explicitly the statements were coded as both, *Cluster* and *Figure* strategy.

We can summarise the following observations:

- Bigger and fewer clusters were favoured. The size and consequently the number of clusters had an effect on the decision making. If there are fewer clusters in the decoys it can happen that they look more homogeneous than the higher autocorrelated plot with, e.g., two clusters having a smooth transition, and, therefore, a higher Moran's I value. This happened, for example, in the small, regular, choropleth map line-up (compare Figure 5.27 plot number 6: one red and one yellow cluster).

- Centred clusters were favoured. The position of the cluster influenced the decision in some cases and centred clusters had a greater effect than those on the side. The cluster in the middle got over-emphasised, e.g., in the case of small grid high autocorrelated centroid-dot maps (compare the position of the cluster in Figure 5.34 plot number 5).

These strategies seem intuitive and can lead to good decisions, however, they are not reliable for autocorrelation judgements. The clusters in the high autocorrelated regular small choropleth map condition, for example, were often wrongly interpreted and decoy plots seemed to fit better to these strategies than the correct real ones (compare the low success rate in Figure 5.36 and the *Cluster* example in Figure 5.34).

Analysing transitions

This strategy summarises all statements on transitions, where participants looked for smoother changes within each plot.

Examples: *"There is a nice transition to the centre", "This evolves beautifully from light to dark".*

Transition was remarkably well reported in the condition of high autocorrelated regular large choropleth line-ups, which led to a good success rate compared to the centroid-dot (compare Figure 5.36) and fairly confident decisions (compare *11 reg 8 large* in Figure 5.31).

Elimination due to single outliers

Participants used single outliers in a plot as a reason to exclude the plot from the possible range of answers. Hence, this elimination strategy was sometimes heavily applied when no positive example stood out. The maximum of 8 times per line-up, however, was rarely employed, but instead, a switch to a different strategy occurred with the reduced set of plots.

Examples: *"I will exclude this because of these high contrasts here", "Here are also these high contrasts", "There is a very light one right next to a very dark one".*

Emphasis on colour

Dark hues had a greater impact than light ones. Darker colour hues were more often explored than the yellow, light hues when looking for clusters and transitions in both the choropleth map and the centroid-dot map. Some participants reported on looking at orange *"mid-level"* colours, some tried to look at both and change their focus when they were stuck at an impasse, but those were the exception.

Examples: *"I mostly look at dark areas", "Maybe I should look more on light colours.. I will try that now."*.

Regarding outliers light and dark hues were equally distracting, e.g., *"One light in the middle of this dark area"*, as well as *"..but then there is this one dark dot here"*.

Recognising a figure (storytelling)

Storytelling as an option to design visualisations has already been discussed in the visualisation community [KM13, LRIC15]. It has been argued that the development of a storyline supports users in forming connections between disparate facts to make them memorable. Similar issues have been discussed in cognitive psychology too (see, e.g., [ZMG95]). Therein, the emphasis is on the activities of the study participants and how they make sense of the material that is being presented to them. Participants try to construct coherent models based on this material. We called a strategy storytelling when participants identified meaningful shapes in the maps that helped them to solve the tasks. Figure-like cluster arrangements have a big impact.

Examples: *"Here are dark clouds", "A blob or an island", "..like a mountain range", "This looks somewhat like Vienna"*.

If a random plot resembles some kind of figure (a recognisable pattern with some kind of meaning) it has a strong effect on the user's decision although it is unrelated to the task of judging autocorrelation, e.g., *"This looks somehow like a man with a stick"* (compare decoy number 8 of the *Figure* line-up in Figure 5.35). This can, of course, also help if the figure is seen in the correct real, as it was the case in the low autocorrelated large choropleth grid condition where the highest autocorrelated plot was described as the shape of a *pincer* (compare real of condition *8 grid 3 large* in Figure 5.31). Confidence, however, was not high in this case (compare confidence in Figure 5.31).

Trusting a "gut feeling"

This strategy summarised statements about first impressions and *gut feelings*, where no rationale could be verbalised. It is hard to discern aesthetics and other statements from this category, but we wanted to grasp how often participants were stuck at an impasse and had to give up like this, or just liked to trust their guts. We think it is due to the difficulty of the line-up tasks that this strategy was not employed very often.

Examples: *"I don't know, it is just a feeling", "This was my first impression"*.

Although participants mostly could not make a decision at first glance, they often had first choices as a starting point. However, this depends decisively on the participant and the verbal reporting (participants usually do not verbalise in the same amount of detail). For example, participant 17 mentioned a first idea in 16 of the 24 line-ups, but only stuck with this choice in five of the cases. The explanation then was, e.g., *"Here I will stick to my gut feeling"*. More cautious people, on the other hand, stated their gut feeling only when they were more or less sure about it, e.g., participant 19 verbalised eight times a first idea and then stuck with it in seven of the cases.

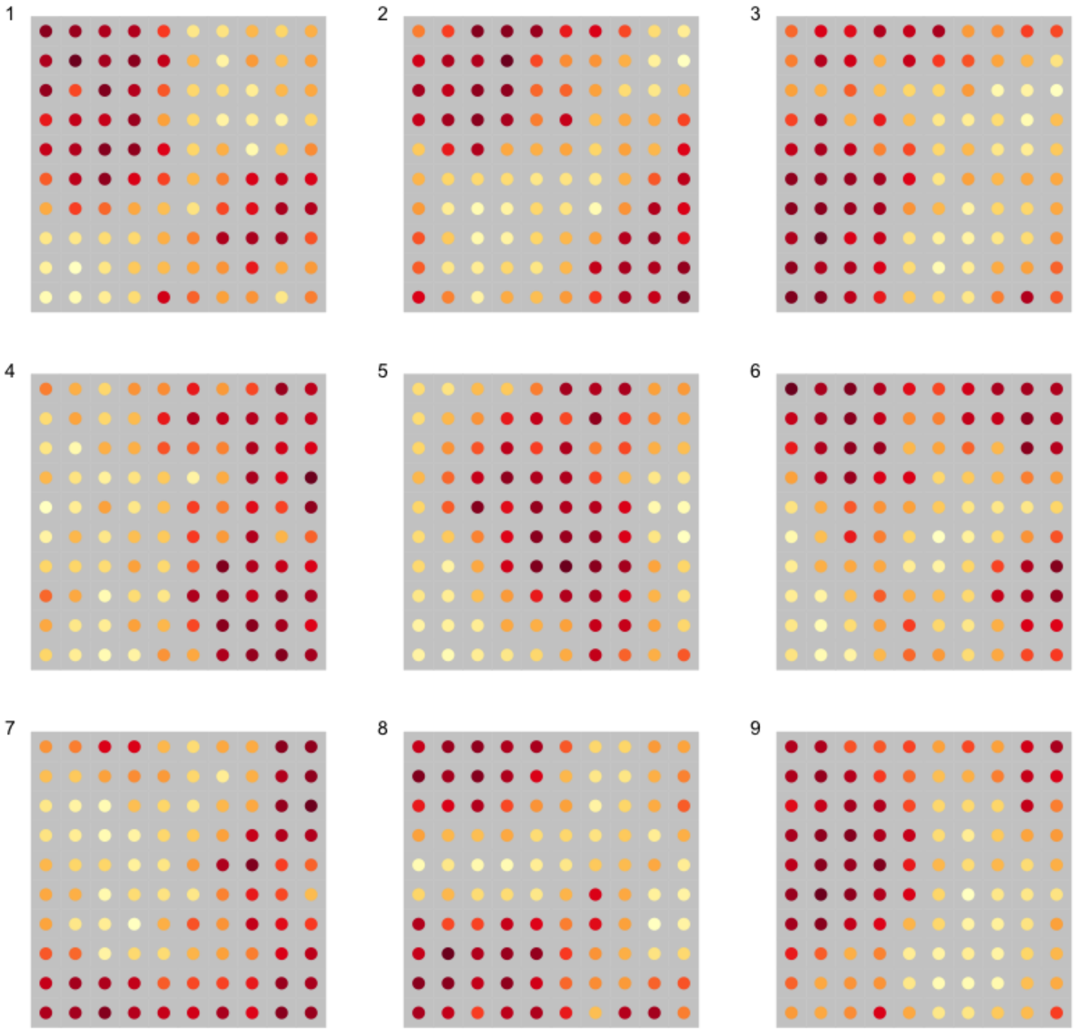


Figure 5.34: Condition *24 grid 8 large centroid-dot*: Most participants chose plot 5 because of the dominant centred cluster instead of the smoother, higher autocorrelated plot 9, leading to the lowest success rate of all tasks [DHBDP19].

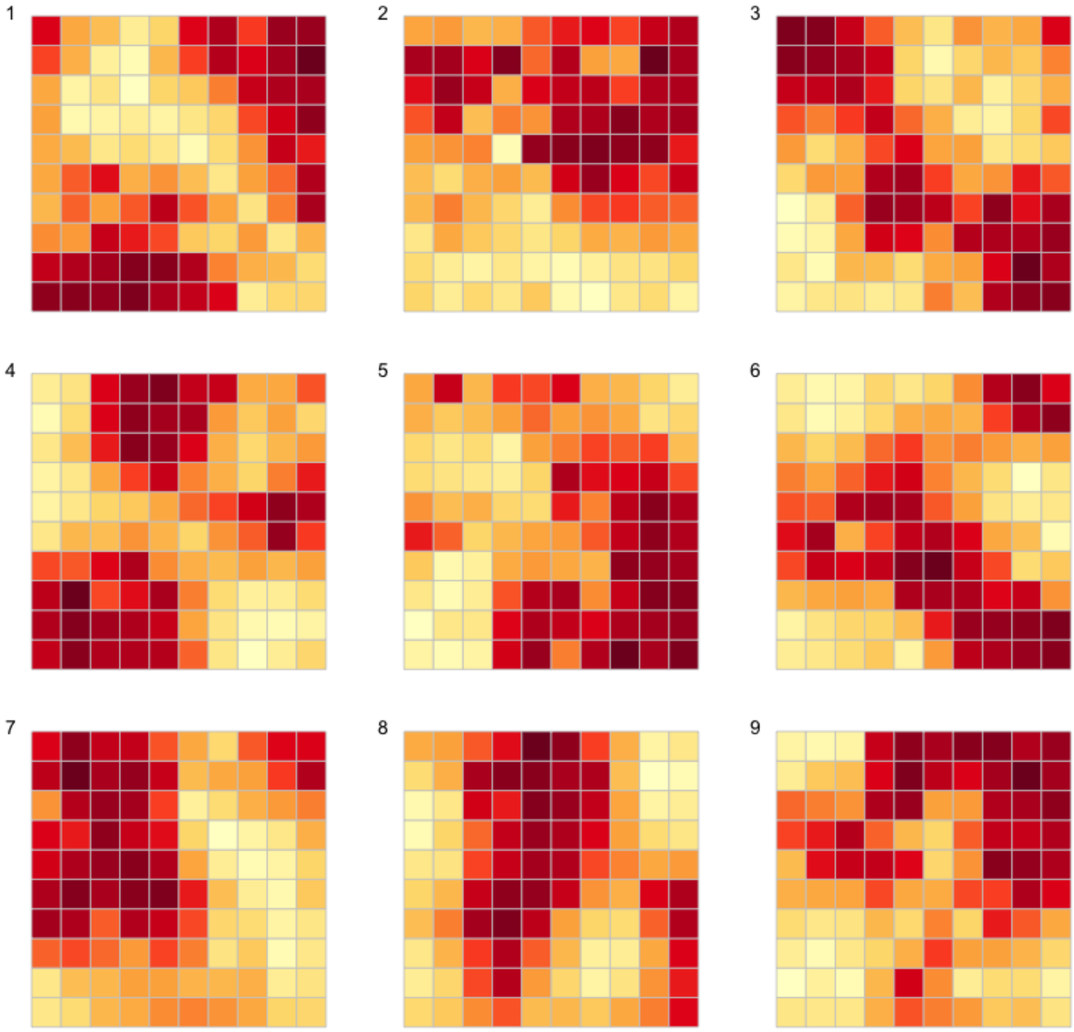


Figure 5.35: Condition *14 grid 8 large choropleth*: Most participants chose plot 8 in this line-up task because they perceived some kind of *Figure*. Plot number 7 corresponds to a higher autocorrelation value and, therefore, would have been the correct answer [DHBDP19].

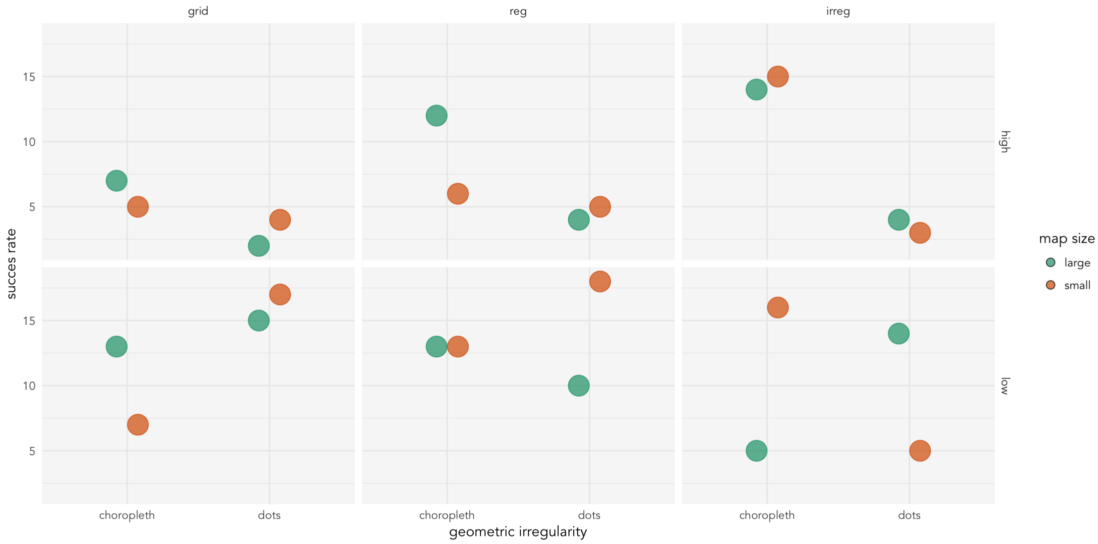


Figure 5.36: Each dot represents a unique line-up, representing each experiment condition. The vertical position of dots varies according to the *success rate* for that line-up, i.e., the number of participants that correctly identified the real from the decoys.

5.5.8 Generalisation of strategies

We assumed that participants would use the transitions most often because it seems to be most appropriate for the task of finding autocorrelation. Instead, they relied most often on the strategy of finding clusters. Some of the strategies are more general and were also identified in other contexts: *Finding clusters, Identifying outliers (Elimination), Storytelling* and *Gut feeling* [DHSP+17, HPH+nt].

Especially finding clusters seems to be a very general strategy. It has some similarities with finding connections in Klein's model [Kle13] where it is an important element. Klein argues that finding connections often occurs in the context of problem solving activities. In our experiment participants described clusters as elements having some connection with each other. From this they concluded that autocorrelation exists. This is also confirmed by Doppler Haider et al. [DHSP+17], where *looking for connections* was a popular strategy.

Finding connections seems to be a kind of default strategy people adopt when they are not sure how to solve a problem. This might be the reason why participants applied it very often in our study although it is probably not the most obvious strategy. Doppler Haider et al. [DHSP+17] also found that elimination is often used when people work with visualisations. People frequently exclude elements from consideration to make the problem space more manageable [NS72].

Other strategies seem to be more specific to the task of detecting autocorrelation - especially analysing transitions, which is the most obvious strategy to use when trying to detect spatial autocorrelation. Nevertheless, it was not the most prominent strategy in our experiment and also not more successful than the other strategies. This was a very surprising result. This might be due to the fact that the participants did not apply this strategy correctly, but more research is necessary to reinforce this interpretation.

5.5.9 Performance and confidence

The quantitative analysis also provides interesting results. The performance of the participants was influenced by level of autocorrelation (high vs. low autocorrelation), but not by map size. This is surprising as we have deliberately designed our line-ups to control for this effect using empirically-informed prior knowledge [BDM+17]. We assume that the Moran's I levels low ($I = .3$) and high ($I = .8$) are equally difficult. That we find a lower success rate associated with the condition of high autocorrelation may suggest that Beecham et al.'s [BDM+17] modelling might have overrepresented the effect of statistical intensity, so that the staircase procedure used in Beecham et al. [BDM+17] disproportionately improves the ability to detect differences in the high autocorrelation cases. Another explanation might be that the 3×3 line-up design is more challenging than the comparison of two maps where the level of statistical effect is low.

There is one approach that is helpful to solve map line-up tasks. This is analysing the material in detail and spending more time on the tasks. This is probably an indication of increased motivation. This also shows that challenging map line-up tasks require the users to study the material repeatedly for a considerable amount of time.

There was no relationship between confidence about the correctness of the results and the performance of the participants. The participants apparently had difficulties to assess the correctness of their solutions, especially when the tasks were very difficult. This probably indicates that the participants need more training in detecting autocorrelation.

Map design

We compared choropleth and centroid-dot maps assuming that centroid-dot maps would be associated with higher success rates, especially in the more irregular geographies, as they negate certain artefacts inherent to choropleth maps. Contrary to our expectation, participants performed significantly better with choropleth maps. This might be due to the fact that people are better acquainted with choropleth maps and feel more comfortable with them. Subjective ratings of the participants indicate that they prefer choropleth maps to dot maps. This could be addressed by adapting the centroid-dot design, e.g., varying size of dots, prominence of region borders or different colour palettes.

Level of spatial autocorrelation

The different autocorrelation levels (low and high) were received with mixed feelings. On the one hand participants liked the higher autocorrelation plots, calling them *"more pleasing"* just to realise that it made the task not easier because *"all look very good now"* (the highest autocorrelated plot was not sticking out). In the high autocorrelated conditions they would quickly find one or more possible candidates and while looking at each plot they would add more and more to this subset, ending up with a long list of possible answers. When participants try to narrow potential candidates down to one solution *(Elimination)* they compared pairs and would often have difficulties to decide. In situations where they were stuck, we could observe strategy changes, e.g., from looking at dark colours to light ones. A participant said *"If I look at the light colours for once, at least one area is connected here and maybe, therefore, this is the better choice"*. In some cases this yielded better results, like when the number of light clusters was smaller than the dark ones, compare, e.g., plot number 6 with the correct plot number 2 in Figure 5.27. When they could not decide at all, they named the remaining choices as possible alternatives.

We observed higher discomfort in the participants when confronted with the lower autocorrelated line-ups. Participants did not like them at first sight because they found them chaotic and could not make out any clusters or good transitions, i.e., the two most frequently applied strategies could not be employed. One participant said, for example, *"I don't like these chaotic one's, I see nothing and it is very painful to look for any autocorrelations"*. Nevertheless, once the participants applied other strategies and started to look at each plot in detail, they could effectively use *elimination* to come to a decision more efficiently than they would have thought, especially with the centroid-dot map. This aligns well with the results for the group comparison of lower and higher correlated line-ups, where performance of the centroid-dot maps is higher in the grid conditions as well as the small regular condition (compare Figure 5.36).

5.5.10 Future work

Further investigations are necessary to clarify the importance of the order of analysed plots in the line-up, as well as to study the number of plots and arrangement might influence the applied strategies. Analysing the order of strategies – left to right, or top to bottom or scan then check – seems like an important next step. We might find that performance is better if the outlier is checked first as a baseline. Using eye-tracking in a quantitative study would facilitate this. Similarly to Beecham et al [BDM$^+$17] we also found a great variability among participants with respect performance and strategies used. The reason for this is not clear.

It would also be interesting to investigate whether people use different strategies for different levels of difficulty of the tasks or different levels of autocorrelation. We also noticed that the confidence about the solution deviated from the performance. Participants were sometimes very confident about incorrect solutions. They were, for example, distracted by large clusters in the middle of the map or by cases where meaningful *Figure* occurred. Training might help to overcome this problem. In general, it is an open question as to whether training can help to improve user performance in challenging map line-up tasks.

We would like to point out that the tasks participants had to solve were quite hard. In many cases it was impossible to detect the correct solution at a glance, but participants had to study the map line-ups repeatedly. In this process they applied several different strategies in succession. This conforms to research from everyday thinking indicating that people do not adhere to fixed algorithms when solving problems, but use strategies pragmatically and apply them depending on the context [Wol01]. In the case of solving map line-up tasks this approach might be inappropriate. Future work should try to clarify the issue whether combining different strategies is helpful in the context of map line-ups or not. Our experiment indicates that challenging tasks in map line-ups are difficult to solve. Drawing statistical inferences under such conditions is not a straight-forward task.

5.5.11 Conclusion

We conducted a study of graphical line-up tests to investigate how statistical inferences are drawn from data graphics. In this context, we are especially interested in the users' sense-making strategies. We conducted an in-depth experiment with 19 graduate students with choropleth and centroid-dot maps. We could identify six strategies users adopt to solve challenging line-up tasks. Some of the strategies are comparable to strategies used when interacting with other visualisations (*Clustering* to *Looking for connections*, *Elimination* and *Storytelling*). There is one strategy that is specific to the interpretation of map line-up tasks (*Transition*).

We also investigated factors influencing performance in map line-up tasks, but, as of yet, there is no clear evidence which factors influence the quality of the solutions. In general, the performance of participants for the challenging map line-up tasks was fairly weak. This indicates that a consistent interpretation of statistical structure for challenging map line-up tasks is not a straightforward process. It certainly requires more contemplation than the quick visual judgements associated with some visual tasks, and whether or not training has an effect on performance remains an open question. Finally, we compared choropleth maps and centroid-dot maps. In general, choropleth maps were superior to centroid-dot maps. We think that our results and the analysis of users' sense-making strategies provides valuable information for designers of (geo)visualisations.

The three main contributions of the study are:

- An exposition of the cognitive strategies users adopt when performing map line-up tasks. We identify six cognitive strategies – most are strategies generalisable across visualisation types, but one is specific to geo-spatial data.
- Findings on the factors influencing performance of map line-up tasks. These factors may result from differences in the stimuli (map size, low/high statistical intensity) or from the strategies the participants adopted (cognitive strategies, time spent on tasks). The most important factor influencing performance is time spent on a task.
- Insight into the role of geovisualisation design in influencing task performance. We compared choropleth maps and centroid-dot maps to find out which of the two supported more reliable judgements of statistical structure in maps. We assume that centroid-dot maps would be associated with more reliable interpretations since visual artefacts resulting from different sizes and shapes of spatial units are minimised. Our investigation indicates that this general assumption is inaccurate. Overall the performance was better when using choropleth maps, but there were some line-ups of lower autocorrelations (*regular grid*, compare Figure 5.31) which could be better solved with the dot-centroid map.

5.5.12 Comments on dissemination

This study was conducted in co-operation with project partners from City University London and a manuscript is prepared for submission [DHBDP19]. My responsibilities included the planning and execution of the study. The idea is based on Beecham et al.'s paper *Map Line-ups: Effects of spatial structure on graphical inference* [BDM+17]. We performed the qualitative analysis and co-operated on the interpretation of the results.

CHAPTER 6

Outcomes

Research on sense-making goes beyond discovering problems in working processes and usability of a system. It digs deeper into explaining usage behaviour – why a tool is used in a certain way – and the reasoning processes of the user. Sense-making has been investigated occasionally in information visualisation and visual analytics, but there is little research about the specific sense-making strategies that users adopt. This kind of research is essential for the design of appropriate visual analytics systems not only in this domain, but also in general.

This thesis describes general sense-making strategies similar to the ones found in the literature, but also task-specific sense-making strategies that were formulated based on our observational studies. We observed that strategies applied with simple systems could be transferred to complex ones. We used realistic tasks, that were created in collaboration with end users, and students with the background of computer science as participants. Additionally, we could assemble a group of experts for an investigation with the goal to evaluate the suitability and usefulness of the system.

Our studies address the following research questions (RQ):

RQ1: How can the design of visual analytics tools support cognitive processes for intelligence analysis?

RQ2: What kinds of sense-making processes get employed in visual analytics?

RQ3: What kinds of interaction processes should be supported by the system?

RQ4: How useful are models of insight generation and does the representation of the problem influence insights?

The results of our investigations indicate how intelligence analysis can be supported by the type of visualisation (RQ1) and how the analysis of sense-making strategies can help to inform system design (RQ2). We derived design recommendations from the investigations, which are described at the beginning of each investigation. To answer RQ1 we provide general design implications in section 6.1. RQ2 is addressed in section 6.2 where we present a framework of sense-making strategies on the basis of our empirical investigations.

The first two studies investigate simple systems, analysing different kinds of visualisations and the transformations between a node-link diagram and a matrix. In the third and fourth investigation the interaction with visual analytics systems is used to identify sense-making strategies, once in a dual visualisation system and once in a complex system with multiple components. The final investigation addresses map line-ups in more detail, comparing two map designs with a task coming from the geography domain.

In the spirit of the evaluated visualisations in investigation I3 and the resulting design recommendation 3.3., we map the research questions to the studies using two techniques, a node-link diagram (Figure 6.1), and a matrix (Table 6.1). The matrix gives a structured overview and the force directed layout algorithm puts the node with the highest in-betweenness in the centre. Note that research question 2 (RQ2) was addressed by most of the studies.

Results from the first investigation clearly show that the 2D map, which is the predominant visualisation for spatial information, is not suitable for the representation of temporal data. The Gantt was less successful than the line chart and the matrix, which yielded the smallest error rate. The matrix was also the preferred visualisation by the participants, thus, we suggest that matrices can give a good overview of temporal data. In the line chart errors were made in the more busy areas where more people meet. There the technique of using pie charts to counteract occlusion was not successful.

The second investigation on node-link diagrams and matrices has direct implications on the design of these visualisations, such as using a clockwise or counter-clockwise arrangement of nodes in node-link diagrams. Results suggest that the mental transformation from node-link diagrams to matrices is cognitively less demanding than vice versa, i.e., it was harder to draw a matrix on the basis of a node-link diagram.

	I1	I2	I3	I4	I5
RQ1	x	x	x		
RQ2		x	x	x	x
RQ3			x	x	x
RQ4	x			x	x

Table 6.1: Mapping of research question RQ1-4 and studies I1-5 in a matrix representation.

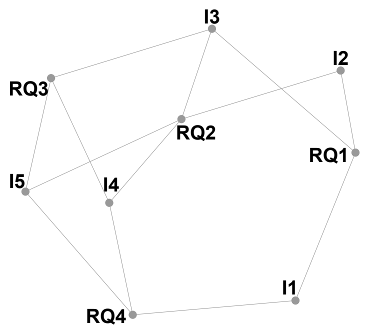


Figure 6.1: Node-link representation of Table 6.1 using a force-directed layout algorithm.

Results from the third investigation I3 suggest to show developments over time at a glance instead of forcing the analyst to jump from one visualisation to another (again and again). An important point the study makes is that participants used the combination of two visualisations for verification purposes making sure that their conclusions from one visualisation hold true under consideration of the other visualisation. This indicates that the juxtaposition of two visualisation of that type can be used to support *Verification* strategies. These are examples for the usefulness of analysing sense-making strategies for design purposes, here, systems for intelligence analysis.

An overview of the design recommendations per investigation I1-5:

1. Identifying location-based meetings (I1).
    1.1. Avoid using maps for the representation of temporal data
    1.2. Use maps in combination with other visualisations
    1.3. Use auxiliary lines carefully
    1.4. Avoid white space between entities of the same group
    1.5. Show duration of meetings explicitly
    1.6. Consider the influence of visual properties
2. Transferring mental map from node-link to matrix and vice versa (I2).
    2.1. Use clockwise or counter-clockwise arrangement of nodes
    2.2. Use $\times$ symbol to represent one connection type
    2.3. Use the same representation for the different connections types in both visualisations
    2.4. Avoid semantic meaning of the order of connections
    2.5. Represent the reading direction of connections along the left side
    2.6. Use intra-cell representations for showing second-degree neighbours in matrices
3. Assessing temporal evolution in node-link vs. matrix representations (I3).
    3.1. Show connections of a social network
    3.2. Show temporal evolution in matrix representations
    3.3. Show data in different views
    3.4. Support *Trend Analysis*
    3.5. Support *Storytelling*
    3.6. Support *Elimination*
4. Analysing crime with the VALCRI prototype (I4).
    4.1. Offer simple visualisations

4.2. Support *Verification*

4.3. Motivate *Trend Analysis*

4.4. Motivate using complementing views

4.5. Emphasise training of the tool use

5. Assessing spatial autocorrelation with map line-ups (I5).

   5.1. Use choropleth maps to show spatial autocorrelation

   5.2. Train participants on the importance of transitions

   5.3. Train participants in spending more time on a line-up task

   5.4. Use dot-centroid maps carefully for lower autocorrelated data and more regular geographies

Through the thinking aloud protocol and self-reported insights we investigate the analysis process in investigation 3 and 4. To address RQ3 we chose challenging tasks, which have no clear answers. In this way we could observe interaction processes and the level of the resulting insight. Notably is, that interacting with different tools induced strategy changes.

Our research findings concerning insights (RQ4) confirm Friel et al.'s [FCB01] three level model for making sense with graphs. We found low-level insights reporting simple facts without explanations (which goes in line with (1) reading the data), moderate insights reporting explanations (which aligns to (2) reading between the data), and high-level insights reporting recommendations (which equals (3) reading beyond the data). We think the analysis of tool use and interactions related to reported insights is a promising approach. With future work we aim to retrieve interaction patterns, which might differ depending on the quality of insights, i.e., patterns of success or failure in the sense-making and insight processes.

In the final investigation we address patterns in different map designs, which yielded a clear preference for the traditional choropleth design to represent spatial autocorrelation. Other evaluated parameters (size, geometric irregularity and baseline autocorrelation) showed no significant differences. Investigating the success of sense-making strategies proved to be a difficult endeavour. While a variety of strategies was applied per task, the distribution of correct answers resembles those of wrong answers, thus, we cannot say that one strategy was more successful than another. While it is unlikely that all strategies influence the task at the same rate it might be challenging to retrieve their individual influences. We further assume that the difficulty of the task has an influence on the number of strategies. We, thus, suggest to follow up on this kind of research and describe future work in this area.

6.1 Design implications

By combining guidelines from literature (chapter 4) and the results from our investigations (chapter 5) we derive implications for the design of visualisations and visual analytics systems. We can generalise the results of our four research questions by formulating seven recommendations for the design of visual analytics systems.

6.1.1 Provide different perspectives on the data

Data from the end user interviews show that analysts are actively looking for contradictions in a hypothesis. They challenge their hypotheses against contradicting data and they go after each piece of evidence to form a coherent story. To support this a system should provide different perspectives on the data and let the user work with multiple visualisations. A design that supports relationship mapping in the data and form a coherent mental model can be achieved by multiple coordinated views. Multiple coordinated views combine different views on the data in one visual representation like, e.g., Microsoft's Windows Explorer combines an outline view of the folders, a tabular view of the files in the selected folder and a quick details view of the selected file.

Empirical data shows that the ability to see data in multiple linked views significantly speeds up easy as well as difficult tasks in comparison to a single scrolling view [NS03]. The requirement analysis further shows that analysts want to work with different visualisations that support cross filtering. They want to have a visual representation of spatial events on maps and additional information of the events in the investigated area to get ideas in which direction an inquiry can go. A multiple view approach supports that because unexpected or less expected ideas can be shown as well.

6.1.2 Provide open exploration and allow the redefinition of the goal and the methods

Research in everyday reasoning indicates that so-called wicked problems are solved differently than clear-cut problems. Wicked problems have no clear method or path to the solution and it is sometimes not clear how the solution might look like. It is typical for such problems that users explore the problem space and often redefine the problem while working on the solution [Kle13, DS03] A trade-off decision is needed in order to support exploration in a way that does not lead users astray but helps to brainstorm on the one hand, and on the other hand, to focus at the end of the exploration process. The two higher objectives for a better system are saving time and reducing uncertainty. A more concrete goal would be to identify relations between crimes.

A general hypothesis is that a crime is seldom a single event. Search is especially in the context of CCA a particularly interesting topic due to its breadth in functionality. One might be interested to find similarities and look for patterns or find oddities that stand out from the others. Control over threshold parameters should be given to the user as one can explore the data more openly and is able to bring in one's experience in that way. Analysts have different privileges in different systems and sometimes have to ask for access to different databases which results in delays of different magnitudes. Analysts feel the barrier that this procedure brings to their investigation process. Though restricted, a vast amount of data is available and uncertainties as well as gaps (missing data) need to be represented in equal terms. This needs to be addressed in distinct guidelines from other perspectives but is also considered in these sense-making guidelines.

6.1.3 Support holistic sense-making processes by structuring information in a coherent manner

Gestalt psychology indicates that sense-making is a holistic process where structure plays an important role and empirical evidence shows that everyone develops his or her own structure [May95]. Hence, users should be able to generate their own structures of knowledge where this is possible, e.g., by being able to structure network representations such as social networks themselves. However, anything too complicated and time-consuming is unlikely to be used in an environment characterised by significant time-pressure. The requirement analysis supported this because the analysts prefer to start from scratch rather than remodel an auto-generated representation. In addition to this, program response times are important for sense-making and reasoning. Analysts stated that slow programs are very frustrating. Responses like, e.g., search results, should ideally be available to the user at the speed of thought. Otherwise the slower response might influence reasoning, as one might have no time to follow an idea or forget a sparking idea in the meantime [HSP12].

6.1.4 Support reuse and provide graphical histories

Users should be able to reuse previously performed working steps and get a fluid transition between different visualisations to save time. A keyword search on text documents should be transferable onto other data so that results can be textual, but, for example, also spatial data on a map. Furthermore, an investigation can last over days and weeks and an analyst might end up looking for the same thing over and over again.

To provide a graphical history on the one hand improves a quick reuse of previous searches but also provides an overview over the steps taken. In addition to facilitating analysis and communication, graphical history may help in teaching analysis by example [HMSA08]. The problem with visual histories is that they do not scale. It might be faster to do a new search than to look up an entry in a long search history. For example, the history in current web browsers is implemented as a simple list with every site you visited in the past. Since this list quickly gets long in a short period of time, searching for a specific site can become a time-consuming task. An entry in the history must be identifiable and in reusing a previous search it must be visible what the result set means. To provide a graphical history is a challenging task but users appreciate the feature if they can reuse done work.

6.1.5 Provide interaction possibilities

Providing interaction techniques supports sense-making. A single image can only answer a small set of question. Visual analysts gain insight by repeatedly looking at the data, exploring, refining the views iteratively and are consequently developing insights [HSP12]. The theory of *Distributed Cognition* assumes that interaction with cognitive tools (e.g., the VALCRI system) plays an important role for sense-making processes [HHK00].

Semantic zooming is one interaction technique that should be provided for temporal and spatial data to enable the exploration of more details as a chart is enlarged in the area where the zoom happens. The granularity in time can vary from years to weeks to days, time of day, etc. The granularity of spatial data occurs on different levels such as zip code, grid, and address – with address being the most common practise in crime analysis. Geocoding is a critical issue because the visualisation reveals information that is not obvious by looking at a street name or address. However, using coordinates from GPS data is the most accurate and reliable way to locate incident data, especially when they happen in parks or on vacant land [BS13].

6.1.6 Enable the users to cluster similar cases

Another interaction technique is to offer an easy clustering and hot spots generation possibility to show similarities in the data. There are various mathematical methods to cluster data that can support analysts in finding similarities, e.g., in MOs of criminal gangs.

The results of the mathematical methods can be visualised on the screen so that analysts can see at a glance which data is similar. It should be possible for the users to influence the clustering process, for example, by choosing how many clusters should be presented. Therefore, we distinguish between machine and human created clusters. Cluster visualisation can be used for rapid identification of relevant documents. Allan et al. [ALS97] support the cluster hypothesis and show how feedback techniques enhance the effect to help users separate relevant from non-relevant documents. Data can be clustered in different ways, depending on the criteria for the clustering procedure. It is sometimes difficult to identify a clustering procedure that helps the user to gain insights. For example, in node-link diagrams (networks) the users could cluster persons with similar attributes or the users could cluster crimes with a similar MO. It might sometimes be difficult to identify the criteria of what constitutes a similar MO which indicates that the same group of persons is responsible for a series of crimes. In alignment with the guideline to support reuse, users ideally should be able to save a cluster selection for future analysis.

6.1.7 Do not overwhelm the user with too many connections

Displays that visualise too many individual data points become cluttered, which affects analysis because trends and patterns are harder to spot. There are different techniques to measure and reduce clutter in a visualisation [PWaR04]. Crime maps need to convey a lot of attributes in different granularities, like the incident numbers per year (as well as per hour a day, etc.), the crime types, boroughs or boundaries. It is sometimes difficult to identify what constitutes clutter. How much information is needed might depend on the data or the task. Possible solutions to this problem are to enable filtering or highlighting interesting data. The representation of large datasets can be simplified with the help of aggregates, which represent a group of data points so that fewer markers are needed.

A challenge in the construction of aggregates is to choose the right granularity. This is an important decision for the effective visualisation and depends on the application domain and on the task at hand [FNPS99b, WBB03]. The problem with different aggregation levels is that anomalies are possibly hidden [RFF+08]. On the other hand effective maps have to be highly generalised so that important trends and features emerge [Har03]. To address the problem of the unit of analysis exploratory tools should allow the selection and combination of aggregates and give several perspectives on the data. To reduce this threat the system needs to give control to the user and raise curiosity to explore the different levels.

Further, in the context of map visualisations, spatial exaggeration is a relevant factor. The data of interest needs to be exaggerated in order to make connections and contradictions visible. An example that came up during the requirement analysis is that telephone records and the routes of buses might be connected and that an exaggeration of antenna regions may reveal that the mobile phone data is following a certain bus route.

6.2 Sense-making strategies

Designers can also adapt the system according to the sense-making strategies potential users adopt. For that purpose, we exposed sense-making strategies in various contexts and tasks. An overview of the observed strategies is given in Table 6.2. They were analysed with regard to the insights that were generated in the tasks and some conclusions on the efficiency of strategies could be made. *Looking for trends*, for example, yielded higher level insights than *Profiling*, which in that sense was an especially unsuccessful strategy.

We identified nine, eleven and six sense-making strategies in the last three studies respectively. The strategies we identified in the third and the fourth investigation are quite similar. This is an indication that these strategies are more general and apply to different visualisation systems. Nevertheless, the strategies have to be validated with other systems in future work.

We also related the observed sense-making strategies to the number and quality of insights in investigation I3. The following results were the most important: Applying fewer sense-making strategies is correlated with more insights in general and high-level insights, which are the conclusions we want to be able to make with visual analytic systems. When such high-level insights were reached analysts used significantly more verification. Similarly, they analysed more trends and it seems that there is a tendency of looking for trends being favourable for high-level insight generation. *Pattern* and *Profiling*, on the other hand, do not support insight generation very well and mostly lead to low-level insights.

Table 6.2: Strategies derived from the user investigations I3-I5.

<table>
<tr><th>Purpose</th><th>Strategy</th><th>Insight level</th></tr>
<tr><td>Overview</td><td>Looking for connections (I3 and I4) / Clusters (I5)</td><td rowspan="2">Make connections between data points</td></tr>
<tr><td rowspan="2">Elaborate</td><td>Storytelling (I3-I5)
Comparing, Summarising, Laddering (I3)
Colour (I5)</td></tr>
<tr><td>Looking for an increase in criminal activity (I3) / Trends (I4) / Transitions (I5)</td><td>Identify single data points</td></tr>
<tr><td>–</td><td>Creative Desperation (I4)</td><td>Change strategy</td></tr>
<tr><td>Conclude</td><td>Verification (I3-I5)
Elimination (I3-I4) / Elimination incl. trend (I4) / Elimination due to outliers (I5)
Coincidental Aha's (I4)
Contradictions (I4)
Gut feeling (I5)</td><td>Make inferences</td></tr>
</table>

The contribution of our studies, therefore, is twofold:

- First, the analysis of sense-making strategies in a complex system with a mixed approach has not been done before. We investigated sense-making strategies adopted by the participants to interact with a visual analytics system consisting of multiple components. The system enables users to explore large datasets at great length by providing different perspectives on the data. We also studied cognitive sense-making strategies in investigation 2 in a dual visualisation tool revealing user expectations and preference of the visualisations, a node-link diagram and a matrix. We can partly confirm previous results where simple systems were used. They seem to be transferable from simple systems to complex, realistic systems in general, and differences depend rather on the semantics of the study.

- Second, we related sense-making strategies to the insight generation process to identify successful strategies, where *success* is not only measured by the number of insights but also the quality of the insights in investigation 4 and 5. Strategies frequently adopted by users are sometimes not the most efficient ones. Therefore, we wanted to identify strategies generating appropriate results and observed that a faster change of strategies led to more low-level insights and spending more time on one strategy led to more high-level insights.

CHAPTER 7

Critical reflection

This work contributes to the related work in two ways: 1) by developing specific recommendations for visual analytics through the analysis of existing research and 2) by adding empirical results to this research area through five sense-making investigations and, additionally, by describing strategies during tool use.

7.1 Comparison to existing research

The systematic analysis of how analysts make sense with visual analytics tools on a sense-making level is a novel approach and related work in this area is rare. The literature review did not reveal investigations comparable to our empirical work. Existing systems are rather evaluated with regard to efficiency and accuracy of the analysis results. Interactions are studied by relying on automatic analysis methods to find patterns in the user behaviour. Guo et al. [GGZL16], for example, related insights to interaction by using an algorithm to analyse interaction logs, being able to derive only two general design recommendations.

Identifying sense-making processes is challenging. Blandford et al. [BFA14] and Sedig et al. [SPLM16] have addressed this issue in great detail. Early research on sense-making was done by Pirolli and Card (e.g., [PC95, PC05]), but their approach got criticised due to fundamental reasons (such as tool use and user interaction). Since then Klein's perspective on sense-making and the Triple Path Model of Insight [Kle13] was applied to support naturalistic decision making by, e.g., Wong [WV12, Won14], Attfield [AB10, AHW10], or Kodagoda [KPS+17].

Reda et al. [RJLP14] point out that interaction processes with information visualisations are analysed by focussing on the outcomes rather than on the sense-making process itself. Notably, the research by Sedig et al. [SPLM16] is concerned about interactions that occur in conjunction with each other in order to improve the user performance. In another investigation by Sedig et al. [SRL05] could show that linked, juxtaposed views were more effective than single views by addressing the cognitive processes involved to understand change over time.

7.2 Understanding of sense-making

Our first investigation looked at visualisations that support the detection of meetings and our results suggest that the matrix is more effective than a line diagram. A similar evaluation by Tversky et al. [TGC$^+$16] concludes that lines are recommended to track persons. They compared two visual stimuli in matrices, once putting the dots per person next to each other in the cells and once by using parallel lines that connect the respective cells. As a result they describe how different kinds of insights can be gained by the different visualisations. We want to note that their evaluation uses a smaller dataset with four persons and time steps. The question is if their result holds with a three to ten times larger dataset or if the apprehension of the lines will reach their limit as they did in our investigation. Research conducted independently by different groups highlights the relevance of this topic in the community.

In Doppler Haider et al. [DHSP$^+$17] we developed sense-making strategies based on Klein's research [Kle13] and a bottom-up analysis approach and a detailed analysis of thinking aloud protocols. In the third investigation we identified several general sense-making strategies: *Comparing (to find connections), Laddering, Storytelling, Summarising, Eliminating*, and *Verifying.* We also found task specific sense-making strategies: *Increase in criminal severity, Pattern recognition, Relationships,* and *Profiling.* In this study we were interested if matrices can be designed in a way to show connections as good as node-link diagrams and compare the usefulness of two visualisation techniques in real life scenarios. For that purpose we adapted the basic node-link and matrix visualisations to represent structures over time in a superimposed static visualisation. We developed fairly complex, explorative tasks with intelligence analysts to determine whether such tasks yield similar results to previous research using generic tasks. Such realistic tasks would strengthen the guidelines on when to use node-link diagrams and when a matrix representation. However, we found contradictory evidence and we show that principles from previous results are not as clear cut as suggested [DHGP$^+$18].

In the third investigation the evaluated system was a simple one, consisting of two visualisations. In the fourth investigation we identified similar strategies derived from an analysis of sense-making with a complex system. An in-depth analysis of eleven strategies contrasted the descriptive analysis of the strategies with the interactions of strategies over time. The three layer model by Friel et al. [FCB01] conforms to our research findings concerning insights in a complex system (Investigation 4). We found low-level insights reporting simple facts without explanations, moderate insights reporting explanations, and high-level insights reporting recommendations. The switching of strategies during tool use was evaluated quantitatively, and it could be shown that switching strategy often decreased the performance. We observed in our studies that focusing on fewer strategies was related to an increased amount of insights. A possible explanation is that more motivated participants used strategies for in-depth analysis which takes longer, which results in fewer switches as all participants had about the same time for the task. In contrast to that, superficial examinations, what we call "the easy way", is to switch views to see if something pops up" at first sight".

We investigated whether the representation of the problem influences insights by looking at basic visualisations, such as different types of maps or diagrams and advanced analytical tools like the SPC, S3 or CCT. More insights were communicated with the basic representations. Basic representations often have only one job, which is well-known, e.g., maps for spatial data, or timelines for temporal data. Although participants used all tools in their analysis, they reported high-level insights more often with a simple timeline, most moderate insights in timeline and bar chart and most low-level insights in the map. The traditional choropleth map also led to more insights than the centroid-dot alternative. The new design was not convincing and participants strongly preferred the choropleth map. A problem with novel designs might be that participants prefer the more familiar design and longitudinal studies would be of benefit so that users can get more familiar with new designs and tools.

In the final investigation we wanted to elicit sense-making strategies with challenging tasks offering clear answers. The assumption that one strategy would be superior to others could not be proven as we observed similar strategies for correct and wrong answers. The only significant factor leading to success was spending more time on a task. It is likely that different motivation levels affect sense-making, thus, more motivated participants perform better in difficult tasks. We assume that participants who studied the tasks longer were more motivated. Another idea is to determine which strategy was most dominant per task or which strategy had most influential power. Additionally, it might be possible to gain more insight on the sense-making processes with the line-up task, which we used in our investigation for the detection of patterns of higher spatial autocorrelation. These results do not rule out the influence of other factors in the use of strategies and we, therefore, see a definite need for future work in this area.

7.3 Development of guidelines

Guidelines for the design of visualisations have been questioned critically in the past [PRS$^+$94]. We think the value of guidelines lies in their application, and yet oftentimes guidelines are very general and not easily adaptable for specific use cases. Many guidelines stop at a certain level because they shy away from the problem of not being valid ubiquitously or solid enough to hold in various situations. The more specific individual guidelines get they more likely they will fail to satisfy the tricky problem of managing design trade-offs. From an HCI perspective the application of very general guidelines sometimes poses serious problems. One of the eight golden rules defined by Ben Shneiderman [Shn10] is that systems should provide informative feedback. This is certainly sensible advice but difficult to achieve because it is not always clear what informative feedback really is. Because the guideline is quite general it is sometimes difficult to apply it in a specific context at hand. It makes sense to break down such guidelines into more specific ones which specify how to achieve the goal defined in a general guideline more clearly. There are similar problems with respect to the design of visual analytics interfaces. The guideline specifying that the identification of connections by the users should be supported is certainly sound but, at the same time, also difficult to realise because this can be done in many different ways. Such a general guideline does not provide system designers with concrete advice how to develop systems for intelligence analysis. There are many different methods which could be used here, but it is rarely clear which methods fits a certain context best. Methods like multiple views combined with linking and brushing or the usage of network visualisations proved to be useful in general. Again, it is necessary to interact with prospective users to make sure that these methods are adapted to their needs and requirements. Typical for applications in visual analytics is the necessity of trade-off decisions.

Individual guidelines might sometimes be contradictory. One guideline might have consequences which prevents the realisation of another. One example from the VALCRI guidelines concerns the issues of privacy versus providing the user with as much related material as possible. Privacy principles sometimes prevent access to required information which can help to identify connections between information and patterns in the data. On the other hand, privacy is an important issue, and sensitive data has to be treated with care. In this context legal considerations have to be taken into account. Within these legal constraints, access has to be provided to all the relevant information. Guidelines concerning the issue of filtering data could be also problematic in terms of the occurrence of confirmation bias. Such a selection of data can be based on initial thoughts, hypotheses, but also unconscious prejudice, which may lead to wrong decisions.

Sometimes trade-offs are necessary when applying guidelines [HPH+nt]. One guideline which can be derived from the sense-making theory of Klein [Kle13] is to support the identification of connections. This can be supported by providing the user with access to a large amount of related material to enable the users seeing connections in the material. This represents a trade-off since users in intelligence analysis usually only have a limited amount of time to come to a conclusion about some specific issue. If they are overwhelmed by a large amount of information this will become increasingly difficult. Apparently, there is a trade-off between supporting the users to identify unexpected but relevant connections and overwhelming them with information. The question here is how much information is enough to enable the user to come to valuable insights in a reasonable amount of time, but still providing him or her with enough information to be able to see unexpected patterns in the data. This problem can only be solved by a detailed analysis of the work environment of the user. Designers have to know the time frame in which analysts have to come to a conclusion and the possible importance of finding unexpected patterns. They then have to weigh these two necessities to come to a viable solution. In the context of cognitive biases and their mitigation, the application of a guideline that is used to mitigate a specific cognitive bias might induce another one. Taking as example the guideline on providing the same information in different visualisation formats. The underlying idea is that presenting the same information from different perspectives can mitigate the confirmation bias, as it allows for building alternative hypothesis on the same background information. However, when using visualisations dependent on the visualisation format, a framing effect can occur if the same information is visualised differently.

Despite the problems encountered when using guidelines, they are a very valuable tool for designers that can help to improve the design of interfaces tremendously. An advantage of dealing with guidelines during the design process might be a better understanding of the characteristics of the specific design problem. A thorough understanding may emerge over time through iterative analysis (repeatedly looking at user studies) and synthesis (advancing the prototype design).

7.4 Limitations and future work

Our explorative studies mainly yield first tentative results. More systematic investigations are necessary in the future. One limitation when investigating complex, realistic systems is that a formative evaluation will never be able to simulate realistic usage conditions. Long-term investigations in which users are acquainted with the system will probably yield different results. Such a long-term investigation is not possible within the context of a formative evaluation. Possible topics for future research would be, for example, to investigate how to support the *Verification* strategy and looking for *Trends* where applicable). In the cognitive bias literature some suggestions have been made, but more research in this area is still necessary.

The last study aimed to identify sense-making strategies in line-up tasks in a qualitative manner. We were successful in exposing sense-making strategies that were used in the line-up tasks, but the relationship between usage of strategies and performance is not straightforward. We noticed that participants adopted several different strategies for the same task, while whether or not such combinations are beneficial remains unclear. The analysis of decisive strategies is a natural next step.

Thinking aloud is a good method to analyse thought processes while a task is solved or a specific tool is used. It can show up what participants have in mind while they are working [BR00, ES93]. However, this information cannot be complete because humans are not thinking about everything they do and look at, and only verbalise certain information at a time. As a consequence, the thinking aloud method has its limits. The analysis of our verbal protocols indicate, for example, that reading directions of line-ups were employed in different ways. The starting position was described by some participants: if the top most left plot or the centre position was more investigated than other. One participant was concerned, for example, that she was always coming back to the centre. The participant stated *"I don't really know why I decided for plot 5, maybe because it's in the middle"*. From the protocols we could observe starting from the top left to the right line-by-line, chaotically jumping between plots as well as pairwise comparison between likely candidates for the answer. Further analysis of such strategies requires another method, such as eye-tracking, to see in which order the plots get fixated. Although eye-tracking was suggested to complement thinking aloud, we think the verbalisation is influencing eye gazes in a way that alters the interpretation of the tracked eye movements. Therefore, we suggest to rather use eye-tracking in isolation and combine the results with the results from thinking aloud studies. This additional information might provide more insights on the specifics of the strategies, which are used to solve line-ups and help to determine their influence on the task performance. It would also provide detail on whether the position of the target plot makes a difference when judging spatial autocorrelation in map line-up tasks.

CHAPTER 8

Conclusion

The aim of my thesis was in one sense to make the invisible visible. First, we uncover the knowledge about sense-making with visualisations that is already contained somewhere hidden within research results. By making this knowledge visible and understandable for system designers, they can readily incorporate this knowledge in their design decisions. This first empirical part constitutes a guideline catalogue in the form of design actions, which can be found in the Appendix. Second, we try to shed more light on sense-making processes with five empirical studies. We investigate the insight model of Klein [Kle13] by studying sense-making with the thinking aloud protocol and, thus, provide dearly needed empirical data on the topic of visual analysis by looking at the user interaction with traditional visualisations [MSSW16]. The results yield further design recommendations as well as a general set of sense-making strategies that can be applied in future work.

We could identify that *looking for connections* and more generally, *patterns*, is a prevalent activity in the explorative analysis process. This allows to conclude that humans naturally thrive to gain insights by looking for new connections in the data. There are different ways in which such connections can be shown. Social networks or other network-like visualisations proved to be suitable for emphasising connections in the data. However, such visualisations can easily get too large for standard screen sizes. If the network contains too many nodes and links, it is difficult to identify connections between single entities. Interaction, as suggested in the literature, is key to overcome this problem. Thus, a considerable amount of research has been published trying to explain user interaction [WGK10, PSM12, Tom15, PWK16]. A well-known example of interaction for clutter reduction is to enable filtering. This also enables the user to look at the data from different viewpoints, which supports making new connections and, therefore, the insight generation process. Another possibility to show connections is via multiple views, which also show the data from different points of view. Here, we face again the scalability problem as each view needs to fit on the screen. These examples show that there are up- and downsides to each decision, inevitably leading to trade-off decisions. To be able to make informed design decisions we collected this kind of information and put it in a concise form, readily available to designers and developers. The goal of the format is to provide straightforward information and references to further reading.

The research of Klein [KMH06a, KMH06b, Kle13] shapes our understanding of sense-making and mainly influenced our research questions, while it was not clear how to support seeing *Connections, Contradictions* and overcoming *Creative desperation* at the same time. Klein points out that contradictory evidence can be a powerful motivation for getting novel insights. We describe ways from literature how contradictory evidence can be made visible, such as patterns from semi-automatic analysis methods and investigate how such patterns can be made visible in a visualisation, e.g., by overlaying this information on a map. A human analyst is necessary to assess the significance of these patterns, and if the story the patterns tell makes sense or if new explanations and insights are needed. It is not always clear how the human in the loop is best supported by visualisations since literature suggests different theories of the analysis process, none of which fully describe the processes of visual analytics. Our investigations on sense-making strategies in the domain of criminal intelligence analysis contribute another piece to the puzzle of how analysts think and how they can be supported by design.

Appendix

Guideline Framework

ID	G1 Contradictions
Title	Identification of contradictions between observed patterns and ideas that just don't make sense.
Description	The system should identify differences between observed patterns (e.g., in the modus operandi of criminals). Visualisations could be used to present these differences to the users for inspection (human in the loop). Analysts working with a cognitive Tool like VALCRI should be supported in detecting stories that don't make sense. Contradictions warn the analysts that something is seriously wrong with the stories we are telling ourselves. More importantly, it tells us that we have weak anchors – important beliefs that are not fully supported.
Trade-off (optional)	If the system emphasises contradictions it might be difficult for users to reach a coherent conclusion.
Context	3.4 Patterns, trends, relationships, outliers detection 18.2 Volume Crime-Related Tasks
Theory	Gestalt principles, e.g., the law of proximity says that objects that are near to each other are seen as belonging to each other. The law of similarity says that objects that are similar to each other are seen as belonging together. Gestalt principles are powerful tools to indicate differences and similarities to users. See, e.g., Pinker 1999 [Pin90], Carpenter and Shah 1998 [CS98], Klein 2013 [Kle13]
Example	Deviations from normal day-to-day activities which can be interpreted as indication of suspicious behaviour. Contradictions cause the analyst to think, "that can't be right! That doesn't make sense!".

ID	G2 Clustering / hot spots
Title	Enable the users to cluster similar and different cases.
Description	The system should enable easy clustering and hot spots generation. Visualisations can be used to achieve such clustering.
Trade-off (optional)	Different granularities can result in clutter (compare G3.2.4 Clutter). Data can be clustered in different ways, depending on the criteria for the clustering procedure. It is sometimes difficult to identify a clustering procedure which helps the user to gain insights.
Context	13. Data Annotations / Data Extraction / Metadata 26.3 Link/Network Charts
Theory	Allan et al. 1997 [ALS97]
Example	Clustering is a possibility to show similarities in the data. There are various mathematical methods to cluster data. Such methods can support analysts in finding similarities, e.g., in MOs of criminal gangs. The results of the mathematical methods can be visualised on the screen so that analysts can see at a glance which data is similar. It should be possible for the users to influence the clustering process (e.g., too choose how many clusters should be presented), so that different views on the data are shown. In node-link-diagrams, e.g., actors with similar attributes could be clustered or the user could be enabled to cluster crimes with a similar MO (comment: it might sometimes be difficult to identify the criteria for what constitutes a similar MO which indicates that the same group of persons is responsible for a series of crimes).
Evaluation Criteria	Clustering can be enabled in various ways. Users could be given the possibility to let the system perform clustering algorithms or similar mathematical procedures. It is also possible to let the users cluster data manually in a visual form (e.g., in a social network). The evaluation criteria would be whether such possibilities exist in the system or not.

ID	G3 Facilitate the making of connections
Title	Indicate connections in the data in an appropriate way and show new or other pieces of information and non-information.
Description	The system should indicate connections between data in an appropriate way. Connections between people, e.g., can be shown as social networks. Such networks can indicate connections between people, which are not immediately obvious. The system should enable the user to explore such networks. The system should help the analyst distinguish between dots, non-dots, anti-dots, and other ambiguities. "Dots" are information that is relevant to the problem; "non-dots" are information that are irrelevant messages; "anti-dots" are data that are different, contradicting, conflicting, such as situational information that may make the ideas seem less likely, though perhaps plausible. These new connections could create new ideas or open up new pathways to consider or investigate.
Trade-off (optional)	Too many connections are confusing and lead to clutter. Restriction of screen size
Context	3.4 Patterns, trends, relationships, outliers detection 13. Data Annotations / Data Extraction / Metadata 14. Data Presentation/Briefing/Reporting 26.1 Visualisation
Theory	Klein et al. 2007 [KPRP07], Klein 2013 [Kle13]
Example	E.g., show connections between people as social networks.
Evaluation Criteria	Does multiple views and linking and brushing exist or not? Does the system support the visualisation of and interaction with social network representations? These are just a few evaluation criteria.

ID	G4 Facilitate the identification of coincidences
Title	Enable the analyst to re-organise and present data in multiple ways that can reveal repetition or co-occurrence.
Description	The visualisation and the interaction techniques should enable the analyst to see repeating events, patterns or co-occurring similar events, in context. These are coincidences and associations we don't fully understand; relationships we cannot (yet) articulate – but they look like they are related, somehow. This can cause us to break from initial stories or to break away from initial anchors, and to initiate alternate stories.
Trade-off (optional)	
Context	3.4 Patterns, trends, relationships, outliers detection
Theory	Klein 2013 [Kle13], North & Shneiderman 2005 [NS03]
Example	Use multiple views
Evaluation Criteria	Does multiple views and linking and brushing exist or not? This is just one possible evaluation criterion.

ID	G5 Presentation of unexpected/less expected ideas.
Title	Enable the analyst to persistently see the ideas created by combining and grouping data.
Description	The visualisation and the interaction techniques should enable the analyst to be persistently exposed to other ideas, unexpected ideas, and less-expected ideas.
Trade-off (optional)	
Context	3.4 Patterns, trends, relationships, outliers detection
Theory	Klein 2013 [Kle13]
Example	Use multiple views.

ID	G6 Facilitate the perception of curiosities
Title	Raise curiosity in the analyst for unusual single events that stands apart from the trend, such as an outlier.
Description	Curiosities make the analyst wonder about something. This reaction is not the insight, but starts the person on the road to gaining insight. Curiosities are sparked by a single event, while coincidences are observed as a repetition of pattern.
Trade-off (optional)	Curiosities might distract analysts from other important patterns.
Context	3.4 Patterns, trends, relationships, outliers detection 25. User Interface / System Functionalities
Theory	Klein et al. 2007 [KPRP07], Klein & Jarosz 2011 [KJ11], Klein 2013 [Kle13]
Example	Curiosities cause the analyst to think, "what is going on here?" and to comment, e.g., "that's funny" / "that's strange" / "that's odd". Visualisation should emphasise outliers or other unusual data.

ID	G7 Provoke questions
Title	The visual representation should provoke the analyst to ask questions.
Description	The intention is to create situations that prompt the analysts to ask questions such as, "That's interesting - why did that happen?" or "why are there missing elements?"
Trade-off (optional)	How to point out the unknown unknowns?
Context	3.4 Patterns, trends, relationships, outliers detection 13. Data Annotations / Data Extraction / Metadata 25. User Interface / System Functionalities 26.1 Visualisation
Theory	This idea is based on Klein's Data-Frame Model ("Question a frame"), and Klein's Triple Path Model of Insight (Klein, 2013 [Kle13]; Klein & Jarosz, 2011 [KJ11]); Asking questions see, e.g., Wong & Varga, 2012 [WV12]
Example	"Blackholes" or showing gaps where data is missing in a series.

ID	G8 Holistic sense-making processes
Title	Information should be structured in a coherent manner to support holistic sense-making processes.
Description	Gestalt psychology indicates that sense-making is a holistic process where structure plays an important role. There is empirical evidence that everyone develops his or her own structure therefore it makes sense to allow users to generate their own structures of knowledge where this is possible.
Trade-off (optional)	Anything too complicated and time-consuming is unlikely to be used in the environment characterised by a significant time-pressure.
Context	11.2 Data Processing / Management 13. Data Annotations / Data Extraction / Metadata 14. Data Presentation/Briefing/Reporting 16.2 Organizational Knowledge Management 25. User Interface / System Functionalities Ability to freely organize different views
Theory	Pretz et al. 2003 [PNS03]
Example	Users should be able to structure network representations (e.g., social networks) themselves if it makes sense in the context of their work.

ID	G9 Open exploration
Title	Allow for open exploration and the redefinition of the goal and methods.
Description	Research in everyday reasoning indicates that so-called wicked problems are solved differently than clear-cut problems. Wicked problems have no clear method or path to the solution and it is sometimes not clear how the solution might look like. It is typical for such problems that users explore the problem space and often redefine the problem while working on the solution. To support such problem solving processes, systems need to be open and allow the redefinition of the goal and the methods to reach a solution.
Trade-off (optional)	Exploration should be supported in a way which does not lead users astray but helps to brainstorm on the one hand and to focus at the end of the exploration process on the other hand.
Context	3.1 Manual, semi-automatic, and automatic analysis Be uninhibited in creatively exploring the data 11.2 Data Processing / Management 16.3 Personal Knowledge Management
Theory	Davidson & Sternberg 2003 [DS03] , Klein 2013 [Kle13]

ID	G10 Facilitate the identification of assumptions
Title	Enable the analyst to identify and change assumptions that lock-in.
Description	In some situations, we can become fixated on or trapped by our assumptions or habits that are the basis of our understanding of our world or events in that world.
Trade-off (optional)	
Context	3.4 Patterns, trends, relationships, outliers detection 13. Data Annotations / Data Extraction / Metadata 14. Data Presentation/Briefing/Reporting 25. User Interface / System Functionalities 26.1 Visualisation
Theory	Klein 2013 [Kle13]

ID	G11 Graph design
Title	Design of simple graphical representations.
Description	There is a systematic body of research on the design of simple graphs, including guidelines that are simple to apply. They can be easily used for the design process of simple graphical representations.
Trade-off (optional)	
Context	14. Data Presentation/Briefing/Reporting 26.1 Visualisation
Theory	See e.g., Kosslyn 1989 [Kos89]
Example	Line graphs are, e.g., usually interpreted as trends even if the data does not represent trends. Lines are better able to represent values than other visualisations.

ID	G12 Time
Title	Use different timelines for the visualisation of processes in time.
Description	Linear timelines are only one possibility to present temporal data. To describe recurring events, circular visualisations are used to enable the users to identify recurring patterns easily. The system should probably recommend a certain visualisation, which is most appropriate for the data.
Trade-off (optional)	The design of several co-occurring patterns is a special challenge. Another challenge in this context is the fact that time has different and irregular granularities (years, months, weeks, days, hours etc.). Formats for dates vary for different countries.
Context	3.4 Patterns, trends, relationships, outliers detection Automatically determine similarities and highlight differences in MOs from same crime types, different crime types, different times (season variation), regions, etc. 26.1 Visualisation
Theory	Johnson-Laird 1996 [Joh96], Keim et al. 2010 [KKEM10], Andrienko et al. 2010 [AAD+10]
Example	There might be more crimes at specific times/days of the week, which rather pops out in a circular visualisation than in a linear one.
Evaluation Criteria	Are there any visualisations representing time? Are cyclical processes supported by specific visualisations? Can users interact with the visualisations?

ID	G13 Clutter
Title	Do not overwhelm the users with too many connections.
Description	Possible solutions for this problem could be filtering or highlighting interesting connections. Filtering mechanisms should be appropriate for the user's task.
Trade-off (optional)	It is sometimes difficult to identify what constitutes clutter. It depends on the users and their tasks how much information they need.
Context	13. Data Annotations / Data Extraction / Metadata 14. Data Presentation/Briefing/Reporting 26.1 Visualisation
Theory	Peng et al. 2004 [PWaR04]
Example	Crime maps need to convey a lot of attributes in different granularities, like the incident numbers per year (as well as per hour a day, etc.), the crime types, boroughs, boundaries, etc. Automated clutter reduction techniques could be used in combination with hierarchical data visualisation. Spatial exaggeration is a critical factor. The data of interest needs to be exaggerated in order to make connections and contradictions visible. Telephone records and the routes of buses might be connected and an exaggeration of antenna regions may reveal that the mobile phone data is following a bus route.
Evaluation Criteria	Use an appropriate measure of visual clutter.

ID	G14 Interaction processes
Title	Provide interaction to support sense-making.
Description	The theory of Distributed Cognition assumes that interaction with cognitive tools (e.g., the VALCRI system) plays an important role for sense-making processes. These interaction processes should be designed carefully.
Trade-off (optional)	This can mean that not all technical possibilities should be implemented and better be left out.
Context	11.2 Data Processing / Management 25. User Interface / System Functionalities
Theory	See, e.g., Hollan et al. 2000
Example	When to offer filtering; Which attributes are used for filtering, etc.
Evaluation Criteria	Does the visualisation support various interaction processes?

ID	G15 Design of node-link diagrams
Title	Design node-link diagrams so that users can develop a coherent mental model of the relationships in the data.
Description	Node-link diagrams can represent connections in the data. They should be designed in a way so that significant relationships in the data can easily be perceived and processed. The following design recommendations should be adopted: - reduce crossings - cluster nodes which belong to the same group - highlight important nodes, put them on top or in the centre - preserve mental map in a dynamically evolving node-link diagram to increase memorability
Trade-off (optional)	
Context	3.4 Patterns, trends, relationships, outliers detection 14. Data Presentation/Briefing/Reporting 25. User Interface / System Functionalities 26.1 Visualisation 26.3 Link/Network Charts
Theory	Huang et al. 2006 [HHE06], Archambault & Purchase 2012 [AP12]

ID	G16 Repeated search
Title	Provide a history of queried searches for quick reuse.
Description	In principle, users appreciate the feature of saved search queries because they save time if they can reuse done work (selection of keywords).
Trade-off (optional)	The design of a search history is challenging because users can easily be overwhelmed by the amount of the accumulated information. The search for a saved query could result in extra work for the users and may take longer than to start a new search. Another problem is that this could inhibit the users' motivation to make new searches.
Context	1.5 Saving Search Results 25. User Interface / System Functionalities
Theory	
Example	Structured search results may be a solution approach to overcome the problems of a lot of accumulated results.

ID	G17 Multiple views
Title	Use multiple views to get different views on the data.
Description	This helps users to assess the value of different solutions and might even help to overcome some forms of biases. It encourages the analyst to address the problem holistically and examine more than one type of the data presentation method.
Trade-off (optional)	Restrictions due to screen size.
Context	5. Bias Mitigation 25. User Interface / System Functionalities Fluid transition between various data views Ability to see data in multiple linked views that support cross filtering
Theory	Robertson et al. 2008 [RFF+08], North & Shneiderman 2000 [NS03], Kang et al. 2011 [KGS11]
Example	In the car theft case there were multiple related events in different locations at different times. Multiple linked views would help to get a better picture
Evaluation Criteria	Does the system support multiple views and linking and brushing?

List of Figures

List of Tables

Index

Acronyms

2D	Two-dimensional
3D	Three-dimensional
APA	American Psychological Association
CCA	Comparative Case Analysis
CCT	Crime Classification Table
CSR	Complete Spatial Randomness
EC	European Commission
fMRI	functional Magnetic Resonance Imaging
GPS	Global Positioning System
HCI	Human Computer Interaction
JND	Just Noticeable Difference
MO	Modus Operandi
MRI	Magnetic resonance imaging
POI	Point of Interest
R&D	Research & Development
S3	Space Similarity Selector
SPC	Statistical Process Chart
VA	Visual Analytics
VALCRI	Visual Analytics for Sense-making in Criminal Intelligence Analysis
VAST	Visual Analytics Science and Technology

Bibliography

[AA08] Gennady Andrienko and Natalia Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium On*, pages 51–58. IEEE, 2008.

[AA11] Natalia Adrienko and Gennady Adrienko. Spatial Generalization and Aggregation of Massive Movement Data. *IEEE Transactions on Visualization and Computer Graphics*, 17(2):205–219, 2011.

[AAD+10] Gennady Andrienko, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. Space, Time and Visual Analytics. *International Journal of Geographical Information Science*, 24(10):1577–1600, October 2010.

[AAK+14] Daniel Archambault, James Abello, Jessie Kennedy, Stephen Kobourov, Kwan-Liu Ma, Silvia Miksch, Chris Muelder, and Alexandru C. Telea. Temporal Multivariate Networks. In Andreas Kerren, Helen C. Purchase, and Matthew O. Ward, editors, *Multivariate Network Visualization: Dagstuhl Seminar #13201, Dagstuhl Castle, Germany, May 12-17, 2013, Revised Discussions*, pages 151–174. Springer International Publishing, Cham, 2014.

[AB10] Simon Attfield and Ann Blandford. Discovery-led refinement in e-discovery investigations: Sensemaking, cognitive ergonomics and system design. *Artificial Intelligence and Law*, pages 1–23, 2010.

[ABHR+13] Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 483–492, New York, 2013. ACM.

[ABW08] Richard Adderley, Atta Badii, and Chaoxin Wu. The automatic identification and prioritisation of criminal networks from police crime data. In *Intelligence and Security Informatics*, pages 5–14. Springer, Berlin, Heidelberg, 2008.

[AF11] Martin A. Andresen and Marcus Felson. Co-Offending and the Diversification of Crime Types. *International Journal of Offender Therapy and Comparative Criminology*, 56(5):811–29, 2011.

[AHW10] Simon Attfield, Sukhvinder Hara, and B. L. William Wong. Sense-making in visual analytics: Processes and challenges. In *EuroVAST 2010: The 1st European Symposium on Visual Analytics Science and Technology.*, Bordeaux, France, 2010.

[ALS97] J Allan, AV Leouski, and RC Swan. Interactive cluster visualization for information retrieval. *In Proceedings of ECDL'98*, 1997.

[AMST11] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. *Visualization of Time-Oriented Data*. Human-Computer Interaction Series. Springer London, London, 2011.

[AP12] Daniel Archambault and Helen C Purchase. The mental map and memorability in dynamic graphs. In *Visualization Symposium (PacificVis), 2012 IEEE Pacific*, pages 89–96. IEEE, 2012.

[AP16] Daniel Archambault and Helen C. Purchase. Can animation support the visualisation of dynamic graphs? *Information Sciences*, 330:495–509, February 2016.

[Ass06] Association of Chief Police Officers in Scotland. *Major Incident Manual*. Association of Chief Police Officers in Scotland (ACPOS), Centrex, (Revised Edition, 2006).

[Bar97] Lyn Bartram. Perceptual and interpretative properties of motion for information visualization. In *Proceedings of the 1997 Workshop on New Paradigms in Information Visualization and Manipulation*, NPIV '97, pages 3–7, Las Vegas, Nevada, United States, 1997. ACM.

[BBDW14] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. The State of the Art in Visualizing Dynamic Graphs. In *EuroVis - STARs*, EuroVis, pages 83–103, Swansea, 2014. Eurographics Association.

[BDM+17] Roger Beecham, Jason Dykes, Wouter Meulemans, Aidan Slingsby, Cagatay Turkay, and Jo Wood. Map LineUps: Effects of spatial structure on graphical inference. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):391–400, January 2017.

[BFA14] Ann Blandford, Sarah Faisal, and Simon Attfield. Conceptual design for sensemaking. In *Handbook of Human Centric Visualization*, pages 253–283. Springer, 2014.

[BG10] R Rensink Baldridge and G. The perception of correlation in scatterplots. *Computer Graphics Forum*, 29(3):1203–1210, 2010.

[BHD+15] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J. D. Fekete, and T. Grabowski. Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks. In *Computer Graphics Forum*, volume 34, pages 31–40, 2015.

[BM15] Michael Burch and Tanja Munz. Edge-stacked timelines for visualizing dynamic weighted digraphs. *Proc. of the 6th International Conference on Information Visualization Theory and Applications; VISIGRAPP*, pages 93–100, 2015.

[Bog05] Robert L. Bogue. Use S.M.A.R.T. goals to launch management by objectives plan. *TechRepublic*, April 2005. [Online; Last accessed: 2016-03-23].

[BPF14a] Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. GraphDiaries: Animated Transitions andTemporal Navigation for Dynamic Networks. *IEEE Transactions on Visualization and Computer Graphics*, 20(5):740–754, May 2014.

[BPF14b] Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. Visualizing Dynamic Networks with Matrix Cubes. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 877–886, New York, NY, USA, 2014. ACM.

[BPWv17] Kathrin Ballweg, Margit Pohl, Günter Wallner, and Tatiana von Landesberger. Visual Similarity Perception of Directed Acyclic Graphs: A Study on Influencing Factors. *arXiv preprint arXiv:1709.01007*, 2017.

[BR00] T. Boren and J. Ramey. Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3):261–278, September 2000.

[BS13] Rachel Boba Santos. *Crime Analysis with Crime Mapping.* Sage, Los Angeles, London, New Delhi, 2013.

[Bux10] Bill Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design.* Morgan Kaufmann, 2010.

[BVB+11] Michael Burch, Corinna Vehlow, Fabian Beck, Stephan Diehl, and Daniel Weiskopf. Parallel edge splatting for scalable dynamic graph visualization. *IEEE transactions on visualization and computer graphics*, 17(12):2344–2353, December 2011.

[Cai07] Paul Cairns. HCI... Not As It Should Be: Inferential Statistics in HCI Research. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It - Volume 1*, BCS-HCI '07, pages 195–201, University of Lancaster, United Kingdom, 2007. British Computer Society.

[Car08] Sheelagh Carpendale. Evaluating Information Visualizations. In *Information Visualization: Human-Centered Issues and Perspectives*, volume 4950/2008 of *Lecture Notes in Computer Science*, pages 19–45. Springer Berlin / Heidelberg, 2008.

[CFL10] Arzu Coltekin, Sara Irina Fabrikant, and Martin Lacayo. Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *Int. J. Geogr. Inf. Sci.*, 24(10):1559–1575, October 2010.

[CGW14] Kris Cook, Georges Grinstein, and Mark Whiting. VAST Challenge 2014: The Kronos Incident. http://vacommunity.org/VAST+Challenge+2014, 2014. (Last accessed: 2017-01-07).

[Cla98] Andy Clark. *Being There: Putting Brain, Body, and World Together Again.* MIT press, 1998.

[Cla13] Robert M Clark. *Intelligence Analysis: A Target-Centric Approach.* SAGE Publications, Thousand Oaks, CA: CQ Press, 2013.

[CN06] Dempsey Chang and Keith V. Nesbitt. Developing Gestalt-based design guidelines for multi-sensory displays. In *Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop - Volume 57*, MMUI '05, pages 9–16, Sydney, Australia, 2006. Australian Computer Society, Inc.

[COD07] Code of Conduct - Rules to Ensure Good Scientific Practice. https://www.tuwien.ac.at/dle/datenschutz_und_dokumentenmanagement/richtlinien_und 2007. (Last accessed: 2017-01-08).

[CS73] William G Chase and Herbert A Simon. Perception in chess. *Cognitive psychology*, 4(1):55–81, 1973.

[CS98] Patricia A Carpenter and Priti Shah. A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2):75, 1998.

[Cum14] Geoff Cumming. The new statistics: Why and how. *Psychological science*, 25(1):7–29, 2014.

[CYFR14] Lane Harrison Chang, Fumeng Yang, Steven Franconeri, and Remco. Ranking Visualizations of correlation using Weber's Law. *IEEE Conference on Information Visualization (InfoVis)*, 20(12):1943–1952, 2014.

[CZGR09] Remco Chang, Caroline Ziemkiewicz, Tera Marie Green, and William Ribarsky. Defining Insight for Visual Analytics. *IEEE Comput. Graph. Appl.*, 29(2):14–17, March 2009.

[Dag] O. Dag. George Orwell: In Front of Your Nose. http://orwell.ru/library/articles/nose/english/e_nose. (Last accessed:10-01-2019).

[DBS08] Oscar De Bruijn and Robert Spence. A new framework for theory-based interaction design applied to serendipitous information retrieval. *ACM transactions on computer-human interaction (TOCHI)*, 15(1):5, 2008.

[DD95] Roger L. Dominowski and Pamela Dallob. Insight and problem solving. In *The Nature of Insight*, pages 33–62. The MIT Press, Cambridge, MA, US, 1995.

[DGK01] Stephan Diehl, Carsten Görg, and Andreas Kerren. Preserving the Mental Map using Foresighted Layout. In David S. Ebert, Jean M. Favre, and Ronald Peikert, editors, *Data Visualization 2001*, pages 175–184, Vienna, 2001. Springer Vienna.

[DHBDP19] Johanna Doppler Haider, Roger Beecham, Jason Dykes, and Margit Pohl. Detecting Difference in map line-up tasks. submitted 2019.

[DHDP+19] Johanna Doppler Haider, Isabel Dallinger, Margit Pohl, Patrick Seidler, Neesha Kodagoda, and B.L. William Wong. Sensemaking Strategies vs. Quality of Insights: Investigating Analysis Processes Using Multiple Visualizations. submitted 2019.

[DHGP+18] Johanna Doppler Haider, Bettina Gastecker, Margit Pohl, Patrick Seidler, Neesha Kodagoda, and B. L. William Wong. Sense-making Strategies in Explorative Intelligence Analysis of Network Evolutions. *Behaviour & Information Technology*, 2018.

[DHSP+17] Johanna Doppler Haider, Patrick Seidler, Margit Pohl, Neesha Kodagoda, Rick Adderley, and B. L. William Wong. How Analysts Think: Sense-making Strategies in the Analysis of Temporal Evolution and Criminal Network Structures and Activities. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 61, pages 193–197, September 2017.

[Dir95] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, November 1995.

[Dir16] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, May 2016.

[DL94] Norman K. Denzin and Yvonna S. Lincoln, editors. *Handbook of Qualitative Research.* Handbook of qualitative research. Sage Publications, Inc, Thousand Oaks, CA, US, 1994.

[Dra16] Pierre Dragicevic. Fair Statistical Communication in HCI. In Judy Robertson and Maurits Kaptein, editors, *Modern Statistical Methods for HCI*, pages 291–330. Springer International Publishing, Cham, 2016.

[DS03] JE Davidson and RJ Sternberg, editors. *The Psychology of Problem Solving.* Cambridge University Press, Cambridge, 2003.

[DS09] Cody Dunne and Ben Shneiderman. Improving graph drawing readability by incorporating readability metrics: A software tool for network analysts. *University of Maryland, HCIL Tech Report HCIL-2009-13*, 2009.

[EBB05] Dominik Engel, Sven Bertel, and Thomas Barkowsky. Spatial principles in control of focus in reasoning with mental representations, images, and diagrams. In *In*, pages 181–203. Springer, 2005.

[Ein55] A Einstein. Old Man's Advice to Youth:'Never Lose a Holy Curiosity. *Life Magazine*, page 64, 1955.

[ES93] Karl Anders Ericsson and Herbert Alexander Simon. *Protocol Analysis.* MIT press Cambridge, MA, 1993.

[FCB01] Susan N Friel, Frances R Curcio, and George W Bright. Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in mathematics Education*, pages 124–158, 2001.

[FHQ11] Michael Farrugia, Neil Hurley, and Aaron Quigley. Exploring temporal ego networks using small multiples and tree-ring layouts. *Proc. ACHI*, 2011:23–28, 2011.

[FL09] Fiona Fidler and Geoffrey R Loftus. Why figures with error bars should replace p values: Some conceptual arguments and empirical demonstrations. *Zeitschrift für Psychologie/Journal of Psychology*, 217(1):27–37, 2009.

[FNPS99a] Anna Fredrikson, Chris North, Catherine Plaisant, and Ben Shneiderman. Temporal, geographical and categorical aggregations viewed through coordinated displays: A case study with highway incident data. In *Proceedings of the 1999 Workshop on New Paradigms in Information Visualization and Manipulation in Conjunction with the Eighth ACM Internation Conference on Information and Knowledge Management*, NPIVM '99, pages 26–34, Kansas City, Missouri, United States, 1999. ACM.

[FNPS99b] Anna Fredrikson, Chris North, Catherine Plaisant, and Ben Shneiderman. Temporal, Geographical and Categorical Aggregations Viewed through Coordinated Displays: A Case Study with Highway Incident Data. *Workshop on New Paradigms in information Visualization and Manipulation*, pages 26–34, 1999.

[FQ11] Michael Farrugia and Aaron Quigley. Effective Temporal Graph Layout: A Comparative Study of Animation versus Static Display Methods. *Information Visualization*, 10(1):47–64, January 2011.

[FW04] Jessica I. Fleck and Robert W. Weisberg. The use of verbal protocols as data: An analysis of insight in the candle problem. *Memory & Cognition*, 32(6):990–1006, September 2004.

[GB98] Gene Golovchinsky and Nicholas J. Belkin. Innovation and Evaluation in Information Exploration Interfaces. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, CHI '98, pages 196–197, Los Angeles, California, USA, 1998. ACM.

[GF09] Tera Marie Green and Brian Fisher. The personal equation of complex individual cognition during visual interface interaction. In *Workshop on Human-Computer Interaction and Visualization*, pages 38–57. Springer, 2009.

[GFC04] Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. A Comparison of the Readability of Graphs using Node-Link and Matrix-Based Representations. In *IEEE Symposium on Information Visualization, 2004. INFOVIS 2004.*, pages 17–24. Ieee, 2004.

[GGZL16] Hua Guo, Steven R Gomez, Caroline Ziemkiewicz, and David H Laidlaw. A case study using visualization interaction logs and insight metrics to understand how analysts arrive at insights. *IEEE transactions on visualization and computer graphics*, 22(1):51–60, 2016.

[Gig07] Gerd Gigerenzer. *Gut Feelings: The Intelligence of the Unconscious*. Penguin, 2007.

[GM05] KJ Gilhooly and P Murphy. Differentiating insight from non-insight problems. *Thinking & Reasoning*, 11(3):279–302, 2005.

[GS13] Andreas Gegenfurtner and Marko Seppänen. Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, 63:393–403, 2013.

[GZ08] David Gotz and Michelle X Zhou. An empirical study of user interaction behavior during visual analysis. *Technical Report RC24525*, 2008.

[Har03] M Harrower. Tips for Designing Effective Animated Maps. *Cartographic Perspectives*, 44:63–65, 2003.

[HB03] Mark Harrower and Cynthia A. Brewer. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, 40(1):27–37, June 2003.

[HB11] Mark Harrower and Cynthia A. Brewer. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. In Martin Dodge, Rob Kitchin, and Chris Perkins, editors, *The Map Reader*, pages 261–268. John Wiley & Sons, Ltd, 2011.

[HCS98] Robert R Hoffman, Beth Crandall, and Nigel Shadbolt. Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40(2):254–276, 1998.

[HE12] Christopher G. Healey and James T. Enns. Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18:1170–1188, 2012.

[HF06] Nathalie Henry and Jean-Daniel Fekete. MatrixExplorer: A Dual-Representation System to Explore Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):677–684, September 2006.

[HF07] Nathalie Henry and Jean-Daniel Fekete. MatLink: Enhanced Matrix Visualization for Analyzing Social Networks. In Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa, editors, *Human-Computer Interaction – INTERACT 2007*, number 4663 in Lecture Notes in Computer Science, pages 288–302. Springer, Berlin Heidelberg, September 2007.

[HF08] Mark Harrower and Sara Fabrikant. The role of map animation for geographic visualization. *Geographic visualization: concepts, tools and applications*, pages 49–65, 2008.

[HFM07] N. Henry, J. Fekete, and M.J. McGuffin. NodeTrix: A Hybrid Visualization of Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1302–1309, November 2007.

[HFMC12] H. Hofmann, L. Follett, M. Majumder, and D. Cook. Graphical Tests for Power Comparison of Competing Designs. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2441–2448, December 2012.

[HHE06] Weidong Huang, Seok-Hee Hong, and Peter Eades. How people read sociograms: A questionnaire study. In *Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation-Volume 60*, pages 199–206. Australian Computer Society, Inc., 2006.

[HHK00] James Hollan, Edwin Hutchins, and David Kirsh. Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, 7(2):174–196, June 2000.

[HK18] Gillian Hill and Shelly M. Kemp. Uh-Oh! What Have We Missed? A Qualitative Investigation into Everyday Insight Experience. *The Journal of Creative Behavior*, 52(3):201–211, 2018.

[HMSA08] Jeffrey Heer, Jock D. Mackinlay, Chris Stolte, and Maneesh Agrawala. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1189–1196, 2008.

[HNA+11] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures.* OUP Oxford, 2011.

[Hog01] Robin M. Hogarth. *Educating Intuition*, volume xii. University of Chicago Press, Chicago, IL, US, 2001.

[HP14] R. J. Jr Heuer and R. H. Pherson. *Structured Analytic Techniques for Intelligence Analysis (2nd Ed.).* Los Angeles, CA: SAGE CQ Press, 2014.

[HPF13] Johanna Haider, Margit Pohl, and Peter Fröhlich. Defining Visual User Interface Design Recommendations for Highway Traffic Management Centres. In *17th International Conference*, pages 204–209, 2013.

[HPH+nt] Johanna Haider, Margit Pohl, Eva-Catherine Hillemann, Alexander Nussbaumer, Simon Attfield, Peter Passmore, and B. L. William Wong. Exploring the Challenges of Implementing Guidelines for the Design of Visual Analytics Systems. In *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting - 2015*, Los Angeles, 2015, in print.

[HPPW15] Johanna Haider, Margit Pohl, Chris Pallaris, and William B.L. Wong. Supporting Sense-Making and Insight Processes in Visual Analytics by Deriving Guidelines from Empirical Results. In *Proceedings of the Internation Summer School on Visual Computing 2015*, pages 59–68. Fraunhofer IGD, Rostock, Germany, 2015. Vortrag: International Summer School on Visual Computing 2015, Rostock, Germany; 2015-08-17 – 2015-08-21.

[HS74] J. R. Hayes and H. A. Simon. Understanding written problem instructions. In *Knowledge and Cognition*, pages ix, 321–ix, 321. Lawrence Erlbaum, Oxford, England, 1974.

[HS12] Jeffrey Heer and Ben Shneiderman. Interactive Dynamics for Visual Analysis. *Commun. ACM*, 55(4):45–54, April 2012.

[HSP12] Jeffrey Heer, Ben Shneiderman, and College Park. A Taxonomy of Tools That Support the Fluent and Flexible Use of Visualizations. *Interactive Dynamics for Visual Analysis*, 10:1–26, 2012.

[JH95] Brigitte Jordan and Austin Henderson. Interaction analysis: Foundations and practice. *The journal of the learning sciences*, 4(1):39–103, 1995.

[Joh80] Philip Nicholas Johnson-Laird. Mental models in cognitive science. *Cognitive science*, 4(1):71–115, 1980.

[Joh96] Philip N Johnson-Laird. Space to think. *Language and space*, pages 437–462, 1996.

[Kah11] Daniel Kahneman. *Thinking, Fast and Slow.* Macmillan, 2011.

[KAW+14] Sungahnn Ko, Shehzad Afzal, Simon Walton, Yang Yang, Junghoon Chae, Abish Malik, Yun Jang, Min Chen, and David Ebert. Analyzing high-dimensional multivariate network links with integrated anomaly detection, highlighting and exploration. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 83–92, 2014.

[KBB+10] Daniel A Keim, Peter Bak, Enrico Bertini, Daniela Oelke, David Spretke, and Hartmut Ziegler. Advanced visual analytics interfaces. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 3–10. ACM, 2010.

[KCM89] Gary A Klein, Roberta Calderwood, and Donald Macgregor. Critical decision method for eliciting knowledge. *IEEE Transactions on systems, man, and cybernetics*, 19(3):462–472, 1989.

[KDA+09] Per Ola Kristensson, Nils Dahlback, Daniel Anundi, Marius Bjornstad, Hanna Gillberg, Jonas Haraldsson, Ingrid Martensson, Mathias Nordvall, and Josefine Stahl. An Evaluation of Space Time Cube Representation of Spatiotemporal Patterns. *IEEE Transactions on Visualization and Computer Graphics*, 15(4):696–702, 2009.

[KEC06] René Keller, Claudia M. Eckert, and P. John Clarkson. Matrices or Node-link Diagrams: Which Visual Representation is Better for Visualising Connectivity Models? *Information Visualization*, 5(1):62–76, March 2006.

[KEM07] Andreas Kerren, Achim Ebert, and Jörg Meyer, editors. *Human-Centered Visualization Environments*, volume 4417 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, Heidelberg, 2007.

[KGS11] Youn-ah Kang, Carsten Gorg, and John Stasko. How Can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 17(5):570–583, 2011.

[KHWP16] Simone Kriglstein, Johanna Haider, Günter Wallner, and Margit Pohl. Who, Where, When and with Whom? Evaluation of Group Meeting Visualizations. In Mateja Jamnik, Yuri Uesaka, and Stephanie Elzer Schwartz, editors, *Diagrammatic Representation and Inference: 9th International Conference, Diagrams 2016, Philadelphia, PA, USA, August 7-10, 2016, Proceedings*, pages 235–249. Springer International Publishing, Cham, 2016.

[KJ11] Gary Klein and Andrea Jarosz. A Naturalistic Study of Insight. *Journal of Cognitive Engineering and Decision Making*, 5(4):335–351, December 2011.

[KKEM10] DA Keim, J Kohlhammer, G Ellis, and F Mansmann. *Mastering The Information Age-Solving Problems with Visual Analytics*. Goslar: Eurographics Association, 2010.

[Kle13] Gary Klein. *Seeing What Others Don't: The Remarkable Ways We Gain Insights*. PublicAffairs, a Member of the Perseus Book Group, New York, USA, 2013.

[KM13] Robert Kosara and Jock Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[KMH06a] Gary Klein, Brian Moon, and Robert R. Hoffman. Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intlelligent Systems*, 21(4):70–73, July 2006.

[KMH06b] Gary Klein, Brian Moon, and Robert R. Hoffman. Making Sense of Sensemaking 2: Alternative Perspectives. *IEEE Intlelligent Systems*, 21(5):88–92, September 2006.

[KNC+11] Udayan Khurana, Viet-An Nguyen, Hsueh-Chien Cheng, Jae-wook Ahn, Xi Chen, and Ben Shneiderman. Visual analysis of temporal trends in social networks using edge color coding and metric timelines. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference On*, pages 549–554. IEEE, 2011.

[Kof35] Kurt Koffka. *Principles of Gestalt Psychology*. New York: Harcourt, Brace, 1935.

[Kos89] Stephen M Kosslyn. Understanding charts and graphs. *Applied cognitive psychology*, 3(3):185–225, 1989.

[KP15] Simone Kriglstein and Margit Pohl. Choosing the Right Sample? Experiences of Selecting Participants for Visualization Evaluation. In W. Aigner, P. Rosenthal, and C. Scheidegger, editors, *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*. The Eurographics Association, 2015.

[KPDH18] Simone Kriglstein, Margit Pohl, and Johanna Doppler Haider. How Users Transform Node-Link Diagrams to Matrices and Vice Versa. In Peter Chapman, Gem Stapleton, Amirouche Moktefi, Sarah Perez-Kriz, and Francesco Bellucci, editors, *Diagrammatic Representation and Inference*, pages 526–534, Cham, 2018. Springer International Publishing.

[KPRP07] Gary Klein, Jennifer K Phillips, Erica L Rall, and Deborah A Peluso. A data-frame theory of sensemaking. In *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, pages 113–155. New York, NY, USA: Lawrence Erlbaum, 2007.

[KPS12] Simone Kriglstein, Margit Pohl, and Claus Stachl. Animation for Time-oriented Data: An Overview of Empirical Research. In Ebad Banissi, Stefan Bertschi, Remo Aslak Burkhard, Urska Cvek, Martin J. Eppler, Camilla Forsell, Georges G. Grinstein, Jimmy Johansson, Sarah Kenderdine, Francis T. Marchese, Carsten Maple, Marjan Trutschl, Muhammad Sarfraz, Liz J. Stuart, Anna Ursyn, and Theodor G. Wyeld, editors, *IV*, pages 30–35. IEEE Computer Society, 2012.

[KPS+17] Neesha Kodagoda, Sheila Pontis, Donal Simmie, Simon Attfield, BL William Wong, Ann Blandford, and Chris Hankin. Using Machine Learning to Infer Reasoning Provenance From User Interaction Log Data: Based on the Data/Frame Theory of Sensemaking. *Journal of Cognitive Engineering and Decision Making*, 11(1):23–41, 2017.

[KPSL08] Andreas Kjellin, Lars Winkler Pettersson, Stefan Seipel, and Mats Lind. Evaluating 2D and 3D visualizations of spatiotemporal information. *ACM Trans. Appl. Percept.*, 7(3):19:1–19:23, June 2008.

[Kri99] Lisa Krizan. *Intelligence Essentials for Everyone.* Joint Military Intelligence College / DTIC Document, Washington, DC, 1999.

[KYW+07] Sung-il Kim, M. Yoon, S.-M. Whang, Barbara Tversky, and Julie Bauer Morrison. The effect of animation on comprehension and interest: Animation and interest. *Journal of Computer Assisted Learning*, 23(3):260–270, January 2007.

[LF05] Patrick Lemaire and Ludovic Fabre. Strategic aspects of human cognition: Implications for understanding human reasoning. *Methods of though: Individual differences in reasoning strategies*, pages 11–56, 2005.

[LRIC15] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. More than telling a story: Transforming data into visually shared stories. *IEEE computer graphics and applications*, 35(5):84–90, 2015.

[LS15] Xiaotong Liu and Han-Wei Shen. The Effects of Representation and Juxtaposition on Graphical Perception of Matrix Visualization. In *Proceedings of* *the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 269–278, New York, NY, USA, 2015. ACM.

[May95] Richard E. Mayer. The search for insight: Grappling with Gestalt psychology's unanswered questions. In *The Nature of Insight.*, pages 3–32. The MIT Press, Cambridge, MA, US, 1995.

[May03] Philipp Mayring. *Qualitative Inhaltsanalyse: Grundlagen Und Techniken.* Beltz Verlag Weinheim, 2003.

[Maz09] Riccardo Mazza. *Introduction to Information Visualization.* Springer Publishing Company, Incorporated, 1 edition, 2009.

[Mil56] George A Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[Moo85] James H. Moor. What is computer ethics? *Metaphilosophy*, 16(4):266–275, October 1985.

[Mor50] Patrick AP Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950.

[MSSW16] Eva Mayr, Günther Schreder, Michael Smuc, and Florian Windhager. Looking at the Representations in Our Mind: Measuring Mental Models of Information Visualizations. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, pages 96–103, New York, NY, USA, 2016. ACM.

[Nat08] National Policing Improvement Agency on behalf of the Association of Chief Police Officers. *Practice Advice on Analysis.* National Policing Improvement Agency on behalf of the Association of Chief Police Officers, UK, 2008.

[Nat11] United Nations. *United Nations Office on Drugs and Crime. Criminal Intelligence.* United Nations, New York, 2011.

[NB05] Laura R Novick and Miriam Bassok. *Problem Solving.* Cambridge University Press, 2005.

[Nes05] Keith V. Nesbitt. Using guidelines to assist in the visualisation design process. In *Proceedings of the 2005 Asia-Pacific Symposium on Information Visualisation - Volume 45*, APVis '05, pages 115–123, Sydney, Australia, 2005. Australian Computer Society, Inc.

[NHT01] L. Nowell, E. Hetzler, and T. Tanasse. Change blindness in information visualization: A case study. *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.*, 2001:15–22, 2001.

[Nie92] Jakob Nielsen. Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 373–380. ACM, 1992.

[Nor06] Chris North. Toward measuring visualization insight. *IEEE computer graphics and applications*, 26(3):6–9, 2006.

[NS72] Allen Newell and Herbert Alexander Simon. *Human Problem Solving*, volume 104. Prentice-Hall Englewood Cliffs, NJ, 1972.

[NS03] Chris North and Ben Shneiderman. Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata. In BENJAMIN B. BEDERSON and BEN SHNEIDERMAN, editors, *The Craft of Information Visualization*, Interactive Technologies, pages 341 – 348. Morgan Kaufmann, San Francisco, 2003.

[NSH02] Lucy Nowell, Robert Schulman, and Deborah Hix. Graphical encoding for information visualization: An empirical study. In *IEEE Symposium on Information Visualization*, pages 43–50, 2002.

[NVH+16] Alexander Nussbaumer, Katrien Verbert, Eva-Catherine Hillemann, Michael A Bedek, and Dietrich Albert. A Framework for Cognitive Bias Detection and Feedback in a Visual Analytics Environment. In *Intelligence and Security Informatics Conference (EISIC), 2016 European*, pages 148–151. IEEE, 2016.

[OD92] Claire O'Malley and Steve Draper. Representation and interaction: Are mental models all in the mind. *Models In the Mind. San Diego, CA: Academic Press Ltd*, 1992.

[ON11] T. Opach and A. Nossum. Evaluating the Usability of Cartographic Animations with Eye-Movement Analysis. In *Proceedings of the 25th International Cartographic Conference*, Paris, 2011.

[OU10] D O'Sullivan and D Unwin. *Geographic Information Analysis*. John Wiley & Sons, New Jersey, USA, 2 edition, 2010.

[PAKC15] Robert Pienta, James Abello, Minsuk Kahng, and Duen Horng Chau. Scalable graph exploration and visualization: Sensemaking challenges and opportunities. In *Big Data and Smart Computing (BigComp), 2015 International Conference On*, pages 271–278. IEEE, 2015.

[PC95] Peter Pirolli and Stuart Card. Information Foraging in Information Access Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 51–58. ACM Press/Addison-Wesley Publishing Co., 1995.

[PC05] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, volume 5, pages 2–4, 2005.

[PDE10] Margit Pohl, Alan Dix, and Geoffrey Ellis. Perception and Cognitive Aspects. In *Mastering The Information Age - Solving Problems with Visual Analytics*, chapter 7. Eurographics, November 2010.

[PDH17] Margit Pohl and Johanna Doppler Haider. Sense-making Strategies for the Interpretation of Visualizations–Bridging the Gap between Theory and Empirical Research. *Multimodal Technologies and Interaction*, 1(3):16, 2017.

[PHPW15] Margit Pohl, Johanna Haider, Chris Pallaris, and William B.L. Wong. Guidelines for Sense-making in Intelligence Analysis. In *European Intelligence and Security Informatics Conference*, page 177, 2015.

[Pin90] Steven Pinker. A theory of graph comprehension. *Artificial intelligence and the future of testing*, pages 73–126, 1990.

[PNS03] Jean E Pretz, Adam J Naples, and Robert J Sternberg. Recognizing, defining, and representing problems. *The psychology of problem solving*, 30(3), 2003.

[PRS+94] Jenny Preece, Yvonne Rogers, Helen Sharp, David Benyon, Simon Holland, and Tom Carey. *Human-Computer Interaction.* Addison-Wesley Longman Ltd., 1994.

[PSCO09] William A. Pike, John Stasko, Remco Chang, and Theresa A. O'Connell. The Science of Interaction. *Information Visualization*, 8(4):263–274, 2009.

[PSM12] Margit Pohl, Michael Smuc, and Eva Mayr. The User Puzzle - Explaining the Interaction with Visual Analytics Systems. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2908–2916, 2012.

[Pur97] Helen Purchase. Which aesthetic has the greatest effect on human understanding? In *International Symposium on Graph Drawing*, pages 248–261. Springer, 1997.

[Pur12] Helen C. Purchase. *Experimental Human-Computer Interaction: A Practical Guide with Visual Examples.* Cambridge University Press, New York, NY, USA, 1st edition, 2012.

[PWaR04] Wei Peng, M O Ward, and E a Rundensteiner. Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering. *IEEE Symposium on Information Visualization*, pages 89 – 96, 2004.

[PWK16] Margit Pohl, Günter Wallner, and Simone Kriglstein. Using lag-sequential analysis for understanding interaction sequences in visualizations. *International Journal of Human-Computer Studies*, 96:54–66, 2016.

[PWM+12] Margit Pohl, Sylvia Wiltner, Silvia Miksch, Wolfgang Aigner, and Alexander Rind. Analysing interactivity in information visualisation. *KI-Künstliche Intelligenz*, 26(2):151–159, 2012.

[PWR04] Wei Peng, Matthew O Ward, and Elke A Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On*, pages 89–96. IEEE, 2004.

[PWR+11] Margit Pohl, Sylvia Wiltner, Alexander Rind, Wolfgang Aigner, Silvia Miksch, Thomas Turic, and Felix Drexler. Patient development at a glance: An evaluation of a medical data visualization. In *IFIP Conference on Human-Computer Interaction*, pages 292–299. Springer, 2011.

[Pyl03] Zenon Pylyshyn. Return of the mental image: Are there really pictures in the brain? *Trends in cognitive sciences*, 7(3):113–118, 2003.

[R C13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[Rap05] David N Rapp. Mental models: Theoretical issues for visualizations in science education. In *Visualization in Science Education*, pages 43–60. Springer, 2005.

[Rei10] Florian Reitz. A Framework for an Ego-centered and Time-aware Visualization of Relations in Arbitrary Data Repositories. *arXiv:1009.5183 [cs]*, September 2010.

[RFF+08] George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko. Effectiveness of animation in trend visualization. *IEEE transactions on visualization and computer graphics*, 14(6):1325–32, 2008.

[RJLP14] Khairi Reda, Andrew E. Johnson, Jason Leigh, and Michael E. Papka. Evaluating User Behavior and Strategy During Visual Exploration. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pages 41–45, Paris, France, 2014. ACM.

[RK16] Judy Robertson and Maurits Kaptein, editors. *Modern Statistical Methods for HCI*. Human–Computer Interaction Series. Springer International Publishing, 2016.

[RM14] S. Rufiange and G. Melancon. AniMatrix: A Matrix-Based Visualization of Software Evolution. In *2014 Second IEEE Working Conference on Software Visualization (VISSOFT)*, pages 137–146, September 2014.

[Rot13] Robert E. Roth. Interactive maps: What we know and what we need to know. *Journal of Spatial Information Science*, 6, June 2013.

[RSC15] Peter Rodgers, Gem Stapleton, and Peter Chapman. Visualizing sets with linear diagrams. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(6):27, 2015.

[RSPC93] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, pages 269–276. ACM, 1993.

[RTJ+11] Khairi Reda, Chayant Tantipathananandh, Andrew Johnson, Jason Leigh, and Tanya Berger-Wolf. Visualizing the Evolution of Community Structures in Dynamic Social Networks. *Computer Graphics Forum*, 30(3):1061–1070, June 2011.

[RW73] Horst WJ Rittel and Melvin M Webber. Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169, 1973.

[SC99] Daniel J Simons and Christopher F Chabris. Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28:1059–1074, 1999.

[Sch09] Steven H Schwartz. *Visual Perception: A Clinical Orientation*. McGraw Hill Professional, 2009.

[Sch12] Margrit Schreier. *Qualitative Content Analysis in Practice*. Sage Publications, London, 2012.

[SHK+16] P. Seidler, J. Haider, N. Kodagoda, B. L. W. Wong, M. Pohl, and R. Adderley. Design for Intelligence Analysis of Complex Systems: Evolution of Criminal Networks. In *2016 European Intelligence and Security Informatics Conference (EISIC)*, pages 140–143, August 2016.

[Shn96] Ben Shneiderman. The Eyes Have It : A Task by Data Type Taxonomy for Information Visualizations. In *Proc. VL'96*, pages 336–343. IEEE Computer Society, 1996.

[Shn10] Ben Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Education India, 2010.

[Sn15] Johnny Saldaña. *The Coding Manual for Qualitative Researchers*. Sage, 2015.

[SND05] Purvi Saraiya, Chris North, and Karen Duca. An Insight-Based Methodology for Evaluating Bioinformatics Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):443–456, July 2005.

[Spe11] Bob Spence. The Broker. In A Ebert, A Dix, N Gershon, and Margit Pohl, editors, *Human Aspects of Visualization*, pages 10–22. Springer, 2011.

[Spe14] Robert Spence. *Information Visualization*. Springer International Publishing, Cham, 2014.

[SPLM16] Kamran Sedig, Paul Parsons, Hai-Ning Liang, and Jim Morey. Supporting Sensemaking of Complex Objects with Visualizations: Visibility and Complementarity of Interactions. In *Informatics*, volume 3, page 20. Multidisciplinary Digital Publishing Institute, 2016.

[SRL05] Kamran Sedig, Sonja Rowhani, and Hai-Ning Liang. Designing interfaces that support formation of cognitive maps of transitional processes: An empirical study. *Interacting with computers*, 17(4):419–452, 2005.

[SSS10] Daniel Schmeiß, Ansgar Scherp, and Steffen Staab. Integrated mobile visualization and interaction of events and POIs. In *Proceedings of the International Conference on Multimedia*, MM '10, pages 1567–1570, Firenze, Italy, 2010. ACM.

[Sut00] Alistair Sutcliffe. On the effective use and reuse of HCI knowledge. *ACM Transactions on Computer-Human Interaction*, 7:197–221, 2000.

[Swe94] John Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312, 1994.

[TAB14] Tableau Desktop. https://www.tableau.com/products/desktop, 2014. (Last accessed: 2017-01-08).

[TC05] J J Thomas and K A Cook, editors. *Iluminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE CS Press, 2005.

[TGC+16] Barbara Tversky, Jie Gao, James E. Corter, Yuko Tanaka, and Jeffrey V. Nickerson. People, Place, and Time: Inferences from Diagrams. In Mateja Jamnik, Yuri Uesaka, and Stephanie Elzer Schwartz, editors, *Diagrammatic Representation and Inference: 9th International Conference, Diagrams 2016, Philadelphia, PA, USA, August 7-10, 2016, Proceedings*, pages 258–264. Springer International Publishing, Cham, 2016.

[TMB02] Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. Animation: Can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.

[Tob70] Waldo R Tobler. A computer movie simulating urban growth in the Detroit region. *Economic geography*, 46(sup1):234–240, 1970.

[Tom15] Christian Tominski. Interaction for visualization. *Synthesis Lectures on Visualization*, 3(1):1–107, 2015.

[TS09] Barbara Tversky and Masaki Suwa. Thinking with sketches. In *Tools for Innovation: The Science behind the Practical Methods That Drive New Ideas*, pages 75–84. Oxford University Press, New York, NY, US, 2009.

[TSL$^+$18] Martin Tik, Ronald Sladky, Caroline Di Bernardi Luft, David Willinger, André Hoffmann, Michael J Banissy, Joydeep Bhattacharya, and Christian Windischberger. Ultra-high-field fMRI insights on insight: Neural correlates of the Aha!-moment. *Human brain mapping*, 2018.

[Tve05] Barbara Tversky. Visuospatial reasoning. *The Cambridge handbook of thinking and reasoning*, pages 209–240, 2005.

[Tve11] Barbara Tversky. Visualizing Thought. *Topics in Cognitive Science*, 3(3):499–535, July 2011.

[Urd11] Timothy C Urdan. *Statistics in Plain English.* Routledge, 2011.

[VAL18] Visual Analytics for Sense-making in Criminal Intelligence Analysis. http://www.valcri.org/, 2018. (Last accessed: 2019-01-08).

[VH16] S. VanderPlas and H. Hofmann. Spatial Reasoning and Data Displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):459–468, January 2016.

[War04] Colin Ware. *Information Visualization: Perception for Design.* Morgan Kaufmann Publishers Inc., 2004.

[WB02] B. L. William Wong and Ann Blandford. Analysing Ambulance Dispatcher Decision Making: Trialing Emergent Themes Analysis. In Frank Vatere, Lorraine Johnston, and Ros Kushinsky, editors, *Proc. of the HF2002 Human Factors Conference Design for the Whole Person - Integrating Physical, Cognitive and Social Aspects*, Canberra, 2002. Ergonomics Society of Australia.

[WBB03] David Weisburd, Wim Bernasco, and Gerben Bruinsma, editors. *Putting Crime in Its Place.* Springer New York, 2003.

[WBWK00] Michelle Q Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 110–119. ACM, 2000.

[WCHB10] Hadley Wickham, Dianne Cook, Heike Hofmann, and Andreas Buja. Graphical Inference for Infovis. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):973–979, 2010.

[Wei95] Karl E Weick. *Sensemaking in Organizations*, volume 3. Sage, 1995.

[WGK10] Matthew O Ward, Georges Grinstein, and Daniel Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications.* CRC Press, 2010.

[WGLL98] Christopher D Wickens, Sallie E Gordon, Yili Liu, and John Lee. *An Introduction to Human Factors Engineering.* 1998.

[WHC15] Jagoda Walny, Samuel Huron, and Sheelagh Carpendale. An exploratory study of data sketching for visual representation. *Computer Graphics Forum*, 34(3):231–240, 2015.

[Wil99] Leland Wilkinson. Statistical methods in psychology journals: Guidelines and explanations. *American psychologist*, 54(8):594, 1999.

[Wol01] Stanley Woll. *Everyday Thinking: Memory, Reasoning, and Judgment in the Real World.* L. Erlbaum Associates, Mahwah, N.J, 2001.

[Won14] B L Wong. How Analysts Think (?): Early Observations. In *IEEE Joint Intelligence and Security Informatics Conf. (JISIC)*, pages 296–299, 2014.

[WPCM02] Colin Ware, Helen Purchase, Linda Colpoys, and Matthew McGill. Cognitive Measurements of Graph Aesthetics. *Information Visualization*, 1(2):103–110, June 2002.

[WSO05] Karl E Weick, Kathleen M Sutcliffe, and David Obstfeld. Organizing and the process of sensemaking. *Organization science*, 16(4):409–421, 2005.

[WV12] B.L. William Wong and Margaret Varga. Black Holes, Keyholes And Brown Worms: Challenges In Sense Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1):287–291, September 2012.

[YaS07] Ji Soo Yi, Youn ah Kang, and John Stasko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics*, 13(6):1224–1231, 2007.

[YS10] Paul Yoder and Frank Symons. *Observational Measurement of Behavior.* Springer Publishing Company, 2010.

[ZK10] Caroline Ziemkiewicz and Robert Kosara. Laws of attraction: From perceptual forces to conceptual similarity. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1009–1016, 2010.

[ZMG95] Rolf A Zwaan, Joseph P Magliano, and Arthur C Graesser. Dimensions of situation model construction in narrative comprehension. *Journal of experimental psychology: Learning, memory, and cognition*, 21(2):386, 1995.

Curriculum Vitae

Johanna Viktoria Doppler (Haider)
Vally-Weigl-Gasse 2/3/19
1100 Vienna, Austria
doppler.johanna@gmail.com

Personal data

Nationality:	Austria
Date of birth:	January 22nd, 1987
Place of birth:	Eisenstadt, Austria
Marital status:	Married

Education

2014/06 - 2019/01:	Doctoral programme in Engineering Sciences Institute of Visual Computing and Human-Centered Technology, TU Wien, Vienna
2010/11 - 2013/01:	Master degree in Computer Science (Dipl.-Ing.), TU Wien Thesis: "Design Guidelines for Effective Visualisations on the Basis of Empirical Studies"
2006/10 - 2010/11:	Bachelor degree in Computer Science (BSc.), TU Wien Thesis: "Integrating a what you see is what you get editor"
2010/01 - 2010/06:	Erasmus Student Exchange at the Polytechnic University of Valencia, School of Computer Engineering, Spain
2001/09 - 2006/06:	Polytechnic School with a 5-years programme on Computing and Organisation, HTBLVA Spengergasse, Vienna

Scholarships and awards

2015/09:	*Best Poster Award* at the European Intelligence and Security Informatics Conference EISIC, Manchester, United Kingdom
2010/01:	Erasmus scholarship at the UPV Valencia, Spain

Work experience

2014 - 2019:	Project assistant at the Institute of Visual Computing and Human-Centered Technology in the Human Computer Interaction group, TU Wien. Field of research: Visualisation
2013 - 2014:	Backend developer at Wienfluss information.design.solutions, Vienna. Field of work: web applications (Roxen/Pike, Ajax, PHP) and web design (HTML, CSS, JS)
2007 - 2013:	Technical support during studies as part-time project assistant at Wienfluss information.design.solutions, Vienna