

Seeking patterns of visual pattern discovery for knowledge building

N. Andrienko^{1,2}, G. Andrienko^{1,2}, S. Chen³, B. Fisher⁴

¹Fraunhofer Institute IAIS, Germany

²City University London, UK

³School of Data Science, Fudan University, China

⁴Simon Fraser University, Vancouver, Canada

Abstract

Currently, the methodological and technical developments in visual analytics, as well as the existing theories, are not sufficiently grounded by empirical studies that can provide an understanding of the processes of visual data analysis, analytical reasoning, and derivation of new knowledge by humans. We conducted an exploratory empirical study in which participants analysed complex and data-rich visualisations by detecting salient visual patterns, translating them into conceptual information structures, and reasoning about those structures to construct an overall understanding of the analysis subject. Eye tracking and voice recording were used to capture this process. We analysed how the data we had collected match several existing theoretical models intended to describe visualisation-supported reasoning, knowledge building, decision making, or use and development of mental models. We found that none of these theoretical models alone is sufficient for describing the processes of visual analysis and knowledge generation that we observed in our experiments, whereas a combination of three particular models could be apposite. We also pondered whether empirical studies like ours can be used to derive implications and recommendations for possible ways to support users of visual analytics systems. Our approaches to designing and conducting the experiments and analysing the empirical data were appropriate to the goals of the study and can be recommended for use in other empirical studies in visual analytics.

Categories and Subject Descriptors (according to ACM CCS): [Human-centered computing → Visual analytics]: Visualization application domains—Visual analytics

1. Introduction

The primary task of information visualisation is to represent information in a visual form so that users can perceive it accurately and efficiently. Expressiveness and effectiveness of visual encoding of information are the main criteria of visualisation design [Mac86]. Accordingly, empirical studies in information visualisation test visual encoding techniques for expressiveness and effectiveness of representing different kinds of information. Results of these studies get incorporated in fundamental principles of visual representation of information, such as choosing appropriate visual variables [Ber83], as well as design guidelines concerning specific visualisation methods (VisGuides [DAREA*18]).

Visual analytics, having stemmed from information visualisation, developed into a research discipline of its own. Naturally, it embraces the criteria and principles of visuali-

sation design developed in information visualisation. Visual analytics, however, has a different goal, which is to facilitate analytical reasoning and generation of new knowledge using information extracted from data [TC05, KAF*08]. Accordingly, the design of visual analytics methods and software needs to be underpinned by empirical research of how people use visualisations for reasoning and knowledge building. This kind of research is also required to verify the existing theoretical models intended to describe different aspects of the process of visual analysis and knowledge building. However, such empirical research is currently lacking, whereas it should be an organic part of the visual analytics discipline.

It is, certainly, much more difficult to test how well visualisation supports knowledge building than how accurately people decode information encoded visually. Knowledge is not explicitly represented in a visual display, but it needs to

be *constructed* by a person viewing the display. The process of knowledge building essentially involves abstraction and generalisation from multiple pieces of information extracted from a display. Consequently, empirical studies must be designed so that the participants are supposed to perform generalisation, and their generalisation-directed activities can be observed. This means that the tasks given to the participants must be synoptic [AA06] and open-ended.

According to these considerations, we conducted an explorative experimental study the goals of which were, first, to observe and study how the participants use visualisations to generalise from specific data instances to overall understanding of the phenomena reflected in the data and, second, analyse how our observations and findings match some of the existing theoretical models intended to describe visualisation-supported reasoning [KWKH19], knowledge building [SSS*14, AAM*21], decision making [PBG*14, PCRHS18], or use and development of mental models [ALA*18, KHL21].

To achieve the goals that we had set for our experimental study, we did the following:

- We designed the tasks for the experiments so that the participants were oriented to derivation of general knowledge.
- We asked the participants to think aloud during their analysis process, and we applied eye-tracking technology to collect data about gaze movements and fixations.
- We studied the audio recordings of the participants' utterances in conjunction with replaying the eye movements of the participants.
- We analysed the eye tracks using a combination of methods that allowed us to consider the data from different perspectives and at different levels of abstraction.
- In analysing the materials collected, we assessed the potential of such studies for improving the general understanding of the process of visual analysis and deriving implications for visual analytics design.

In this paper, we begin with formulating our research questions in Section 2. After an overview of the related work (Section 3), we describe and explain the design of our experiments (Section 4), the methods that we used to analyse the empirical data collected (Section 5), and the results of the analysis (Section 6). We then discuss different aspects of our work (Section 7) linking it to the theoretical research and drawing implications for designing visual analytics systems.

2. Research questions

Throughout the paper we use the terms *abstraction* (to *abstract*) and *generalisation* (to *generalise*). The definitions and implied meanings of these terms vary among dictionaries and contexts of use. In our paper, the term 'abstraction' is used in the sense of considering multiple things together as a single thing and the term 'generalisation' in the sense

of making a statement or forming a concept that refers to multiple things taken together.

As mentioned in the introduction, our research goal was to observe and study how participants of our experiments use visualisations to generalise from specific data instances to overall understanding of the phenomena reflected in the data. The study we conducted was exploratory by nature; therefore, we did not formulate any prior hypotheses. We gave the participants open-ended synoptic tasks requiring them to generalise. There was no ground truth for assessing the participants' performance, and there was no point in measuring the response time. We assumed that each participant may have an individual way and pace of reasoning, derive an individual mental model of the task subject, and come to own conclusions. It was not planned to compare and assess the conclusions of different participants. What we wanted was to observe the analysis processes and try to find commonalities among them, without making prior assumptions about the possible kinds of commonalities.

In line with the overall goal of our study, we posed several more specific research questions. We formulated these questions in relation to particular theoretical models describing the process of visual data analysis or some of its aspects. Among the large number of existing theories and frameworks intended to describe human cognitive activities, we selected a subset of theories that (1) explicitly address the use of visual representations of information and (2) consider high-level activities beyond attention and perception, namely, reasoning, decision making, and generation of new knowledge. In the following, we briefly recount the selected theories and models and state the questions referring to each of them. Please note that the questions and the answers we wanted to obtain apply only to our experimental study. We did not aspire to come to conclusions that would be immediately valid for visual analytics in general.

Visualisation-based decision making. Pinker [PF90] proposed a theoretical model representing comprehension of graphical displays as a combination of bottom-up and top-down processes. Bottom-up processing means that viewers' attention is involuntarily attracted to salient features in an image. In top-down processing, viewers purposefully direct their attention to what is relevant to their analysis task. This theory provided the foundation for the frameworks describing visualisation-supported decision making [PBG*14, PCRHS18, Pad18]. According to these frameworks, bottom-up perception prompts fast, automatic decision making whereas top-down processes are involved in slow, more contemplative decisions. Patterson et al. [PBG*14] emphasise the key role of the top-down processing, which guides the way in which the bottom-up information is processed. We wanted to observe the roles of the two types of processing in knowledge construction, which requires generalisation of perceived information.

RQ1: *What type of decision making prevails in data analysis aiming at generalisation?*

Visual analysis as Bayesian inference. Kim et al. [KWKH19, KKGH21] adopt the cognitive science approach in which individual cognition is modelled as Bayesian inference. Individuals have certain beliefs, which may be updated after observing new data. The previous beliefs are considered as Bayesian “priors” and the new beliefs as “posteriors”. However, experiments [KWKH19] showed that individuals often deviate from a rational behaviour that can be expected according to the model.

RQ2: *To what extent is the Bayesian inference model applicable to the process of analysis involving abstraction and generalisation?*

Knowledge generation model. The key assertion of the model proposed by Sacha et al. [SSS*14] is that the visual analysis process is guided by hypotheses. The process is a combination of three loops: exploration, verification, and knowledge generation. In the exploration, analysts strive to obtain findings that verify or falsify current hypotheses. A finding may be a pattern in data but also a lack of a pattern. Some findings may give rise to new hypotheses. In the verification loop, analysts evaluate the hypotheses based on the findings. In the knowledge generation loop, verified hypotheses are adopted as new knowledge.

RQ3: *Are these three types of activities (exploration, verification, and knowledge generation) observable in the analysis sessions?*

Visual analysis as model building. The knowledge constructed by a data analyst is seen as a mental model of the analysis subject, which is the phenomenon reflected in the data [ALA*18]. A model is a generalised representation of relationships between components of the analysis subject. The authors emphasise that the model represents the subject, not the data as such. In the process of analysis, the analyst always has some version of a model, which is repeatedly evaluated and updated until the analyst finds it appropriate and thus finishes the process.

RQ4: *Are model evaluation and modification activities identifiable in the analysis?* **RQ5:** *What are the termination conditions for the analysis process?*

Pattern discovery. Collins et al. [CAS*18] argue that derivation of a mental model from data necessitates abstraction, in which analysts consider multiple data items together as a meaningful whole, which is called a *pattern*. According to Andrienko et al. [AAM*21], unification of multiple data items becomes possible due to elementary relationships between elements of data components, such as ordering, distances, directions, and equivalence. There are also relationships between patterns enabling integration of smaller patterns into larger ones. Appropriate visualisations enable interpretation of visual patterns as data patterns. Discovered

patterns are involved in various analytical operations, such as refining, comparing, grouping, and uniting.

RQ6: *What types of visual patterns are observed and interpreted as data patterns?* **RQ7:** *What analytical operations are applied to discovered patterns?*

Qualitative visual analysis. Karer et al. [KHL21] criticises contemporary visualisation research as being data-centric. Gaining insights into data is commonly considered as the main purpose of visualisation, whereas it should be provision of insights into the domain reflected in the data (in other words, into the analysis subject [ALA*18]). Obtaining domain insights requires mapping data insights to a conceptual model that a user applies for reasoning. It includes knowledge of the analysis context (i.e., analysis aims and tasks), the user context (user’s background knowledge and skills), and the domain context (structure, properties, relationships, and behaviours of things and phenomena in the domain). The term “qualitative visual analysis” refers to the process of reasoning that transforms data insights into domain insights using different types of context information.

RQ8: *Is the involvement of context information identifiable in the analysis sessions? If so, how is this information used?*

To summarise, our plan was to observe the process of analytical reasoning and knowledge building with the use of data visualisation and to relate our observations to the current theoretical models describing the visual analysis process. By no means we supposed that our observations might be immediately generalisable to a wide range of visual data analysis scenarios. Our experiments should rather be treated as a feasibility study testing the potential of our approach for probing the applicability of current theories. Obtaining general conclusions certainly requires many empirical studies, and we wanted to see whether our approach can be recommended for use in such researches.

3. Related work

3.1. Empirical studies of visual data analysis

Most of the empirical studies conducted so far have been focused on the processes of perceiving and interpreting information encoded visually. The main purpose of the studies have been to draw implications for visualisation design [Heg11]. Padilla et al. [PCRHS18] reviewed the literature describing empirical studies of decision making with visualisations. They extracted four key findings that were common for multiple studies. First, visualisations direct participants’ bottom-up attention to specific visual features. Therefore, visualisations need to be designed so as to attract the viewers’ attention to task-relevant information. Second, the visual encoding technique may provoke certain kinds of biases. Designers need to be aware of such effects and avoid them. Third, visualisations vary in cognitive fit between the

visual encoding, user's knowledge of this kind of graph, and the question to be answered. A poor fit requires the user to perform mental transformations of the image, which increases the cognitive load. Fourth, user's graph-reading skills may interact with the effects of the visual encoding technique; hence, such skills need to be learned and trained.

Abdul-Rahman et al. [ARBC*19] performed an analytical survey of 129 empirical studies reported in the visualisation literature. They discussed a large number of variables that can be used for characterising empirical studies and proposed a taxonomy based on a selected subset of variables, namely, the purpose of a study (fundamental understanding or technical evaluation), scopes of users' prior knowledge (context, pattern, or statistics), visualisation tasks (e.g., retrieve, identify, determine, correlate, cluster, etc.), and functional categories of behaviours (sensing, storing, learning, thinking, etc.). Next, they juxtaposed the empirical research in visualisation with the research topics in psychology. They noted that some topics are actively researched in both disciplines. Examples are visual estimation of correlation and perception of colours. The authors advocate close cooperation between visualisation researchers and psychologists in conducting empirical studies.

While the literature on studying the effects of various factors on perceiving and interpreting visualisations is abundant, there has been not much empirical research focusing on the process of visualisation-supported analytical reasoning. Some work was done on reconstructing cognitive actions of users of interactive data displays from user interaction logs analysed in combination with thinking aloud protocol [KPS*17, PWK16]. The main objective was to find approaches to translating sequences of elementary operations into higher-level activities defined in relevant theories.

There is also rich literature concerning evaluation of information visualisations. Interested readers can be referred to the surveys [LBI*12, IIC*13], where the evaluation works are categorised according to seven scenarios oriented to different research questions. However, all these research questions refer to the use of specific visualisation designs or systems rather than to the analysis processes.

The most related to our work is the exploratory study performed by Isenberg et al. [ITC08]. The participants of the study performed data analysis with the use of diverse visual representations printed on cards, which allowed natural ways of interaction: selecting, comparing, grouping, etc. In individual and collaborative analysis sessions, the participants were supposed to answer both focused and open-ended questions. In the group sessions, the participants naturally discussed their thoughts, actions, and findings. In individual sessions, the participants were asked to talk while working. By observing and analysing the participants' behaviours, the researchers revealed several kinds of analytical activities, such as scanning through data to get a feel for the available information, finding and picking information

relevant to a task, extracting information from a visual representation, and validating partial or complete solutions to the tasks. General findings were absence of common temporal patterns of analytical activities, diversity of the ways in which individuals approached the tasks, variation of the approaches to organising joint work in the groups, and absence of a particular temporal order of performing analytical activities, contrary to assertions of multiple theoretical models. Implications for the design of information visualisation systems were to support flexible temporal sequence of work processes, changing strategies in collaborative work, and flexible organisation of workspaces where analyst can put and arrange relevant information.

In our study, we had the same goal as Isenberg et al. in theirs [ITC08]: to observe and understand the analysis process; however, our experiments were designed differently, as will be described later on. Compared to the other empirical studies, our study differs in its focusing on analytical reasoning and knowledge building rather than graph perception and information extraction.

3.2. Eye tracking-based studies of cognitive activities

Eye tracking enables recording of the gaze movement (saccades) and stops (fixations). Eye tracking technologies are widely used in a variety of applications [Duc07] for understanding human's visual and attentional processes based on how they see things and how they interact with them. Researchers want to know *where* people look, *when* (at which stage of exploration), and for *how long* time. Researchers also analyse if a region of a scene (called *area of interest*, or *AOI*) is attended once or multiple times, how *frequently*, and in what *order of transitions*. Prevailing long saccades often indicate the process of global exploration, while varying length of saccades corresponds to local inspection [FLMB11]. Two major types of attention are top-down goal-driven intent and bottom-up stimulus-driven saliency-based attention [KGRD17], which have been reflected in the Pinker's model of graphics comprehension [PF90] mentioned in Section 2.

Among many applications [Duc07], eye tracking has been used for studying problem solving with maps [KGR13] and graphs [BAA*13]. Specifically for geo-visualisations with data-rich maps, eye tracking was used to investigate the efficiency of single and small multiple maps [FRHA*08], compare 2D and 3D representations [PB13], study the use of dynamic (changes over time) [OGF14] and interactive (changes caused by user) maps. Kiefer et al. [KGR13] identified a set of features of eye movement trajectories by which it was possible to distinguish different tasks performed with maps: free exploration, global search, route planning, focused search, line following, and polygon comparison. These features were used to train a machine learning model for recognising the tasks based on eye-tracking data. Silva et al. [SBJ*19] considered the possibility of using eye

tracking as a tool for interaction with a display. Based on tracking user's gaze, a system can, for example, provide details on demand, suggest related pieces of information, filter, modify symbolisation, etc.

Unlike in the bulk of the studies involving eye tracking, we did not strive to elicit differences in eye movement behaviours among individuals, types of tasks, types of visual displays, or user interfaces. Instead, our goal was to find common patterns that could give us insights into the process of visually supported analytical reasoning.

To fulfill our goal, we combined eye tracking with acquisition of verbal data by asking the participants to think aloud [ES93]. Acknowledging the high value of such data, Holmqvist et al. [HNA*11] warn that if thinking aloud is not completely free, it may interfere with task performance. Besides, speech planning requires from a participant additional time and effort, which affects the eye movement behaviour. The negative impacts of verbalisation can be minimised by creating comfortable conditions for the participants and giving them appropriate instructions. For our study, verbalisation of participants' thoughts was crucial, because we wanted to observe the processes of reasoning and knowledge building going beyond visual attention and perception. At the same time, eye tracking was also necessary, as we wanted to see how the participants used visual displays in their reasoning and knowledge building. To avoid altering the natural behaviour of the participants, we designed our experiments in accord with the recommendations of Holmqvist et al. [HNA*11].

3.3. Methods for analysis of eye tracking data

Standard software provided together with eye-tracking equipment usually transforms raw data into sequences of fixations and saccades. The results are visualised in the form of gaze plots where dot symbols and connecting lines represent the fixations and saccades, respectively [HNA*11, Section 8.1]. Durations of the fixations are often represented by symbol sizes. Weighted densities of fixations form attention maps [HNA*11, Section 7], where higher densities correspond to longer cumulative duration of fixations.

Various methods and measures applied in eye movement analysis are systematically described in the book by Holmqvist et al. [HNA*11]. Blascheck et al. [BKR*17] survey contemporary approaches to visualisation of eye-tracking data in connection to categorisations of stimuli (static versus dynamic, 2D versus 3D, passive versus active context), viewing tasks, collected gaze data (single versus multiple participants, 2D versus 3D gaze recordings, presence or absence of predefined AOIs), and types of visual representations. While spatial aspects of eye tracking data are addressed, to some extent, by traditional gaze plots and attention maps, analysis of the temporal dynamics requires additional tools. A space-time cube [KW13] can be

useful when data volumes are small. A more scalable approach involves data aggregation by time intervals and clustering of the time intervals by similarity of the aggregate features [AABW12].

Attention maps can be used for identifying AOIs at a certain level of detail. Sometimes AOIs identified at different levels are organised hierarchically [BKR*16]. Another approach to identifying AOIs is to apply methods of trajectory aggregation in which fixations are grouped in clusters by spatial proximity [AABW12]. Transformation of continuous coordinates of fixations to a discrete set of AOIs reduces the precision of the data but increases the possibilities for interpreting complex gaze trajectories. Repeated transitions between AOIs can be aggregated in a weighted graph and then represented either as a flow map on top of the stimulus [AABW12] or using different graph layouts without context [BSBE17]. To relate visits of AOIs to periods of analysis, a visualisation technique called scarf plot [RD05, KHW14] is proposed. Distinct colours are assigned to AOIs and then projected onto the timeline, thus depicting the times and durations of fixations within different AOIs and sequences of visits of the AOIs.

According to the goals of our study, exploratory visual analysis of the collected eye tracking data was more appropriate than computing numeric measures and statistical analysis. For the visualisation, we combine advantages of the existing approaches. We adapt scarf plots for representing continuous positions in the stimulus space (Section 5.1.1). We also determine individual AOIs of test participants and analyse them using flow maps and transition graphs (Sections 5.1.2, 5.1.3), separately considering aggregates of long saccades, which are deemed to indicate the process of global exploration [FLMB11].

4. Our experiments

To study the processes of abstractive perception, generalisation, reasoning, and knowledge construction, it was essential to involve participants in realistic scenarios of data analysis, where data and tasks are not trivial and data patterns are not obvious. Therefore, we created visual displays of real data referring to several phenomena and asked the participants to judge whether and how the phenomena are related. *The only purpose of the given tasks was to engage the participants in non-trivial data analysis.* As mentioned in Section 2, we had no ground truth (i.e., correct answers) for assessing the judgements. The answers as such were not of interest for us, but we wanted to observe the process of gaining knowledge needed for giving an answer.

4.1. Visual stimuli and tasks for participants

We prepared three analysis scenarios with the use of maps, where each map portrayed geographic distributions of two or more phenomena (Figs. 1, 2, 3). The data concerning

the phenomena come from organisations that have been our partners in a EU-funded research project. For presenting on the maps, the original detailed data were spatially aggregated by regular grids with square cells. The maps represent different phenomena and different geographic areas, while the visual encoding techniques are similar, so that participants could spend less time on learning the visual encoding and focus more on the data exploration. One phenomenon in each map is represented by means of the choropleth map technique, where values of a numeric attribute are encoded by the degrees of darkness of the colour in which areas (grid cells in our case) are painted. Numeric attribute values referring to another phenomenon are represented by proportionally sized circular symbols drawn inside the grid cells. One map contains simple circles, each representing a value of a single attribute. The other two maps contain pie charts with sector sizes representing values of two or three attributes. In both maps, the attributes represented by the pie charts are summable into meaningful totals; hence, the pie sectors represent not only the values of the attributes but also their proportions in the totals. Besides the grid-based data, each map also includes a geographical background showing other geographic phenomena, such as roads, populated places, waters, shorelines, etc.

4.1.1. Scenario 1

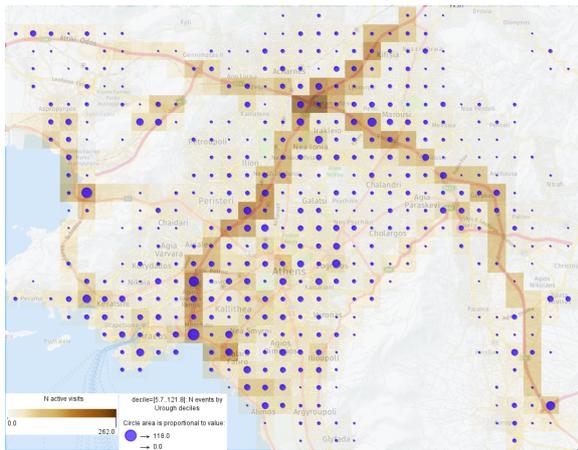


Figure 1: The map of the first scenario. The geographic area is Athens (Greece). The grid cell sizes are 1x1 km. The painting of the cells in shades of brown represents traffic intensity. The sizes of the dark blue circles represent counts of particular driving events (high roughness of speed variation).

The map for the first scenario (Fig. 1) portrays traffic-related phenomena, specifically, aggregated data describing movements of 536 GPS-tracked business vehicles on the territory of Athens (Greece) during one day. The painting of the grid cells represents the daily counts of the visits of the cells

by the tracked cars, which can be treated as the traffic intensity. As can be seen in the map, the spatial distribution of the counts is strongly related to the traffic network. The circular symbols represent the daily counts of a particular kind of driving events, namely, high roughness of the vehicle speed variation.

The task for the first scenario is to judge how much the spatial distributions of the traffic intensity and the driving events are correlated, whether there are significant discrepancies between the distributions and, if so, where.

4.1.2. Scenario 2

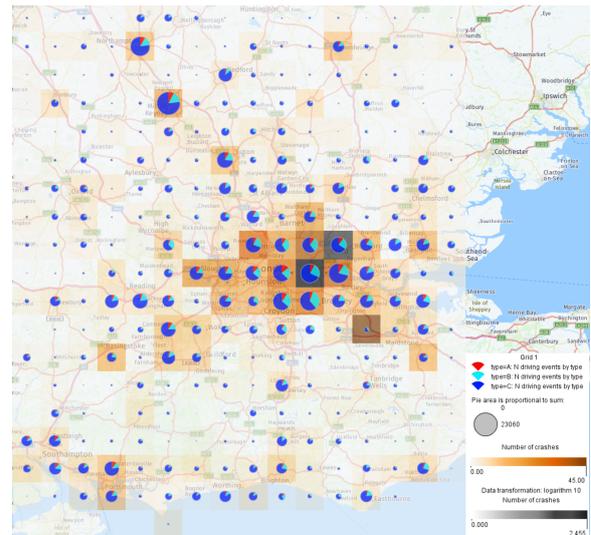


Figure 2: The map of the second analysis scenario. The geographic area is a part of England, including London. The sizes of the grid cells are 10x10 km. The painting of the cells in shades of brown represents counts of vehicle crashes. The pie charts represent counts and proportions of three types of driving events: high acceleration (red), harsh braking (cyan), and sharp cornering (dark blue).

The map for the second scenario (Fig. 2) also portrays traffic-related phenomena. The data come from a car insurance company. The painting of the grid cells represents counts of car crashes. Two cells have exceptionally high numbers of crashes, namely, 285 and 72, while the counts in all other cells range from 0 to 45. Encoding of the counts by proportional degrees of colour darkness would result in a map where all but two cells are nearly white and the spatial distribution of the crashes is not visible. It could be possible to apply logarithmic transformation to the data before encoding them by colour shades; however, this would exaggerate small values, distort the visual appearance of the spatial distribution, and reduce comprehensibility of the representation. To deal with this problem, we encoded the values from 0 to 45 by proportional darkness of shades of brown and

used shades of grey for the two extreme values after applying logarithmic transformation to them. When the map was presented to the test participants, the specifics of the encoding and the reasons for taking this approach were explained to them. Hence, the participants were aware that the grey-painted cells have extremely high numbers of car crashes.

The pie charts represent counts and, at the same time, proportions of three types of driving events registered by special sensors installed in the ensured cars in consent with the car owners. The event types are high acceleration, harsh braking, and sharp cornering.

The *task* is to judge whether the distribution of the three kinds of driving events in relation to the distribution of the crashes suggests that some kinds of events may be associated with a higher probability of a crash, overall or in particular places.

4.1.3. Scenario 3

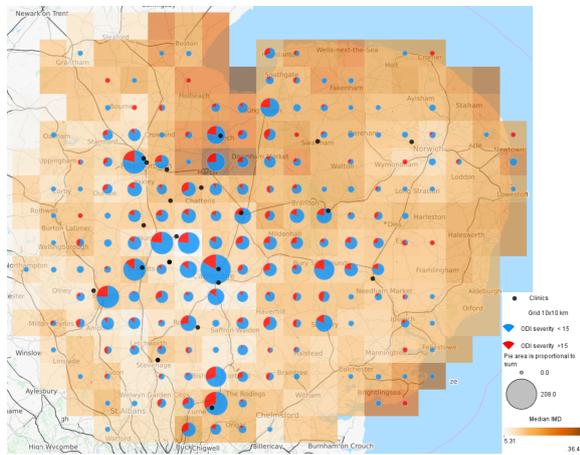


Figure 3: The map of the third analysis scenario. The geographic area is the East of England. The sizes of the grid cells are 10x10 km. The painting of the cells in shades of brown represents the values of the index of multiple deprivation, a relative measure of area poverty. The pie charts represent counts of individuals whose homes are located in the cells and who visited clinics to be tested for a particular health problem. The pie sectors represent counts and proportions of individuals with high (red) and low (blue) severity of the problem. The black dots show the locations of the clinics.

The map for the third scenario (Fig. 3) portrays phenomena related to people's health and well-being. The cell painting represents values of the index of multiple deprivation (IMD), which is a relative measure of area poverty adopted in the UK. Higher values of IMD correspond to more deprived (i.e., poorer) areas. The pie charts represent counts of individuals whose homes are located in the cells and who were tested for presence of a health disorder called obstructive sleep apnoea (OSA). The pies have two sectors,

coloured in red and blue, that represent counts and proportions of patients having high and low severity of OSA symptoms.

The testing for OSA is done using a special device which takes measurements while a person is sleeping. This requires a person to come twice to a clinic providing such devices, first for receiving a device to take it home and then for returning the device with the recorded measurements. Hence, a person who wants to be tested needs to spend time and money for travelling. Health care professionals are interested whether living conditions affect people's willingness to perform a test. They suspect that people living in poor areas may be less disposed to take the trouble to travel twice to a clinic in order to perform the test. The people may tend to ignore OSA symptoms and decide to undergo a test only when they feel serious problems.

In reference to this supposition, the *task* for the participants of our experiments was to judge whether the distribution of the tested people and/or the severity of their OSA condition may be somehow related to the deprivation level in the area in which they live.

In addition to the deprivation and OSA data, the map of scenario 3 includes an information layer with the locations of the clinics where the testing devices are taken. The clinics are represented by black dots. The additional information makes the analysis task quite challenging, since it is hard to figure out whether the distribution of the OSA data is more associated with the area deprivation levels or with the distribution of the clinics.

4.1.4. Commonalities and differences between the scenarios

The maps designed for the three scenarios differ in the level of complexity of the information contained. For scenario 1, the map shows distribution of two phenomena: traffic (choropleth map) and driving events (proportionally sized circles). The map of scenario 2 is more complex. It involves two visualisation techniques, as in the first map, but one of them is more complex (pie charts versus simple circles). The pie charts encode information about events of three different types; hence, there are five spatial distributions to investigate: crashes, each of the three event types, and the joint distribution of all three event types. The map of scenario 3 employs the same visualisation techniques as for scenario 2. The pie charts have fewer sectors than in the map of the second scenario, but the map includes an additional information layer and is therefore more complex.

The analysis scenarios also differ in the difficulty of finding significant patterns and making inferences concerning the relationships between the phenomena. We assess the first scenario as the easiest (the patterns are quite obvious) and the third scenario as the most challenging. The scenarios were presented to the participants in the order 1 – 2 – 3,

so that the complexity level increases as the participants get more familiar and experienced with the visual representations and the tasks.

4.2. Conduct of the experiments

4.2.1. Participants

For participating in the experiments, we recruited (through our network of contacts) eight well-educated professionals having analytical capabilities and willing to perform the tasks earnestly and carefully. The number of the participants may seem too small compared to what is recommended in the literature (e.g., [HNA*11, Section 3.2.7]) for obtaining enough data for a valid statistical analysis of various numeric measures. However, as we did not plan to do statistical analysis of any performance indicators, eight participants were sufficient for our exploratory study, where we aimed at gaining insights through qualitative analysis. This kind of analysis requires thoughtful examination of each analysis session, which limits the number of sessions that can be analysed.

As we wanted to observe diverse approaches to fulfilling the tasks, we involved participants with different kinds of background knowledge, skills, and professional experience (Table 1). Two participants are data science practitioners (ds), four are computer science researchers specialising in machine learning (ml) or visual analytics (va), and two are graphics designers (gd). The duration of the professional activity ranges from 3 to about 30 years. Only two participants had previous experience in analysing spatial data. Nevertheless, all participants were able to understand the contents of the maps and the analysis tasks and to perform sensible analyses. Three participants are female and five are male. The age ranges from 30 to 56 years with the median at 39. Hence, the sample of participants was sufficiently diverse for expecting diverse strategies of analysis. However, we did not intend to investigate how the characteristics of the participants would affect their approaches to the tasks, but we were interested to see and compare the approaches themselves.

We are far from considering this group of participants to be representative for the whole population of data analysts, and we do not claim that our experiments revealed all possible strategies of visually supported data analysis. We expect that similar studies involving various participants will be conducted in the future exposing a large variety of ways to use data visualisations in analytical reasoning.

4.2.2. Materials used in the experiments

For each scenario, we have prepared a sequence of three pages. The first page in each scenario contains a text to be read by the participants at the beginning. It includes an introduction into the analysis subject (i.e., the phenomena to be analysed), a description of the data to be analysed, and an exposition of the analysis task. The second page is meant for the familiarisation of the participants with the map to be used

Table 1: List of participants

N	area	sex	age	experience in	
				data analysis	data vis
1	va	f	50-60	> 20 years	> 20 years
2	va	m	50-60	> 20 years	> 20 years
3	ds	m	50-60	> 20 years	< 5 years
4	ds	m	40-50	> 10 years	< 5 years
5	va	m	30-40	5-10 years	5-10 years
6	ml	f	30-40	5-10 years	< 5 years
7	gd	f	30-40	< 5 years	< 5 years
8	gd	m	30-40	< 5 years	< 5 years

for the analysis. It includes an image of the map and an explanation of the visual encoding. The third page is meant for performing the analysis. It contains the same map image as the second page, a brief instruction for the participants, and a list of five movable markers (Fig. 4, top), with which participants were supposed to mark observed patterns on the map in the course of the analysis. An example of marker placement is shown in Fig. 4, bottom. The tool enabling placement of markers on an image was earlier developed for conducting user evaluations of visualisations [CAA*21].



Figure 4: Top: The appearance of a task page. Enlarged is the fragment containing movable markers for putting on the map. Bottom: An example of marker placements on a map.

Here is an example of an instruction shown on a task page:

“Your task is to examine whether the distribution of the three kinds of driving events in relation to the distribution of the crashes suggests that some kinds of events may be associated with a higher probability of a crash, overall or in particular places. Please identify and mark in the map several spatial patterns you deem important or interesting, or surprising. To mark a pattern, please, drag one of the numeric labels to the approximate location of the pattern on the map.” The first part of an instruction is formulated specifically for each scenario (we have cited the instruction from scenario 2), but the essence is the same: determine whether and how the phenomena represented in the map are related.

The second part of the instruction is the same for all scenarios. A participant was suggested to mark up to five places on the map deemed to exhibit important or interesting patterns. In this way, we wanted to induce the participants to examine the distributions with high attention and to ponder what parts of the visualisation provide the most relevant information for characterising the analysis subject and directing their reasoning. It was not planned to assess the correctness or importance of the patterns marked.

4.2.3. Organisation and implementation of the experiments

For collecting empirical data, we used eye tracking equipment Tobii Pro X2 and accompanying software. The experiments with each participant were conducted separately from others. At the beginning, the participant was informed about the purpose of the study, read the experiment description, and signed the informed consent form. The eye tracking equipment was demonstrated, and then the participant consecutively performed three sessions of analysis in the order of increasing complexity. Before each session, a facilitator explained the analysis scenario: the phenomena to analyse, the data shown on the map, and the research questions concerning relationships between the phenomena. Each session began with calibration of the eye tracker. After that, the participants read the text description of the scenario, then viewed the map and got familiar with the visual encoding of the data. The final step was the visual analysis. The participants were asked to think aloud and explain their process of analysis. The utterances of the participants were recorded using the in-built voice recording tool of the eye tracker software. The facilitator was present during the whole session answering any arising questions about the data, the representation, or the tasks.

In this way, we collected eye tracking data and voice records from 24 analysis sessions, i.e., 8 datasets for each scenario. We provide the eye tracking data to interested researchers for further investigation (<http://geoanalytics.net/and/ETexperiments2020/>), but we refrain from providing the voice records to preserve the anonymity of the study participants.

5. Data analysis methods

In the experiments, we collected two distinct types of data: eye tracks and audio records. The eye tracks are structured data suitable for analysis with the help of computer software. According to the exploratory character of our study, we performed visual analysis of these data using different visualisation techniques described in Section 5.1. The audio records are unstructured data that need to be listened, interpreted, and investigated by an analyst. Our approach to analysing these data is described in Section 5.2.

5.1. Methods for exploration of the eye tracks

Analysis of eye-tracking data usually focuses on examination and comparison of scanpaths, i.e., trajectories (paths) of the eyes while viewing an image or scene. A common approach involves defining Areas Of Interest (AOIs) in the visual stimuli, transforming the trajectories into sequences of visits of the AOIs, and assessing and comparing the sequences. However, the usual approaches were mostly unsuitable for our study for the following reasons:

- We expected that each participants would do the analysis in his/her own way and have his/her own AOIs. It was not our goal to identify common AOIs or to compare AOIs of different participants.
- It was also not relevant to our study to assess the sequences of visiting AOIs and to compare scanpaths of different participants.
- Our main goal in analysing eye tracks was to identify generic analytical operations performed by the participants. By “generic” we mean that an operation is defined as a class, in abstraction from specific AOIs and visiting sequences, and even in abstraction from specific visual stimuli.

Our expectation was that generic analytical operations can be identified by finding repeated patterns of eye movements across the participants and scenarios. As we strove to identify the operations in abstraction from specific AOIs and sequences, we required analysis techniques that would enable different ways and levels of abstraction. We found the following techniques to be suitable for our purposes:

- a combination of a continuously coloured image and a scarf plot representing eye tracks as colour variation (Fig. 5);
- individual gaze flow maps (Figs. 6) with participants’ individual AOIs and flows (i.e., aggregated moves) between them;
- individual transition graphs (Fig. 7) with participants’ individual AOIs represented as abstract graph nodes.

The first technique allowed us to view eye movements in abstraction from positions in the images. In viewing eye tracks as variation of colours, we were interested not in the

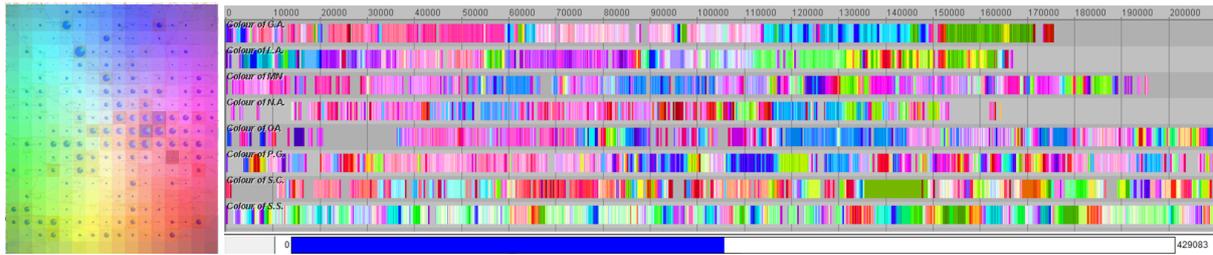


Figure 5: A combination of a continuously coloured image and a scarf plot.

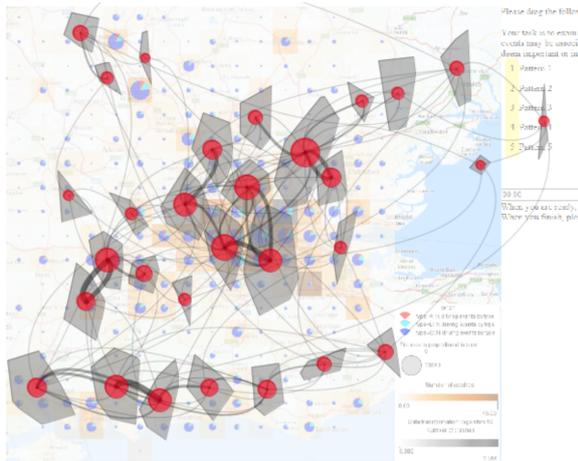


Figure 6: An example of an individual gaze flow map.

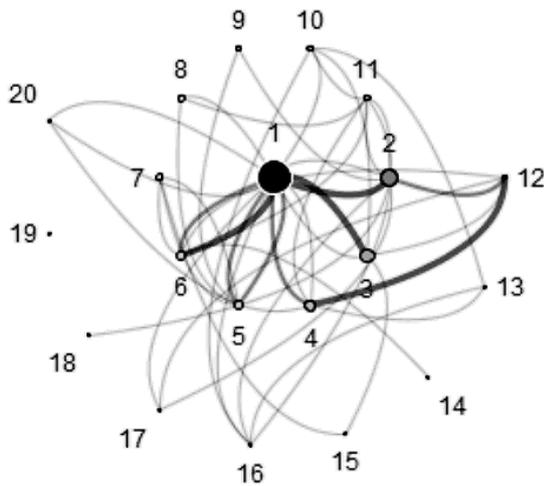


Figure 7: An example of an individual transition graph.

colours themselves but how they change over time. The second technique allowed us to consider eye movements in abstraction from absolute or relative times and from sequences

of visiting AOIs. The third technique enabled yet a higher level of abstraction, in which we could not only disregard specific positions in the images but also unify the individual AOIs of all participants from all analysis scenarios.

5.1.1. Scarf plots

Scarf plots [RD05] is one of traditional techniques for analysing scanpaths. It is a display where one dimension represents time and where scanpaths are represented by segmented bars stretching along the time dimension. Bar segments correspond to visits of different AOIs. The segments are painted in colours corresponding to specific AOIs or to categories or groups of AOIs. In our analysis, we wanted to avoid defining discrete AOIs. Instead, we applied near-continuous colouring to the image area, as shown on the left of Fig. 5. By near-continuous we mean that distinct colours were generated for small squares rather than for each pixel of the image, to avoid generation of an excessive number of unique colours. The bars representing the eye trajectories in a scarf plot are variably coloured according to the positions in the image where the participants were looking over time, as shown on the right of Fig. 5. The colourless fragments correspond to the gaze being outside the map image, when, for example, the participants read the instruction or took one of the pattern markers (Fig. 4) for moving onto the map.

While the colours in the bars indicate where the participants were looking at, the positions as such are not essential for our analysis. We want to discover repeated patterns of eye movements regardless of where specifically the gaze was directed. In a scarf plot, the movements are represented by changes of the colours. Hence, we need to investigate and interpret the ways in which the colours change disregarding the colours themselves. For example, small fluctuations of the colour along a time interval means that a participant focused the attention on some area, an abrupt change of the colour means that the attention was switched to a different area located far from the one attended previously, and gradual changes may mean attentive, systematic scanning. Similarly coloured spots along a bar indicate re-visiting of a place, and alternation of two groups of similar colour shades indicate comparison of two places.

5.1.2. Gaze flow maps

A flow map is a cartographic technique used for representation of aggregated movements [SMKH09, KO10]. Directed linear symbols connect movement origins and destinations, and symbol widths encode amounts or frequencies of the movements. Earlier it was proposed to apply the flow map technique to eye tracking data [AABW12]. To generate a flow map, the image used in eye tracking experiments was partitioned into irregular polygonal areas by means of a method that took into account the distribution of the eye fixation points over the image space [AA11]. For the sake of making comparisons between eye movements of different participants, a common image partitioning was used to aggregate the eye trajectory of each participant.

In our current study, we do not pursue the goal of comparing individual eye movements, and we would like to avoid transforming the individual eye tracks into uniform aggregated representations based on a common image partitioning. We take it for granted that each participant may find his/her own areas of interest in an image and thus generate his/her unique distribution of eye fixation points. We want to generate individual gaze flow maps respecting the individual distributions of the eye fixations. Furthermore, we find that partitioning, in which the image space is fully covered by polygons, is not well suited for our purposes. It is more appropriate to aggregate the eye tracks by areas enclosing spatial clusters of fixation points, which correspond to repeatedly visited individual AOIs. To generate such areas, we adapted the method proposed for extraction of people's individual significant places from trajectories of movement in the geographic space [AAFJ16]. In essence, the method searches for groups of stop points fitting in circles of a specified maximal radius, which is chosen according to the expected average size of a place. Intersections of two or more circles are treated as places of larger sizes. The adaptation of the method to the eye tracking data consisted in finding an appropriate circle radius. After several trials, we found a suitable value, such that the resulting "individual places", i.e., the individual AOIs mostly cover from three to nine adjacent grid cells. The AOIs are represented by polygons, as shown in Fig. 6. Since the AOIs are defined based on clusters of fixation points, some fixation points are not contained in any AOI.

After obtaining the polygons delineating the individual repeatedly visited AOIs, we aggregated each eye track by counting the transitions from each AOI to every other AOI. The aggregated transitions are represented on a gaze flow map by linear flow symbols, such as curved lines in Fig. 6 (other shapes can also be used, e.g., half-arrows [AABW12], as suggested by Tobler [Tob87]). The widths of the lines can encode the absolute counts of the transitions or their relative frequencies, as in Fig. 6. The relative frequencies of visiting the AOIs are represented by proportional sizes of the circular symbols drawn inside the corresponding polygons.

5.1.3. Transition graphs

In the transition graphs (see Fig. 7), the vertices, also called nodes, correspond to the individual AOIs, and the transitions between the AOIs are represented by directed and weighted graph edges, also called links. The vertices, too, have weights representing the frequencies of the visits to the corresponding AOIs. For a uniform representation of eye tracks of different participants from different scenarios, we rank the AOIs of each participant in each scenario by the visit frequency, label them by their ranks, and take top N AOIs (where N is some chosen number) for each combination scenario \times participant together with the transitions between these top N AOIs. We visualise the graphs of different sessions using exactly the same layout of the vertices, namely, a spiral layout where the vertex corresponding to most frequently visited AOI is positioned in the centre and the remaining vertices are positioned on concentric circles around the central vertex in the order of their ranks. The vertices are labelled by the ranks.

Please note that our approach to the construction and use of transition graphs differs from the usual way of using transition graphs for investigation and comparison of eye movements of multiple people, as described, for example, by Blascheck et al. [BSBE17]. Usually, the nodes in multiple graphs correspond to a standard set of AOIs that are assumed to be common for all participants of an eye tracking study. In our approach, the nodes of each graph correspond to a unique set of AOIs, while the uniform ordering and labelling within each set of AOIs make the individual graphs comparable. With such graphs, we can look for abstract structural patterns irrespective of the node semantics.

The transition graphs contain no information for explaining the occurring structural patterns. To obtain explanations, one needs to match the graph nodes and links with the corresponding AOIs and flows in the gaze flow maps and review the session records listening to what the participants were saying about these AOIs. Although this can be done, we do not strive at identifying all individual reasons why some AOIs gained more attention or some transitions occurred more frequently than others. Instead, we exploit the high level of abstraction provided by the transition graphs to find common patterns of eye movements that occurred in different sessions regardless of the distinctions between the visual stimuli and between the individual sets of AOIs. Possible common patterns are, for example, comparison of one or a few AOIs with many others, or careful comparison of two AOIs, or existence of "reference" AOIs to which the attention repeatedly returns after attending other AOIs.

5.2. Investigation of the session records

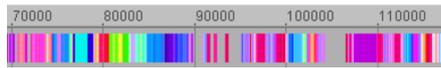
The software accompanying the eye tracker provides the function of replaying a recorded session. The voice comments are reproduced in parallel with replaying the eye

movements; hence, it is possible to identify where a participant was looking when making an utterance. It is also possible to see the times of the utterances and eye movements. We used this possibility to match the session records to the scarf plots of the gaze movement.

The analysis of the session records was done in three steps. In the first step, we replayed each session carefully listening to the recorded comments, viewing the corresponding gaze fixations and movements, and making notes in a text document. The notes characterise what the participants said and how their gaze moved at these and other times, also when they did not speak. We did not transcribe all utterances literally but mostly conveyed their meanings in shorter notes. We also noted the participants' pattern-marking activities. Here is an example of a note: "Focuses on a region; explains the pattern (high number of events and relatively low traffic) referring to the geo background (road density); does not place a marker".

Scans the map (different parts), consults the legend, clarifies the colour encoding of the event types. Says that cornering seems to be the most popular event in London, acceleration is not that much. Braking is also popular. "Do they have so many sharp curves in London?"

- Long eye movements, often beyond the map.



Focuses on the northern part of the map. Finds several places with much more events and accidents than around (finds them surprising because they are not in the centre), marks 3 places.



Investigates the central area again, looking for proportions of different event types. Finds a sub-area with more braking events than in other places, marks the area.

- While describing and marking, looks also at other places, apparently, for comparison and checking if the places in focus are indeed unusual.



Figure 8: An excerpt from a session description document.

The second step was done after we created the scarf plot visualisation for the eye tracking data. We replayed the sessions again. For each note in the session description document, we identified the time interval of the described activity in the session record and then determined the corresponding segment of the scarf plot of this session. We cut this segment from the plot using a screen capture tool and pasted it in the session description document below the note it refers to. As an example, an excerpt from a session description document is shown in Fig. 8.

In the third step, we carefully reviewed, interpreted, and synthesised all notes. We did not do formal coding of the session records, i.e., we did not assign standardised labels to individual notes. Instead, we immediately put our interpretations in a separate document, which enabled us to organise

them systematically. The contents and structure of the analysis document are reproduced in Appendix B. The notes and the corresponding scarf plot fragments from the session description documents were also used for finding and interpreting patterns in the scarf plots.

6. Analysis results

In the following, we report our finding obtained by analysing the eye tracking data and the session records.

6.1. Findings concerning eye movements

During the investigation of the session records (Section 5.2), we found out that the participants verbalised only a small fraction of their analytical activities. Particularly, we found evidences of doing much larger number of assessments and comparisons than it was explicitly mentioned in voice. This means that empirical studies of visual analysis processes cannot rely solely on using the think-aloud protocol. Analysis of eye movements can bring valuable information about tacit analytical activities. We analysed the eye movement data in conjunction with reviewing the session descriptions, where we not only cited the participants' utterances but also made notes about their gaze movements, including the times when they did not speak (see Fig. 8). When necessary, we also replayed fragments of the session records, to double-check our interpretations.

Table 2: Patterns of colour variation and their interpretation

Pattern example	interpretation
	image scanning
	exploration of a region
	examination of a place and around
	widening of the focus region
	narrowing of the focus region
	pairwise comparison
	comparison of one to many
	comparison of multiple places

Table 2 contains examples of re-occurring colour variation patterns we found in the scarf plots. The interpretations of these patterns (validated using the verbal data) show that they correspond to *elementary analytical operations*, but the scarf plots do not help us to identify any kind of higher-level structural patterns involving these elementary operations. As can be noticed in Fig. 5, there is no evidence of any preferred sequences of analytical activities. Therefore, it is reasonable to look for structural patterns in abstraction from the sequences, i.e., from the temporal arrangement of the eye movements. To do this, we constructed and investigated individual gaze flow maps.

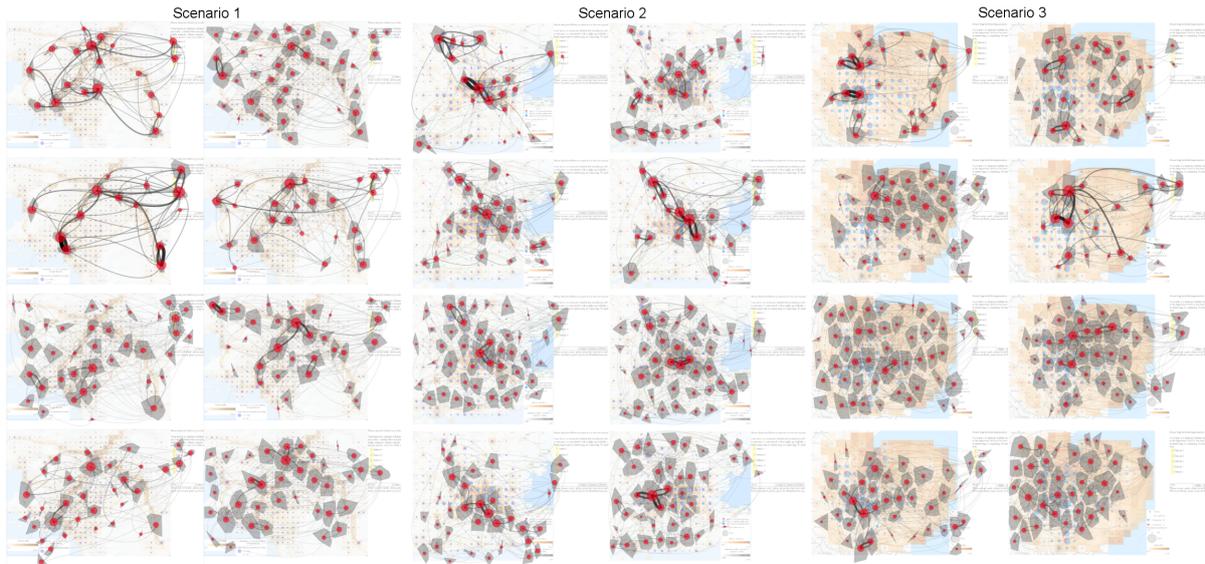


Figure 9: Gaze flow maps of all participants from all analysis scenarios.

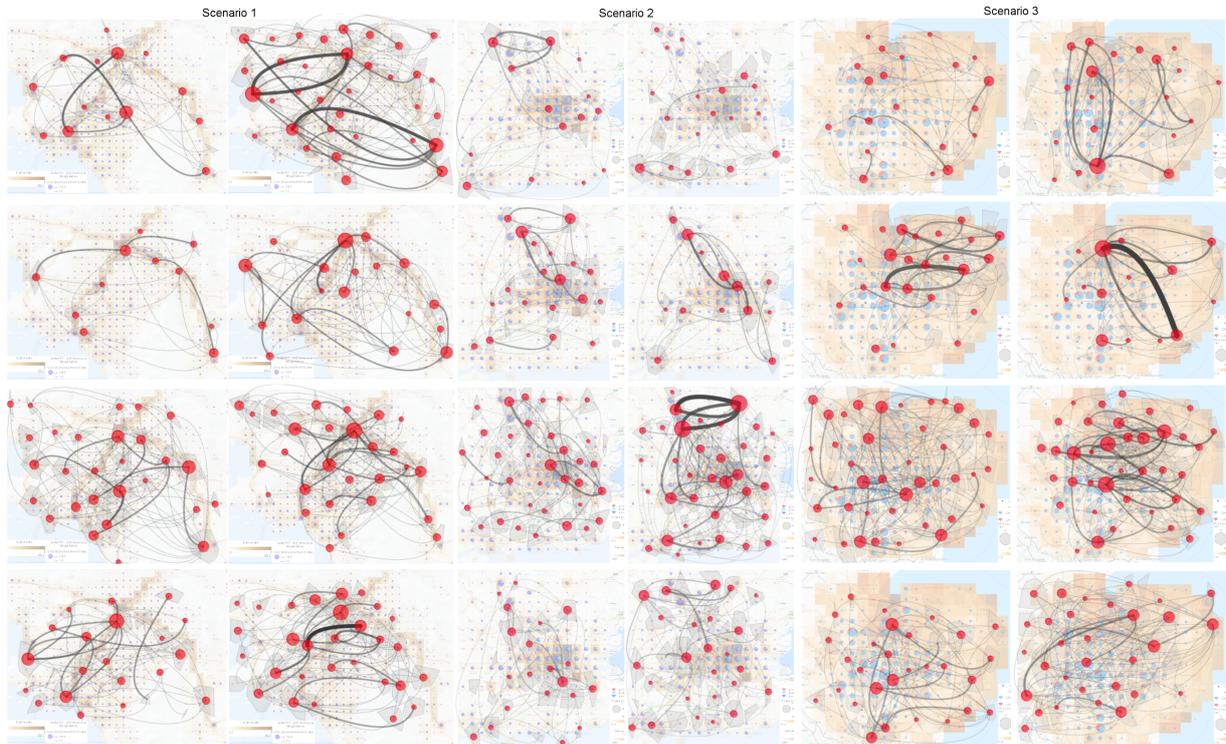


Figure 10: Gaze flow maps built from long saccades.

Figure 9 includes the gaze flow maps of all participants in all scenarios. We immediately notice the existence of two strategies of map exploration, which can be named *compre-*

hensive coverage and *selective focusing*. A map almost completely covered by individual AOIs demonstrates the comprehensive coverage strategy. The strategy of selective fo-

cusing is manifested by notably fewer AOIs and large areas in an image containing no AOIs. There are also several flow maps demonstrating a kind of intermediate strategy, in which a large part of the image is densely covered by AOIs but there are also sparsely covered or uncovered image parts. The choice of the strategy may be related to the analyst's experience and/or the properties of the spatial distribution represented on the map. However, analysis and explanation of individual differences are beyond the focus of our study.

What is also readily noticeable is that some maps contain pairs of strong links connecting two AOIs in two opposite directions. There are also groups where one area is strongly linked to two or even three other areas. Most often such structures include neighbouring AOIs. We investigated the image contents under each group of strongly linked AOIs, and we also checked what the participants were saying while looking at these parts of the images. We came to the conclusion that strongly linked AOIs that are close neighbours correspond to observation and examination of *regional patterns*. A few occurrences of strongly linked AOIs that are spatially separated correspond to observation of *distributed patterns*, i.e., several places with similar data.

In the maps demonstrating the selective focusing strategy we see two types of structures formed by interlinked neighbouring AOIs: cluster and alignment. Both types occur in the example flow map in Fig. 6: a cluster in the centre of the map and an alignment on the south. A cluster indicates exploration of a region. In most cases, AOIs making a cluster are not uniform in terms of the underlying data. From the participants' utterances while keeping their attention in such regions we have identified two types of analytical activities. One was comparing an AOI with its surrounding, which was mostly done for AOIs where data notably differed from the areas around. The other was assessing the differences between the places as possible evidence of the presence or absence of relationships between the phenomena under analysis. Obviously, this requires many comparisons, which are easier to do with spatially close places than with distant ones. This can explain the appearance of clusters of interlinked neighbouring AOIs. In fact, in all maps demonstrating the selective focusing strategy we see much higher frequencies of transitions between nearer AOIs than between more distant ones. What concerns the maps with comprehensive coverage, it is harder to notice significant differences between the frequencies of the near and far transitions, because the transition links are more numerous but mostly quite weak. Still, it can be observed that rarely occurring stronger links connect neighbouring AOIs.

Alignments of AOIs occur in some of the gaze flow maps corresponding to all three scenarios. In scenario 1, where the distributions under analysis are strongly related to the road network, AOIs are often aligned along the major roads. In scenario 2, several flow maps include alignments of AOIs along the southern coast of the territory. Similarly, align-

ments along the eastern coast occur in some flow maps corresponding to scenario 3. However, there are also alignments that are not related to features in the geographic background. Thus, diagonal alignments from the northwest to the southeast can be seen in several maps corresponding to scenario 2. These alignments are induced by the properties of the spatial distributions of the attributes under analysis: there are a few places located along the direction from London to the northwest where the attributes have notably higher values than around. In fact, the geography-related alignments of AOI were also caused not by the geography as such but by patterns in the spatial distribution of the data, which, in turn, were associated with the geography. Hence, we can infer that structures formed by multiple interlinked AOIs may correspond to intermediate-scale spatial patterns in the data distribution.

We also want to find out whether there are any interpretable patterns of the transitions between distant AOIs. Since distant transitions are less frequent than near ones, they are not well visible in the gaze flow maps showing all transitions. To have a clearer view of the distant transitions, we selected the eye saccades that are longer than 6-7 grid cells by means of filtering and created another set of gaze flow maps from these long saccades only. For the aggregation of the long saccades, we used only the AOIs located within the parts of the images showing the data under analysis, i.e., we excluded the AOIs covering the map legends, the pattern markers, and pieces of instruction texts. The resulting flow maps are shown in Fig. 10. The circle sizes represent the numbers of long saccades originating from the AOIs, and the widths of the flow lines are proportional to the counts of the long saccades connecting the AOIs. Please note that the parameters of the visual encoding are specific to each scenario.

In the flow maps of long transitions, we find two major patterns of links between AOIs. The first pattern is linkage by similarity, i.e., transitions between AOIs where data are similar in one or more aspects. Thus, in the first scenario, most of the long transitions are between AOIs with high traffic intensity, and in the second scenario – between AOIs with large total numbers of driving events. In the context of checking the existence of an overall correlation between distributions of two or more attributes, the purpose of such transitions can be to check whether similar values of one attribute correspond to similar values of the other attribute(s). Strong links, i.e., transitions with relatively high frequency, may indicate a lack of consistency between AOIs in terms of the relationships between the attributes. Examples can be seen in several flow maps from scenario 3, such as the topmost and rightmost map in Fig. 10 and the map below it. Here, the participants were comparing several areas with high deprivation and saying in their verbal comments that some of these areas have very few or no tested people while others have quite large numbers.

The second pattern of long transitions indicates laborious

scanning and, possibly, even frustration of a participant who cannot find clear evidence of the presence or absence of a relationship between phenomena. Such scanning behaviour can be seen in several maps from scenario 3, namely, the lowest two maps in the rightmost column in Fig. 10 and the map in the row 2, column 5. The recorded comments confirm that the participants had difficulties with finding clear patterns and making conclusions. Interestingly, the transitions in the scanning pattern mostly have horizontal or just slightly tilted orientation, like it could be during reading a text. This may mean that the participants had no clue where they could find relevant information and therefore scanned the image sequentially.

The gaze flow maps enabled us to discover meaningful spatial patterns in the participants' eye movement behaviours. We wonder whether there exist also common structural patterns of a higher level of abstraction that are not associated with spatial locations and spatial relationships, such as neighbourhood and orientation. We investigate this using transition graphs.

In the investigation of the gaze flow maps, we found out that transitions between neighbouring AOIs prevail over distant transitions. The structures formed by the short-range transitions are induced by the spatial relationships between the AOIs, i.e., these structures are highly space-dependent. It therefore appears unsound to consider these transitions in separation from the spatial relationships. The long-range transitions, on the opposite, are not associated with particular spatial relationships between AOIs. It appears quite reasonable to analyse these transitions disregarding their lengths, spatial orientations, and specific positions in the images. According to these considerations, we generate graphs of long transitions (Fig. 11) from the gaze flow maps that have been built from the long saccades (Fig. 10). For the transition graphs in Fig. 11, we took for each combination scenario \times participant 20 topmost AOIs according to the frequencies of the long saccades originating in these AOIs.

With such graphs, we can look for abstract structural patterns irrespective of the node semantics. Particularly, in Fig. 11, we see the following types of structural patterns:

- patterns of node weights distribution: "heavy centre", "weighty few", "no favourites";
- patterns of link weights distribution: "arterial channels" vs. "spider web";
- patterns of link configurations: "tight coupling", "fork", "confluence", "star", "chain".

"Heavy centre" means that one AOI was visited much more frequently than any other, "weighty few" means that there were two or more such AOIs, and "no favourites" means absence of large differences between the frequencies of visiting different AOIs. "Arterial channels" means that some transitions re-occurred many times while "spider web" denotes a relatively equal distribution of the transition

frequencies. "Tight coupling" denotes a pair of strong oppositely directed links between two nodes, which means that two AOIs were intensively compared to each other. "Fork" and "confluence" denote two or more strong links originating from or ending in one node, a "star" is a configuration where one node has strong links in both directions with several other nodes, and a "chain" is made by two or more strong links such that the destination of one link coincides with the origin of the other. Different patterns may be present in the same graph. For example, the top left graph in Fig. 11 contains a "fork" pattern made by the links 1-4 and 1-7 and a "chain" of links 2-3 and 3-1.

It is important to remember that the graphs show us patterns of long transitions, i.e., when the AOIs that are compared are not close neighbours. High frequencies of some transitions may be due to the difficulty of such distant comparisons. When two or more AOIs that need to be compared cannot be seen simultaneously, an analyst, looking at one AOI, needs to remember what was earlier seen in another AOI. Memorising and recalling is more difficult and less reliable than direct observation of two AOIs together; therefore, multiple transitions between the AOIs may be required. To facilitate distant comparisons, visual analytics systems may equip analysts with special interaction techniques. In designing such techniques, it may be reasonable to take into account the types of patterns of long transitions that can be detected by analysing abstract transition graphs. In the following section, we shall discuss possible implications of different kinds of empirical finding for design of visual analytics systems, tools, and workflows.

6.2. Findings concerning reasoning and knowledge building

The overall goal of the analysis was to understand how the participants use the visualisations in analytical reasoning, particularly, how they transform their observations into an overall understanding of the phenomena reflected in the data. The main source of information required to fulfill this goal is the verbal data. The findings reported here were derived by analysing the session records (see Section 5.2) and, when appropriate, matching our observations to the patterns extracted from the eye tracking data. In describing the findings, we shall refer to the research questions introduced in Section 2. Figure 12 indicates what kinds of evidence have been involved in answering the research questions.

RQ1: *What type of decision making (bottom up or top down) prevails in data analysis aiming at generalisation?*

We remind that the participants of the study were asked to mark up to 5 patterns they deemed important or interesting (Section 4.2.2 and Fig. 4). The placements of the pattern markers (see Appendix A and Fig. 13) suggest that the participants paid quite much attention to salient features of the images, such as large symbols and dark shading. However,

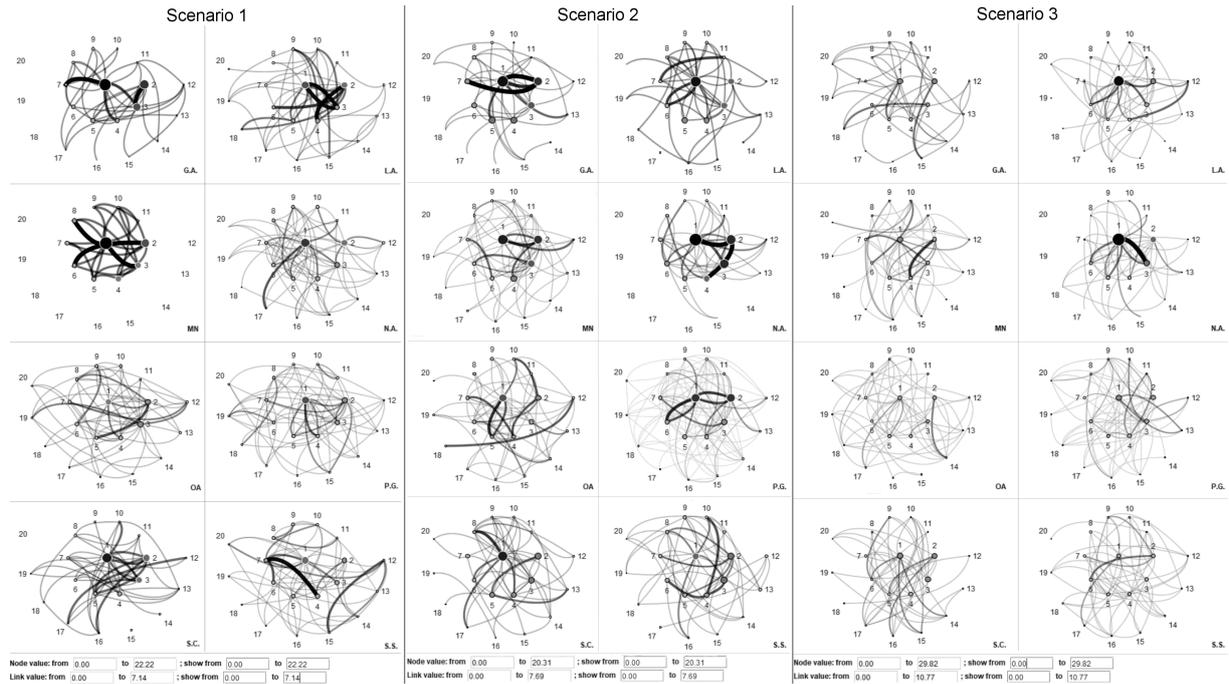


Figure 11: Graphs of long transitions of all participants from all analysis scenarios.

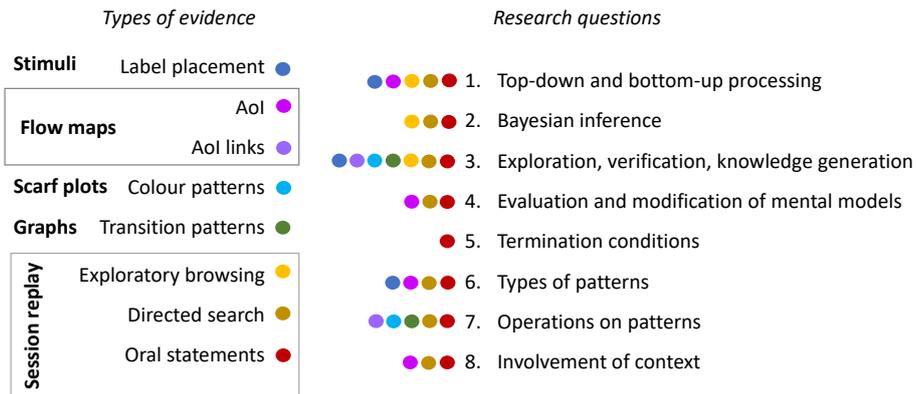


Figure 12: Types of evidence and their involvement in answering the research questions.

none of the participants was quick in drawing conclusions. The minimal duration of an analysis session was about 11 minutes (in scenario 1, the easiest one) and the maximal duration was more than 56 minutes (in scenario 3, the hardest one). In all sessions, we observed that the participants purposefully searched for information that could confirm or contradict the relatedness of the phenomena portrayed. This suggests that bottom-up image processing did not play an essential role in the analysis and did not affect the conclusions drawn.

RQ2: To what extent is the Bayesian inference model ap-

licable to the process of analysis involving abstraction and generalisation?

In none of the sessions the participants expressed their prior beliefs concerning the relationships between the phenomena; it is quite likely that they did not have any. In some sessions (mainly with scenario 1) the participants stated quite soon that the relationship seems to exist, but in the majority of the sessions the participants did not express definitive judgements until almost the end of the analysis. Obviously, these sessions cannot be modelled as Bayesian inference. In all cases when participants did express some

judgements, they always tried to verify their current beliefs by seeking counter-evidence and assessing whether the collected counter-evidence is strong enough to change their opinion. A few counterexamples were typically treated as local exceptions from a globally existing relationship, i.e., the participants did not reverse their beliefs. There was only one case (with scenario 2) when a participant explicitly changed an earlier expressed belief (the presence of weak correlation) saying that the phenomena are not correlated. Still, it cannot be said that the Bayesian inference model provides an adequate description of what we observed, as it does not account for a large part of analytical activities that took place. Thus, the participants made many observations and judgements that were not directly related to the question they were supposed to answer. For example, the participants noted relationships between driving events and street network in scenarios 1 and 2 and between numbers of tested patients and locations of clinics in scenario 3. Such additional observations were obviously important for developing the participants' understanding of the phenomena, but the Bayesian inference model appears too simplistic for describing this development.

RQ3: *Are the three types of activities (exploration, verification of hypotheses, and knowledge generation) observable in the analysis sessions?*

The exploratory activity was observable throughout the entire sessions. As mentioned earlier, we saw no evidence that the participants had prior hypotheses (which is another word for "beliefs") concerning the relationships between the phenomena, most probably, because they had no domain-specific knowledge. Also, there were not many sessions in which participants explicitly formulated the hypotheses they might have derived during map exploration. When this happened (it was in all cases the hypothesis that a correlation between phenomena exists), the participants marked or pointed out a few most prominent visual patterns supporting the hypothesis, then looked for counterexamples. This can certainly be seen as verification of the hypothesis. Finally, the participants concluded whether the counterexamples disprove the existence of the overall correlation, or they can be seen as local deviations from the globally existing relationship. This process of adopting or refuting the hypothesis can be seen as knowledge generation. Hence, the three types of analytical activities are identifiable in some sessions.

However, in the majority of the sessions we did not see a clear evidence that the analysis process was driven by hypotheses. What we observed can be described as progressive gathering of various pieces of information that could contribute to the participants' current understanding of the phenomena even without being directly related to the question the participants needed to answer. This process can be considered as knowledge generation based on exploration, but it would be wrong to say that the generated knowledge consists solely of verified hypotheses.

Hence, we detected two strategies of analysis: hypothesis-driven and compilatory. In the compilatory strategy, the participants considered possible relationships of the phenomena of interest (i.e., mentioned in the questions) with other phenomena, either represented in the maps or known by the participants. Based on all collected evidence about these various relationships, the participants could decide whether the phenomena of interest are related between themselves. The compilatory strategy of analysis was mainly associated with the comprehensive coverage strategy of map exploration detected in the gaze flow maps (Section 6.1 and Fig. 9).

RQ4: *Are model evaluation and modification activities identifiable in the analysis?*

According to [ALA*18], we use the term "mental model" in the sense of the analyst's understanding of the subject of the analysis. The activities on verification of hypotheses can also be categorised as evaluation of parts of the current mental model. In the result, the model was refined (the participants said that the relationship mostly exists but with some exceptions) or modified more radically (the participants concluded that the relationship does not exist). Thus, a single counter-pattern was typically judged as a local exception from a globally existing relationship. After finding several counter-patterns, participants might refine their models, e.g., say "In general, high number of driving events correlates with the number of accidents, except for the coastal area, where more events do not produce many crashes". As mentioned earlier, there was one case when a participant refuted the current model after finding multiple counter-patterns.

Besides evaluating whether some presupposition is true or false, the participants also assessed whether the knowledge they already have allows them to give a definite answer. Their mental models were not only changing due to confirming or refuting hypotheses but also extending due to adding new pieces of information.

Moreover, it appears that the explainability and credibility of a model were also important criteria of model evaluation. This can be deduced from the fact that the participants strove to explain the patterns they observed in the images using their prior knowledge, for example, awareness of the existence of many street crossings in a city or understanding that the number of the patients should depend on the overall amount of population in an area. Hence, the mental models of the analysis subject were also modified by retrieving relevant pieces of the participants' background knowledge and linking these pieces to new information obtained from the visualisations.

RQ5: *What are the termination conditions for the analysis process?*

As we did not limit the time for the analysis, the participants themselves decided when to stop in each session. We observed two types of situations in which the participants stopped: (1) they felt confident of their understanding of the

subject and ready to give an answer, either positive or negative, or (2) they felt that they have not enough information to come to a definite conclusion but did not see further possibilities to improve their understanding (this mostly happened in the sessions with scenario 3, and the participants usually complained about the lack of information about the spatial distribution of the resident population). The first situation corresponds to considering the state of a model as appropriate [ALA*18], whereas the second situation, i.e., a failure of the analysis, is not reflected in the model-building workflow discussed in [ALA*18].

RQ6: *What types of visual patterns are observed and interpreted as data patterns?*

Motivated by the analysis tasks, the participants were focused on checking whether high values of one phenomenon co-occur with high or low values of another phenomenon. Translated in terms of the visualisations that were used, it was looking for correspondences between dark or light painting and large or small symbols. However, the participants formulated their comments mostly in terms of the data and not in terms of the visual elements. For example, it was more typical to note “many rough speed changes without dense traffic” rather than “large circles on light background”. It means that the participants immediately interpreted visual patterns as data patterns and based their reasoning on the data patterns. Only one participant explicitly said that s/he is looking for particular visual patterns: small or large symbols on dark or light background. It does not mean, however, that this participant had a unique approach. It is more likely that the other participants also performed mental translation of their target data patterns into properties of visual elements to look for but did not say this explicitly in their comments.

The visual relationships that formed important visual patterns can be divided in three groups:

- **cross-layer** relationships: co-location of particular visual features from two or more information layers in or near the same place of an image;
- **similarity-difference** relationships between visual features of different places;
- **spatial** relationships between places, in particular, neighbourhood, separation, and alignment.

The cross-layer relationships include not only correspondences between the representations of the phenomena under analysis but also between them and the background geographic features, such as streets, coastlines, and populated places. In scenario 3, the participants also noted patterns of proximity of cells with high numbers of tested people to the clinics providing the testing devices.

The similarity-difference relationships and the spatial relationships between places were jointly involved in forming more complex visual patterns of different spatial scale and configuration:

- **Local** patterns: small spots, sometimes limited to single cells of a grid, differing from the surrounding places.
- **Regional** patterns: areas made of multiple neighbouring cells with similar appearance.
- **Linear** patterns: alignments of cells with similar appearance, often along roads or coasts.
- **Distributed** patterns made of multiple disjoint places with similar appearance.

Hence, meaningful visual patterns arise from interplay of different types of visual relationships. As noted earlier, the participants immediately translated the visual patterns into data patterns, which were used in the reasoning.

RQ7: *What analytical operations are applied to discovered patterns?*

The operation that was usually applied immediately after finding a pattern was to *evaluate* the pattern. The following criteria of evaluation were the most common:

- Does the pattern suggest the presence or absence of the relationship?
- How strong is the pattern? (i.e., how frequently it occurs or how large is the territory where it is observed)
- How does this pattern relate to previously detected patterns? (similar or dissimilar, consistent or discrepant, more or less prominent)
- Is this pattern predictable or surprising?
- Is it explainable?

It is important to note that pattern detection was frequently followed by an attempt to *explain* the pattern, although this was not required by the analysis task. In doing this, the participants referred to the *geographical context* visible in the background map (roads, crossings, cities, etc.) and/or to their *background knowledge*. For example, in scenario 2, many participants associated high numbers of crashes and driving events with road crossings seen in the map background. Concerning the area of London, where street crossings are not visible in the map, participants referred to their background knowledge: “It is a city, in the centre there is usually more traffic and therefore more incidents ... more street crossings”; “There is much braking in the city centre because there are a lot of traffic lights”; “Not many acceleration events – not much space to accelerate”. In scenario 3, the participants noticed areas with high deprivation and very small numbers of OSA-tested people, but they did not immediately interpret this as a relationship between these two phenomena, saying that the small numbers of tests may correspond to small numbers of inhabitants in these areas.

Other analytical operations that we identified by replaying the analysis sessions were:

- *compare* patterns:
 - *pairwise comparison*: similarity or difference in terms of relationships between the phenomena; whether a relationship is expressed more or less prominently;

- *group comparison*: check if several places have a common pattern; select the most representative place;
- *seek patterns*:
 - *similar* to a particular pattern;
 - *opposite* to a particular pattern;
- *group and unite*: consider several similar patterns together as a single regional or distributed pattern;
- *refine*: note local deviations from a large-scale pattern;
- *generalise*: derive a general judgement concerning properties of a phenomenon or relationships between phenomena.

All these operations establish various kinds of relationships between patterns and in this way bring them together constructing an overall mental model of the analysis subject.

RQ8: *Is the involvement of context information identifiable in the analysis sessions? If context information is involved, how is this information used?*

In the sessions, we clearly saw the influences of the analysis context [KHL21], i.e., aims and tasks, on the analysis process: firstly, the participants purposefully sought task-relevant patterns and evaluated them in regard to the questions posed; secondly, as the participants were asked to mark important or interesting patterns (Fig. 4), they also evaluated patterns in terms of their importance or interestingness. We observed the involvement of the user context, namely, background knowledge, and the domain context, namely, the geographic phenomena and features visible in the maps, in the participants' attempts to explain the patterns.

As discussed earlier, we observed that the participants constructed their understanding (i.e., mental model) of the analysis subject not only from facts characterising purely the relationships between the phenomena the tasks referred to but also from information about the relationships of these phenomena to the geographic context and from relevant pieces of the background knowledge. Hence, the context information was involved in model building and was integrated in the mental model.

Moreover, we saw a clear evidence that the evolving mental model of the analysis subject, which can itself be considered as a part of the participant's background knowledge (and, hence, as a part of the user context), affected the further analysis process, particularly, the pattern seeking behaviour. For example, in scenario 3, the participants associated large numbers of OSA-tested people with proximity to clinics and then purposefully searched for other clinics to check this association. Overall, we observed many cases when the current mental model affected the kinds of patterns that were sought and considered by the participants, including targeted search for patterns not complying with the current model.

To summarise, the analysis context and the user context (namely, the current mental model of the analysis subject) directed the behaviours of the participants in the analysis process, and the user context and the domain context were

involved in the knowledge building, i.e., formation of the mental models of the analysis subject.

6.3. Summary of the findings

Let us summarise the most important findings we gained using different approaches to analysing the study data.

Visual exploration and pattern discovery

- There were two major ways of visually exploring a data display: *selective focusing* and *comprehensive coverage*.
- Analysts perceived visual patterns formed by relationships between visual marks.
- Pattern-forming relationships included cross-layer collocation, similarity-difference, and spatial neighbourhood, separation, and alignment (in the image space). Their *interplay* created patterns of different *spatial scale*.
- Analysts immediately translated visual patterns into data patterns.

Analysis and knowledge generation

- Analysts often strove to *explain observed patterns* using available contextual information and/or prior knowledge.
- The analysis process was not always driven by hypotheses. There existed two strategies: *hypothesis-driven* and *compilatory*. They tend to be associated, respectively, with selective focusing and comprehensive coverage in visual exploration of a data display.
- A mental model of the analysis subject was constructed not only from verified hypotheses and data patterns but also from relevant *context information*, including background knowledge.
- Verification of hypotheses or current mental models typically involved *targeted search of counter-evidence*.
- Mental models were evaluated regarding not only their faithfulness to the data but also *explainability* and *credibility*.
- Analysts might decide to *terminate the analysis process* not only when their mental model was deemed appropriate for fulfilling the task but also when they saw no possibility for improvement, e.g., because necessary information was missing. Unsuccessful attempts to improve the model were manifested in laborious scanning of the image.

7. Discussion

7.1. Links to the theoretical models

We explicitly related the research questions of our study to some of the existing theoretical models of the visual analysis process; see Section 2. Our answers to these questions are reported in Section 6.2. Let us now consider how consistent the answers are with the respective theoretical models.

The model of visualisation-based decision making [PBG*14, PCRHS18] differentiated decisions made based on bottom-up and top-down perception. Our eye tracking

data show that visually salient features of the displays, particularly, large symbols, attracted larger amounts of attention than other areas in the display; hence, bottom-up perception took place during the analysis. However, there were no verbalised decisions (i.e., judgements) that would be based solely on seeing the salient features. On the opposite, the participants carefully scanned the display and combined information from different parts of it, which means that the top-down processes played much more important role in the analysis. This agrees with the assertion of Patterson et al. [PBG*14] about the primacy of the top-down processes in visual analysis.

The view of the visual analysis as Bayesian inference [KWKH19, KKGH21] appeared to be over-simplistic in respect to the analysis tasks given to the participants in our experiments. The participants were not just supposed to check and, possibly, update their prior beliefs, but they were expected to make generalisations from the data. In our study, we were not primarily interested in the final results of the participants' analyses and how these results differ from their prior expectations. What we wanted to study was the *process* of analysis and reasoning. The Bayesian inference model cannot be seen as fully adequate for characterising this process.

Regarding the knowledge generation model [SSS*14], the exploration and knowledge generation activities were identifiable in the analysis sessions. However, only a few sessions conformed to the premise of the model that the analysis process is driven by hypotheses. More frequently we saw the compilatory strategy, when the participants collected various facts related in some way to the phenomena under analysis rather than just looked for positive and negative evidence concerning the correlation between the phenomena. While activities on verification of hypotheses were present in the analysis, we saw a wider range of assessment activities using a variety of criteria. Our observations also showed that generated knowledge was not limited to confirmed hypotheses, or "justified beliefs", but included relevant context information and links to prior knowledge. Another note is that the participants were typically not absolutely certain even in their final statements, i.e., their judgements were not fully justified. Hence, the knowledge generation model describes only some aspects of the knowledge generation process. As any model in general, it is a partial and simplified representation of the reality.

The model-building view [ALA*18] does not distinguish knowledge from hypotheses. The term "mental model" refers to the current state of the analyst's understanding of the analysis subject, including already confirmed information and what still needs checking. Some mental model exists at any moment of the analysis process, and it is constantly evaluated and modified. This is what we observed in our experiments. The participants strove to develop their mental models to a state appropriate for answering the ques-

tion. In some sessions, such a state could not be reached. The possibility of such a failure is not reflected in the model building workflow [ALA*18, Fig.2].

Both the knowledge generation model [SSS*14] and the model building view [ALA*18] do not explicitly represent analysts' efforts to explain their findings, the kind of activity that was quite prominent in our experiments. The participants' striving to find explanations can be qualified, according to the qualitative visual analysis model [KHL21, Fig.3], as "mapping the available information to a conceptual model which a user applies for reasoning". In terms of the model-building view [ALA*18], it can be said that the explanations of the findings were included in the participants' mental models. It is probable that the presence of such explanations is essential for a model to be judged as appropriate and should be added to the list of model appropriateness criteria [ALA*18, Definition A.2].

The qualitative visual analysis model [KHL21] emphasises the key role of the user context, analysis context, and domain context in the extraction of meaningful information from data. Our experiments fully agree with the statements of this model and with the representation of the process of qualitative analysis [KHL21, Fig.3]. However, this model does not describe the phenomena of abstraction and generalisation that occur during the analysis. In this regard, the pattern theory [AAM*21] can provide a relevant complement.

The participants of our study performed visual analysis by detecting visual patterns formed by relationships between visual marks and translating them into data patterns consisting of relationships between data items. The patterns can be formally described using the definitions of the pattern theory [AAM*21]. The participants were not given any tools for performing analytical operations on patterns [AAM*21, Section 6.2]; nevertheless, they performed a range of mental operations. The most frequent one was comparison of patterns. By means of the operations 'group and unite' and 'generalise', the participants integrated simple local patterns into patterns of higher scale and level of abstraction, which, in turn, were incorporated into the evolving mental models of the analysis subject. The generalisability of our findings concerning pattern types is discussed in Appendix C.

Hence, we found the model building view [ALA*18], the qualitative visual analysis model [KHL21], and the pattern theory [AAM*21] to be the most suitable for describing the processes of visual analysis and knowledge generation we observed in our study. None of the theoretical models covers all aspects of those processes; rather, the models should be considered complementary and used together.

7.2. Possible implications for visual analytics design

We would like to remind that the aim of our study was not to evaluate a particular visualisation design or technique. We did not intend to check whether maps with circles or pie

charts are good or bad for analysing spatial distributions, but we strove to gain knowledge about the process of visual analysis and to assess the possibility of drawing implications concerning desirable support of this process in visual analytics systems.

Analogously to our study participants, who sought data distribution patterns in order to gain knowledge about the phenomena represented on the maps, we obtain knowledge about the analysis process by finding patterns in the participants' analytical behaviours. The most common behavioural pattern was repeated re-visiting of individual AOIs. Two aspects are important here: first, each individual had a unique set of AOIs having some meaning for this individual and, second, the AOIs were visited repeatedly, some of them quite frequently, as the individuals compared and/or associated the data patterns seen in different AOIs. Since this activity is so common and appears to be very important, it may be good to support it by interactive tools. An analyst may wish to mark or encircle an AOI, at least to find it more easily later on. Another possible operation is to "copy" the content of an AOI and "paste" it in a dedicated workspace. This could be done for several AOIs the analyst wants to compare. The comparison may be easier to perform when AOIs are separated from their surrounding, which distracts the attention, and put closer to each other. This reduces the need in long-distance transitions between AOIs (Figs. 10, 11).

It may be beneficial for the analyst to obtain an automatically created characterisation of the data pattern contained in an AOI, as well as the relationships of this pattern to the overall distribution and to patterns in other AOIs.

The colour variation patterns in the scarf plots (Table 2) suggest that an analyst can spend some effort on establishing the boundaries of a pattern observed by extending and shrinking the current area of the visual focus and examination of its surrounding. This activity can be supported by interactive tools with which an analyst could conveniently vary the extent of an AOI and observe how the summary characterisation changes in response.

The idea of a pattern comparison workspace may not be suited equally well to all kinds of comparison. In the transition graphs (Fig. 11), we saw a recurring pattern of comparing one AOI to many others. It would not be convenient to do such comparison through bringing the contents of all these AOIs to the dedicated workspace. We envisage that an intelligent system could automatically scan the whole display and visually indicate how the contents of different display areas are related (i.e., similar or distinct or opposite, consistent or inconsistent) to the contents of an AOI specified by the user.

Another frequently occurring behavioural pattern was searching for data patterns that were similar to a particular pattern. Identifying such patterns could be supported by automated search. A visual analytics system could also help

users to unite several similar local patterns into regional, linear, or distributed patterns. Another possibility is to enable users to organise patterns in groups or classes based on their similarity or another kind of relationship (e.g., consistency between the patterns or similarity of their contexts).

The participants of our study often attempted to verify their current mental models or hypotheses by purposefully searching for contradicting data patterns. It would be good to facilitate this searching by means of automated techniques. While it may be difficult for an analyst to specify what hypothesis she/he aims to verify, the analyst can select a pattern that is representative of this hypothesis and employ automated facilities to search for contrasting and inconsistent patterns.

We have discovered two patterns of attention distribution: selective focusing and comprehensive scanning. Since the selective focusing behaviour can potentially lead to biases, it could be appropriate to detect situations when analysts tend to neglect some parts of the display and somehow warn them or attract their attention to unexplored data.

The activities on linking data patterns to context and to analyst' prior and earlier gained knowledge can be supported by tools for annotation and for creation of knowledge graphs. However, performing such interactive operations requires additional time and effort of analysts, which may hinder the adoption of the tools. To help analysts, an intelligent system could automatically generate draft annotations. The pattern theory [AAM*21] can be used to predict the possible types of patterns from the types of data components and known basic relationships between their elements. The actual type of a given pattern (e.g., interactively marked by a user) and its synoptic characteristics can be derived through automatic analysis of the specific between-element relationships within the pattern. When a display includes two or more information layers, i.e., shows distributions of several data components over a common base, the system could also establish the relationships of the pattern to data from other layers that are co-located with or close to the pattern in the display. All this automatically derived information can be used to generate a draft annotation of the pattern, which may need to be only slightly edited by the analyst.

Certainly, these or other ideas concerning possible support to analysts need to be checked by further research and experimentation.

7.3. Assessment of the capabilities of the analysis methods used

For analysing the eye-tracking data, we modified standard techniques (namely, scarf plots gaze flow map, and transition graph) to adapt them to the settings and goals of our study. As stated in Section 5.1, an essential feature of our experiments is the absence of standard AOIs. Therefore, we modified the scarf plot technique so that it could be used

without defining AOIs, we extracted individual AOIs of the participants and used them to create gaze flow maps, and we generated comparable transition graphs by establishing equivalence relationships between the specific AOIs from distinct sessions based on the relative frequencies of visiting the AOIs.

We used these techniques in combination due to their complementary capabilities. As Section 5.1 explains, we strove for generic findings and therefore needed to use approaches supporting abstraction. The modified technique of scarf plots allowed us to abstract from particular positions in the display and consider only the manner in which the attention focus changed. The gaze flow maps helped us to abstract from the temporal sequences and reveal spatial configuration patterns of individual AOIs and transitions between them (the term “spatial” refers here to the display space). With the transition graphs, we were able to abstract from the spatial configurations and focus on the relative frequencies of visiting the AOIs and the patterns of links between them. With each method, we obtained a different kind of valuable findings, as described in Section 6.1. We can conclude that the combination of the methods we used was appropriate for our eye-tracking data and analysis tasks.

A very important part of our study was the joint analysis of the “think-aloud” records and the synchronous eye movements. This was done using the session replay function of the standard software provided with the eye tracker. While this work was laborious and requiring high concentration, it was paramount for understanding participants’ analytical reasoning and for interpreting the findings obtained in the analysis of the eye tracking data. On the other hand, the think-aloud protocol alone would not be sufficient for achieving our goals because the participants do not verbalise all their analytical activities. The eye tracking data can reveal what remained unspoken.

We realise that the necessity to talk during the analysis may distract the participants from the analysis itself and that the process of verbalising thoughts may interfere with the processes of perception, interpretation, and reasoning. We designed and conducted our experiments so as to preserve as much as possible the natural behaviour of the participants. This was achieved by creating comfortable conditions for the participants and giving them appropriate instructions as recommended by Holmqvist et al. [HNA*11]. Of course, this cannot guarantee that the participants always gaze at a relevant part of the display while talking. Therefore, it is very important to review sessions observing whether gaze positions during utterances are consistent with the gaze movements preceding the utterances. Attention should also be given to the duration of the utterances and the ratio of the talking time to the time of silent exploration of an image. If it is noticed that talking takes relatively long time and that the gaze position during talking can be arbitrary, it is reasonable to cut the episodes of talking out of the eye track-

ing records before analysing the latter. In our experiments, we did not see a need for such separation. Firstly, the oral statements of the participants always came after periods of silent exploration, which were long enough to provide reliable data about the attention of the participants. Secondly, we did not see significant deviations of the gaze directions during talking from the directions during the preceding exploration. Occasional short-term distractions do not alter the general patterns. Since this may be different in other experiments, adjustments, such as separation of talking from exploration, may be needed to compensate for the possible distortions of eye movement patterns.

8. Conclusion

Visual analytics science is experiencing a deficiency in empirical research on how people do analysis and solve problems using visual representations of information. The research is hindered by multiple objective difficulties: the necessity to use sufficiently complex data and analysis tasks, the need to design experiments so that sufficiently general findings could be gained, the necessity to find commonalities across participants and analysis scenarios while respecting inevitable differences, and the difficulty of validation of findings obtained. We believe that our work described in this paper can provide examples of viable approaches to meeting these challenges, and that new empirical studies will be conducted in near future for strengthening and orienting the visual analytics research.

One of merits of our work is that we linked the research questions and empirical results to existing theoretical and conceptual models describing the process of data analysis and knowledge generation. We have seen that the process involves finding patterns [AAM*21, SSS*14], interpretation of the patterns using analysis context and domain knowledge [KHL21], and repeated evaluation of the current state of the mental model [ALA*18], including current hypotheses [SSS*14], in relation to the patterns observed. The possibility to establish such links allows us to conclude this paper with a kind of “I have a dream” statement.

We imagine an intelligent system as a partner of a human analyst. The two partners conduct the analysis in a discursive manner. The human “thinks aloud” about what is seen, and the machine listens, understands, and complements these thoughts by relevant information that can be computed and inferences that can be made automatically. The machine understands the human well because it (1) can predict what types of patterns can exist in the data under analysis, (2) tracks where the human looks and can determine what data patterns may be there, and (3) knows the terms and expressions in which the human can speak about data patterns that are actually observed and that are expected. The machine applies statistical techniques to check human’s observations, searches through the whole dataset for occurrences of consistent and contrasting patterns, and helps the human to syn-

thesise the collected facts and make general conclusions. The machine also watches the distribution of the human's attention and tries to prevent possible biases by telling the human how the data that have not been yet sufficiently explored relate to what is already known. The machine records and maintains the provenance of all findings and helps the human to organise the results of the analysis and make a final presentation. In this mode of work, the human analyst does not need to spend time for unnatural interactive operations, such as moving the mouse or pushing buttons, but can effectively use the time for seeing and thinking. We believe that current hardware and software technologies are already sufficient for implementing such a dream. What is still missing is sufficient empirical material that can underpin the theoretical development as well as development and training of intelligent software systems.

Acknowledgments

This work was partially funded by Fraunhofer Center for Machine Learning within the Fraunhofer Cluster for Cognitive Internet Technologies, by EU in project SoBigData++ and by German Science Foundation (DFG) in priority research program SPP 1894 on VGI.

References

- [AA06] ANDRIENKO N., ANDRIENKO G.: *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag, Berlin, Heidelberg, 2006. 2
- [AA11] ANDRIENKO N., ANDRIENKO G.: Spatial generalization and aggregation of massive movement data. *IEEE Transactions on Visualization and Computer Graphics* 17, 02 (feb 2011), 205–219. doi:10.1109/TVCG.2010.44. 11
- [AABW12] ANDRIENKO G., ANDRIENKO N., BURCH M., WEISKOPF D.: Visual analytics methodology for eye movement studies. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2889–2898. doi:10.1109/TVCG.2012.276. 5, 11
- [AAFJ16] ANDRIENKO N., ANDRIENKO G., FUCHS G., JANKOWSKI P.: Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Information Visualization* 15, 2 (2016), 117–153. doi:10.1177/1473871615581216. 11
- [AAM*21] ANDRIENKO N., ANDRIENKO G., MIKSCH S., SCHUMANN H., WROBEL S.: A theoretical model for pattern discovery in visual analytics. *Visual Informatics* 5, 1 (2021), 23–42. doi:10.1016/j.visinf.2020.12.002. 2, 3, 20, 21, 22, 28
- [ALA*18] ANDRIENKO N., LAMMARSCH T., ANDRIENKO G., FUCHS G., KEIM D., MIKSCH S., RIND A.: Viewing visual analytics as model building. *Computer Graphics Forum* 37, 6 (2018), 275–299. doi:10.1111/cgf.13324. 2, 3, 17, 18, 20, 22
- [ARBC*19] ABDUL-RAHMAN A., BORGO R., CHEN M., EDWARDS D. J., FISHER B.: Juxtaposing controlled empirical studies in visualization with topic developments in psychology, 2019. arXiv:1909.03786. 4
- [BAA*13] BURCH M., ANDRIENKO G., ANDRIENKO N., HÖFERLIN M., RASCHKE M., WEISKOPF D.: Visual task solution strategies in tree diagrams. In *2013 IEEE Pacific Visualization Symposium (PacificVis)* (2013), pp. 169–176. doi:10.1109/PacificVis.2013.6596142. 4
- [Ber83] BERTIN J.: *Semiology of Graphics*. University of Wisconsin Press, 1983. 1
- [BKR*16] BLASCHECK T., KURZHALS K., RASCHKE M., STROHMAIER S., WEISKOPF D., ERTL T.: AoI hierarchies for visual exploration of fixation sequences. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications* (New York, NY, USA, 2016), ETRA '16, Association for Computing Machinery, p. 111–118. doi:10.1145/2857491.2857524. 5
- [BKR*17] BLASCHECK T., KURZHALS K., RASCHKE M., BURCH M., WEISKOPF D., ERTL T.: Visualization of eye tracking data: A taxonomy and survey. *Computer Graphics Forum* 36, 8 (2017), 260–284. doi:10.1111/cgf.13079. 5
- [BSBE17] BLASCHECK T., SCHWEIZER M., BECK F., ERTL T.: Visual comparison of eye movement patterns. *Computer Graphics Forum* 36, 3 (2017), 87–97. doi:10.1111/cgf.13170. 5, 11
- [CAA*21] CHEN S., ANDRIENKO N., ANDRIENKO G., LI J., YUAN X.: Co-bridges: Pair-wise visual connection and comparison for multi-item data streams. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1612–1622. doi:10.1109/TVCG.2020.3030411. 8
- [CAS*18] COLLINS C., ANDRIENKO N., SCHRECK T., YANG J., CHOO J., ENGELKE U., JENA A., DWYER T.: Guidance in the human—machine analytics process. *Visual Informatics* 2, 3 (2018), 166–180. doi:10.1016/j.visinf.2018.09.003. 3
- [DAREA*18] DIEHL A., ABDUL-RAHMAN A., EL-ASSADY M., BACH B., KEIM D. A., CHEN M.: VisGuides: A forum for discussing visualization guidelines. In *EuroVis (Short Papers)* (2018), pp. 61–65. 1
- [Duc07] DUCHOWSKI A. T.: *Eye tracking methodology: Theory and practice*. Springer, 2007. 4
- [ES93] ERICSSON K. A., SIMON H. A.: *Protocol Analysis: Verbal Reports as Data*. The MIT Press, 04 1993. doi:10.7551/mitpress/5657.001.0001. 5
- [FLMB11] FOLLET B., LE MEUR O., BACCINO T.: New insights into ambient and focal visual fixations using an automatic classification algorithm. *i-Perception* 2, 6 (2011), 592–610. 4, 5
- [FRHA*08] FABRIKANT S. I., REBICH-HESPANHA S., ANDRIENKO N., ANDRIENKO G., MONTELLO D. R.: Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes. *The Cartographic Journal* 45, 3 (2008), 201–215. doi:10.1179/000870408X311396. 4
- [Heg11] HEGARTY M.: The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science* 3, 3 (2011), 446–474. doi:10.1111/j.1756-8765.2011.01150.x. 3
- [HNA*11] HOLMQVIST K., NYSTRÖM M., ANDERSSON R., DEWHURST R., JARODZKA H., VAN DE WEIJER J.: *Eye-tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2011. 5, 8, 22
- [IIC*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMIR M., MÖLLER T.: A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2818–2827. doi:10.1109/TVCG.2013.126. 4

- [ITC08] ISENBERG P., TANG A., CARPENDALE S.: An exploratory study of visual information analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, Association for Computing Machinery, p. 1217–1226. doi:10.1145/1357054.1357245. 4
- [KAF*08] KEIM D., ANDRIENKO G., FEKETE J.-D., GÖRG C., KOHLHAMMER J., MELANÇON G.: Visual analytics: Definition, process, and challenges. In *Information visualization*. Springer, 2008, pp. 154–175. 1
- [KGR13] KIEFER P., GIANNOPOULOS I., RAUBAL M.: Using eye movements to recognize activities on cartographic maps. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2013), SIGSPATIAL'13, Association for Computing Machinery, p. 488–491. doi:10.1145/2525314.2525467. 4
- [KGRD17] KIEFER P., GIANNOPOULOS I., RAUBAL M., DUCHOWSKI A.: Eye tracking for spatial research: Cognition, computation, challenges. *Spatial Cognition & Computation* 17, 1-2 (2017), 1–19. doi:10.1080/13875868.2016.1254634. 4
- [KHL21] KARER B., HAGEN H., LEHMANN D. J.: Insight beyond numbers: The impact of qualitative factors on visual data analysis. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1011–1021. doi:10.1109/TVCG.2020.3030376. 2, 3, 19, 20, 22
- [KHW14] KURZHALS K., HEIMERL F., WEISKOPF D.: ISeeCube: Visual analysis of gaze data for video. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (New York, NY, USA, 2014), ETRA '14, Association for Computing Machinery, p. 43–50. doi:10.1145/2578153.2578158. 5
- [KKGH21] KIM Y. S., KAYONGO P., GRUNDE-MCLAUGHLIN M., HULLMAN J.: Bayesian-assisted inference from visualized data. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 989–999. doi:10.1109/TVCG.2020.3028984. 3, 20
- [KO10] KRAAK M., ORMELING F.: *Cartography: Visualization of spatial data*, 3 ed. Guilford Publications, nov 2010. doi:10.1201/9780429464195. 11
- [KPS*17] KODAGODA N., PONTIS S., SIMMIE D., ATTFIELD S., WONG B., BLANDFORD A., HANKIN C.: Using machine learning to infer reasoning provenance from user interaction log data. *Journal of Cognitive Engineering and Decision Making* 11 (03 2017). 4
- [KW13] KURZHALS K., WEISKOPF D.: Space-time visual analytics of eye-tracking data for dynamic stimuli. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2129–2138. doi:10.1109/TVCG.2013.194. 5
- [KWKH19] KIM Y.-S., WALLS L. A., KRAFFT P., HULLMAN J.: A bayesian cognition approach to improve data visualization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, p. 1–14. doi:10.1145/3290605.3300912. 2, 3, 20
- [LBI*12] LAM H., BERTINI E., ISENBERG P., PLAISANT C., CARPENDALE S.: Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics* 18, 9 (2012), 1520–1536. doi:10.1109/TVCG.2011.279. 4
- [Mac86] MACKINLAY J.: Automating the design of graphical presentations of relational information. *ACM Trans. Graph.* 5, 2 (Apr. 1986), 110–141. doi:10.1145/22949.22950. 1
- [Met06] METZGER W.: *Laws of Seeing*. MIT Press, Cambridge, MA, USA, 2006. 28
- [OGF14] OPACH T., GOLEBIOWSKA I., FABRIKANT S. I.: How do people view multi-component animated maps? *The Cartographic Journal* 51, 4 (2014), 330–342. doi:10.1179/1743277413Y.0000000049. 4
- [Pad18] PADILLA L. M.: A case for cognitive models in visualization research : Position paper. In *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)* (2018), pp. 69–77. doi:10.1109/BELIV.2018.8634267. 2
- [PB13] POPELKA S., BRYCHTOVA A.: Eye-tracking study on different perception of 2d and 3d terrain visualisation. *The Cartographic Journal* 50, 3 (2013), 240–246. doi:10.1179/1743277413Y.0000000058. 4
- [PBG*14] PATTERSON R. E., BLAHA L. M., GRINSTEIN G. G., LIGGETT K. K., KAVENEY D. E., SHELDON K. C., HAVIG P. R., MOORE J. A.: A human cognition framework for information visualization. *Computers & Graphics* 42 (2014), 42–58. doi:https://doi.org/10.1016/j.cag.2014.03.002. 2, 19, 20
- [PCRHS18] PADILLA L., CREEM-REGEHR S., HEGARTY M., STEFANUCCI J.: Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications* 3 (12 2018). doi:10.1186/s41235-018-0120-9. 2, 3, 19
- [PF90] PINKER S., FEEDLE R.: A theory of graph comprehension. In *Artificial Intelligence and the Future of Testing*, Freedle R., (Ed.). Lawrence Erlbaum Associates, Inc., 1990, pp. 73–126. 2, 4
- [PWK16] POHL M., WALLNER G., KRIGLSTEIN S.: Using lag-sequential analysis for understanding interaction sequences in visualizations. *International Journal of Human-Computer Studies* 96 (2016), 54–66. doi:https://doi.org/10.1016/j.ijhcs.2016.07.006. 4
- [RD05] RICHARDSON D. C., DALE R.: Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science* 29, 6 (2005), 1045–1060. doi:10.1207/s15516709cog0000/_29.5. 10
- [SBJ*19] SILVA N., BLASCHECK T., JIANU R., RODRIGUES N., WEISKOPF D., RAUBAL M., SCHRECK T.: Eye tracking support for visual analytics systems: foundations, current applications, and research challenges. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (2019), pp. 1–10. doi:10.1145/3314111.3319919. 4
- [SMKH09] SLOCUM T. A., MACMASTER R. B., KESSLER F. C., HOWARD H. H.: *Thematic cartography and geovisualization, 3rd Edition*. Pearson Education, 2009. 11
- [SSS*14] SACHA D., STOFFEL A., STOFFEL F., KWON B. C., ELLIS G., KEIM D. A.: Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1604–1613. doi:10.1109/TVCG.2014.2346481. 2, 3, 20, 22
- [TC05] THOMAS J. J., COOK K. A.: *Illuminating the path: The research and development agenda for visual analytics*. IEEE, 2005. 1
- [Tob87] TOBLER W.: Experiments in migration mapping by computer. *Cartography and Geographic Information Science* 14 (04 1987), 155–163. doi:10.1559/152304087783875273. 11
- [WEK*12] WAGEMANS J., ELDER J. H., KUBOVY M.,

PALMER S. E., PETERSON M. A., SINGH M., VON DER HEYDT R.: A century of gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological bulletin* 138, 6 (2012), 1172–1217. doi:10.1037/a0029333. 28

Appendix A: Placements of pattern markers

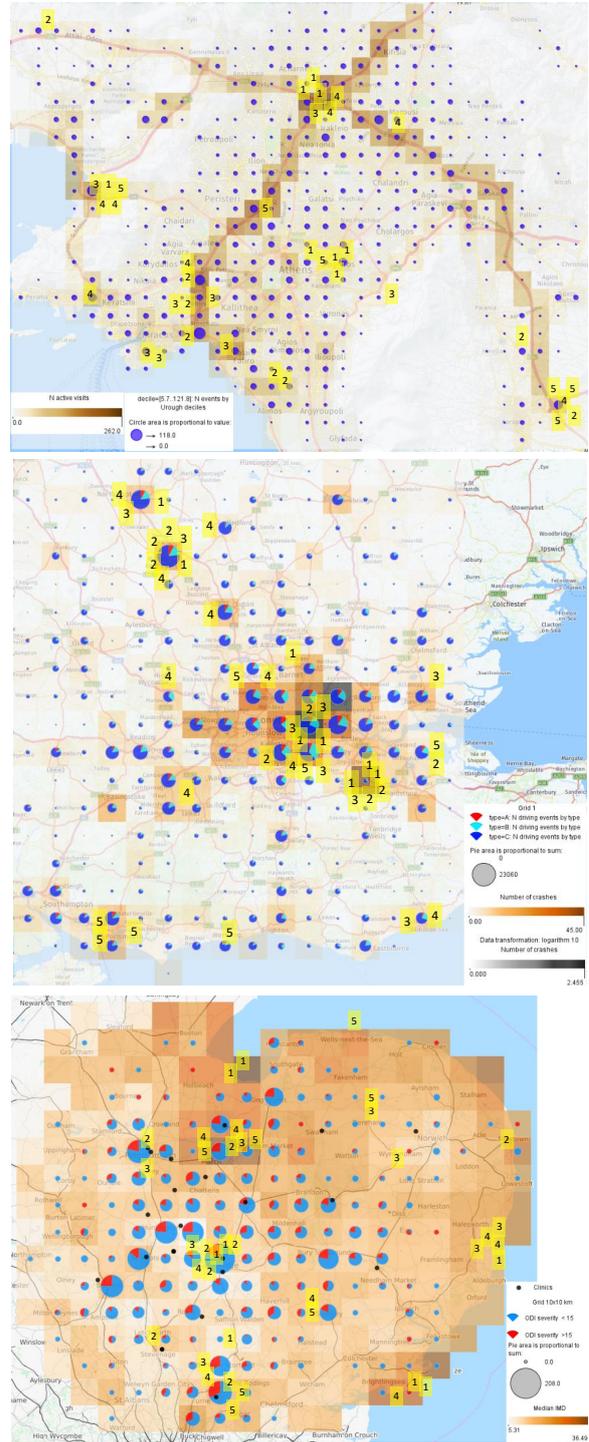


Figure 13: Placements of the pattern markers by all participants in the three scenarios.

Figure 13 shows the compiled placements of the pattern markers by all participants in the three analysis scenarios. The numeric labels from 1 to 5 correspond to the order of the placements made by each participant. They are not meant to indicate the relative importance of the patterns.

Appendix B: Interpretation of session records

Note: Here we reproduce the contents and structure of the session analysis document, in which the interpretations of the session records were systematically organised. The original document was created using another word processor and cannot be attached in its original form.

Scales (spatial extents) of patterns or relationships

- Global, e.g., “Large N of events mostly corresponds to high traffic”.
 - Positive correlation
 - Negative correlation
 - Absence of correlation
- Local: small spots, sometimes singular cells of a grid
- Intermediate
 - Regional: more or less extended areas
 - Linear, e.g., along roads
- Distributed: multiple disjoint places with similar characteristics (i.e., a re-occurring local pattern).

Types of relationships people notice

- Between two phenomena that are visualised
- Between one phenomenon or both and the geographic background

Characteristics of local and intermediate patterns

. What motivates people to consider them as patterns?

- Relationship to a global (observed or expected) pattern/relationship
 - Prominent expression of the global pattern, e.g., two phenomena have very high values
 - Deviation from the global relationship, e.g., “N events is quite high while the traffic is not so high” – weakened association or absence of expected association
 - Disruption of the global relationship, i.e., the local relationship is a kind of opposite to the global one
- High difference from the surrounding context, e.g., values are notably higher or lower than around
- Association with particular features in the geographic background, e.g., road or crossing

Attention scope

- Local (small place and its immediate surrounding)
- Intermediate (region)
- Overall (whole map)

Eye movement patterns

- Scanning: short fixations on locations in different parts of a map
 - Overall scanning (whole map)
 - Regional scanning (fixations are mostly contained within some area)
 - Scanning of a feature in the geographic background, particularly, tracing a linear feature, e.g., a road
- Observation of a global pattern: similar to scanning; requires verbal data to disambiguate.
- Assessing (e.g., assess potential interestingness of a location): relatively short fixation on a location, multiple quick looks at other locations
- Focus place: long fixation or a sequence of fixations on close locations
- Focus region: “dancing” over a region, i.e., short fixations within and around the region; locations of consecutive fixations may not be that close.
- Extension of the focus area: fixations on locations around a location that have been in focus before, not far from the original focus location.
- Contraction of the exploration space: decreasing the area of scanning.
- Comparison of a place in focus with one or more other places: a move from the focus place to another place, short fixation on that place, and return to the focus place; possibly, repeated several times with the same or different places looked at.
- Comparison of two places: moving back and forth between two places with not very long fixations in each of them.
- Comparison of multiple places: repeated visits and movements between the places, approximately equal fixation durations (may signify detection of a common pattern).
- Revisiting of a previously marked pattern (e.g., to compare a new observation with one of previously made)

Behaviour pattern

A common behavioural pattern is that people continue looking at the area where they found a pattern after they have marked it. Possibly, they are checking and justifying their decision.

Types of visual search targets

- Visual contrast (visual properties of a location significantly differ from the neighbourhood)

- Areas with low internal variation (spatial clusters of locations with similar visual properties)
- Features in the geographic background, e.g., roads
- Particular visual properties, e.g., light or dark background shading and large or small symbols
- Places similar to a particular place

Cognitive tasks

- Browse
 - Assess place interestingness
 - Find an interesting place to focus on
 - Select a place to focus from two or more candidate places
- Search
 - Find positive examples of an expected pattern/relationship
 - Find counter-examples or deviations from an expected pattern/relationship
 - Find places similar to a place previously explored
 - Find places differing from a place previously explored
- Examine and interpret
 - Examine properties of a place (i.e., values of the attributes visualised)
 - Assess the properties w.r.t. the expected relationship (i.e., whether it holds or not)
 - Relate the properties to geographical background, e.g., a road or a populated (or unpopulated) area
 - Relate the properties to background knowledge, e.g., how cars may behave on a road crossing
- Compare
 - a place to one or more previously explored places
 - whether the properties are similar or different
 - whether a relationship is expressed more or less prominently
 - a place to the surrounding context (other places in the neighbourhood)
 - several places
 - check if there is a common pattern
 - select a place to focus on from several candidates
- Group
 - Find a region where all or most of the places have similar properties (regional-scale pattern)
 - Find multiple instances of the same pattern
 - Find out how multiple places with similar properties are arranged in the space (e.g., located along a road)
 - Combine multiple disjoint instances into a global pattern

Exploration tasks

- Overall scale
 - Perceive and understand the distribution of a single phenomenon.
 - Determine the presence or absence of a global relationship, assess the strength of the relationship if present. It is usually done by checking if a certain expected relationship or one of several possible relationships holds.
 - Evaluate different places and areas w.r.t. the relationship in question.
 - Look for negative instances (no expected value association, or an opposite association).
 - Judge if the global relationship exists and, if so, how strong it is.
- Local scale: Explore and compare particular places.
 - Determine whether a certain relationship holds or not; assess the strength of the relationship.
 - Assess the similarity or difference of the place to/from the neighbourhood.
 - Compare two or more places w.r.t.
 - Similarity
 - Strength of a relationship
 - Interestingness (e.g., which is more unusual or demonstrates a relationship more prominently)
- Intermediate scale: Unite multiple places.
 - Find regions consisting of similar places.
 - Find specific spatial arrangements of similar places, such as alignments along roads.
 - Find subsets of disjoint similar places, perceive and interpret their spatial distribution.

Strategies

- Assume an expected global relationship, find representative examples, then look for counter-examples
- Identify or assume a global relationship, look for deviations and disruptions
- Search for places with particular visual properties, such as
 - big symbols on light background
 - small symbols on dark background
 - significant difference from the surrounding visual context
- Find multiple local patterns, then search for possible “clusters” of similar places
- After finding a pattern, try to find similar ones

Biases and tendencies

Pattern scale

- Tendency to focus on local, small-scale patterns (particular places)

- Striving to find regional patterns
- Striving to find several places with similar local patterns.

Participants can be classified into “localists” and “regionalists”. “Regionalists” may also detect local patterns, but they tend to strive at finding repeated local patterns.

Data representation

- More attention to darker background
- More attention to larger symbols
- Looking for visual contrast with the surrounding, e.g., large symbols in areas with mostly small ones

Pattern diversity

- Tendency to note patterns of the same kind (e.g., high traffic and many events)
- Striving to find diverse kinds of patterns

Appendix C: Generalisability of the findings concerning patterns

A large part of our findings presented in Section 6 refers to the processes of detecting and interpreting visual patterns. Our studies were done using maps as visual stimuli to analyse. Moreover, the maps were created using the same or similar visualisation techniques. It may be thought that the findings concerning the types, structures, interpretation, and utilisation of the patterns are very specific if not for these visualisation techniques then for maps. However, we believe that this is not the case and can give the following counter-arguments.

First, although particular visualisation techniques were used to represent combinations of values of two or more attributes, participants described their observations and reasoning in terms of the data represented rather than in terms of the visual encoding of the data. They successfully decoded the visual encoding, and this can be done with any visualisation obeying the basic principles of visual representation and providing viewers a legend to enable decoding. Second, although we did our experiments using maps presenting spatial data, only the findings that refer to the spatial (geographic) context can be considered as specific for maps and spatial data. All other findings, in fact, do not refer to geography but refer to the space of the display. Therefore, they can apply not only to maps but also to other kinds of visualisation where some visual elements representing data items are placed in two-dimensional display space in which distances and neighbourhood relationships are meaningful. Examples are scatterplots and projection plots built using dimensionality reduction methods. We deem that similar findings might also be obtained in studies with matrix displays where rows and columns are ordered so that the distances between the cells correspond to some analysis-relevant relationships between the data items represented in the cells.

In the types of patterns found by the study participants,

we can see effects of the general Gestalt laws of visual perception [Met06, WEK*12]. Thus, seeing multiple grid cells jointly as regions correspond to the Law of Closure, finding clusters to the Law of Proximity, finding alignments to the Law of Continuity, and finding distributed patterns to the Law of Similarity. The correspondence to the general principles provides additional evidence of high generality of our findings.

Our analysis showed that the types of data patterns discovered by participants could be predicted using the definitions of the pattern theory [AAM*21]. We believe that this can also be done for experiments with other types of data and other types of display that appropriately represent relationships between elements of data components. We also think that participants in other empirical studies of the processes of visual data analysis may perform the same types of operations on data patterns they have discovered as the participants in our study.