Big Data for Mobility Tracking Knowledge Extraction in Urban Areas

D5.4 Results of eye tracking evaluation

Document Summary Information

Grant Agreement No	780754	Acronym	TRACK & KNOW
Full Title	Big Data for Mobility Tracking Knowledge Extraction in Urban Areas		
Start Date	01/01/2018	Duration	36 months
Project URL	trackandknowproject.eu		
Deliverable	D5.4		
Work Package	WP5		
Contractual due date	M36	Actual submission date	...
Nature	Report	Dissemination Level	Public
Lead Beneficiary	FRHF		
Responsible Author	Gennady Andrienko (FRHF)		
Contributions from	Natalia Andrienko (FRHF), Mirco Nanni (CNR), Omid Isfahanialamdari (CNR), Athanasios Koumparos (VFI), Leonardo Longhi (SIS), Toni Staykova (PAP), Kieran Lee (PAP), Ian Smith (PAP)		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 780754.

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete[1]	Changes	Contributor(s)
V0.1	15.09.20	1%	Initialization / TOC	FRHF
V0.2	05.11.20	15%	Introduction, State of the art	FRHF
V0.3	13.11.20	30%	Conceptualization	FRHF
V0.4	20.11.20	50%	Description of experiments	FRHF
V0.5	03.12.20	70%	Description of analysis	FRHF
V0.6	03.12.20	90%	Analysis of IMNs	CNR
V1.0	07.12.20	100%	Final integrated document	FRHF, CNR

[1] According to TRACK&KNOW's Quality Assurance Process:
Due Date - 6 Weeks: Peer Review
Due Date – 2 Weeks: Quality Manager Review (INTRASOFT)
Due Date - 2 Days: Sent to Inlecom for Submission to the EC after addressing all comments by Quality Manager and Peer Reviewers

Table of Contents

List of Figures

Glossary of terms and abbreviations used

Abbreviation / Term	Description
VA	Visual Analytics
AOI	Area of Interest
UI	User Interface
CNR	Consiglio Nazionale Delle Ricerche
SIS	SISTEMATICA
PAP	NHS Royal Papworth Hospital
NHS	United Kingdom National Health Service
FRHF	Fraunhofer Gesellschaft zur Foerderung der Angewandten Forschung EV
IMN	Individual Mobility Network

1 Introduction

Eye tracking is a process of measuring and recording individual's gaze positions and eye movements. This technology is being increasingly used not only in psychology and product design but also in visualization and human-computer interaction sciences for evaluation of visual displays and user interfaces (UI). Researchers employ eye tracking to understand how their designs are actually used and, possibly, even get insights into users' ways of reasoning and problem solving [1]. In evaluating one design, researchers want to check if users' behaviours correspond to the supposed ways of use, see where and in which situations the users may have difficulties, and understand how the design can be improved. In evaluating two or more alternative designs, researchers want to know not only which design is better in terms of task completion times and error rates but also why it is better: How does the use of this design differ from the use of the others? What is more difficult, confusing, or inconvenient in the other designs?

The eye tracking experiments in Track&Know had quite a different aim than in most of the previous studies. Instead of evaluating particular visualisation and/or UI designs, we wanted to reveal more general patterns of analysts' behaviour in the process of using visualisations of non-trivial data for gaining understanding of the phenomena standing behind the data.

The visual analysis of data collected in eye tracking studies (referred to as "eye-tracking data" in the following text) has become an emerging field of research leading to many new visualization techniques in recent years. These techniques provide insight beyond what is facilitated by traditional attention maps and gaze plots, providing important means to support statistical analysis and hypothesis building. In Track&Know, our secondary aim was to find effective approaches to analysing eye-tracking data that can allow us and, potentially, other researchers to achieve the principal aim of understanding analysts' behaviours and general strategies in visual data analysis.

1.1 Purpose and scope

As already mentioned, the purpose of eye tracking experiments in Track&Know was to gain general knowledge about analysts' behaviours and strategies in the process of visual analysis of non-trivial data. This purpose corresponds to the aims of the project as a whole, namely, development of novel problem solving methods rather than implementation of ready-to-use software to be immediately applied in industrial settings. Like in the entire project, the intended result of eye tracking experiments has been generally valuable new knowledge and new methods and/or analytical workflows that help extract this kind of knowledge from data.

1.2 Approach for the Work package and relation to other Deliverables

The underlying idea of the approach that was used can be stated as follows:

- Take the essence of the analytical visual representations that are developed within the 3 pilots of the project, as presented in deliverables D6.2 – D6.4, and generalize them to a limited set of archetypal visual representations.
- Create instances of these archetypal representations and use them as visual stimuli in experiments in which domain experts and/or data scientists analyse data with the use of these representations.
- Observe, record, and study the processes of data analysis in order to extract general behavioural patterns of the analysts and the approaches and strategies they employ in doing visual exploration.

As previously mentioned, the secondary goal has been to find suitable methods and defining workflows for analysis of collected eye-tracking data. Taking into account that eye-tracking data can be considered as a special case of movement data, it was logical to combine standard analysis methods focussed on fixations in combination with methods for visual and computational analysis of movement data, including methods described in deliverable D5.1 and computational methods of movement analysis developed in WP4.

1.3 Mapping Track and Know outputs

In this document, we present

1. State of the art in visually driven analysis of eye-tracking data.
2. A description of problem solving experiments that were developed based on the 3 pilots of the project and performed with 8 professionals.
3. Analysis of the collected data using visual analytics approaches.
4. Analysis of the collected data using computational methods for movement data analysis.

This deliverable is based of deliverables D5.1 (state of the art in movement data analysis), D5.3 (visual analytics dashboard) and D6.2-D6.4 (pilots 1-3, respectively), as well as on multiple deliverables of WP4. This deliverable is the main output of task T5.4 entitled “Evaluating visual analytics procedures through eye tracking”.

1.4 Delivery plan and deviations

The task description is the following:

VA methods developed in tasks T5.1-T5.4 will be evaluated in corresponding usage scenarios with appropriate categories of professional users. A part of evaluation will be done with computer science specialists focussed on data processing and analysis, another part will be performed with domain experts from use cases. The evaluation will address objective user performance (how fast/correct are users in problem solving) and subjective user satisfaction. For selected critical tasks, eye tracking evaluation of problem solving activity [13] will be performed for identifying common mistakes and suboptimal problem solving strategies. The result of the task will be documented in deliverable D5.4

A preparatory work in task T5.4 started during the implementation of the dashboard (Task T5.3) for the purposes of pilots 1-3 (Tasks T6.2-T6.4). A series of consultations with project partners was performed. In the result of these consultations, a set of representative tasks on analysing relationships between two or more phenomena co-existing in the geographic space was identified, and most appropriate visual representations that are able to support data analysis were selected. These representations generalize the dashboard visualisation designs that were implemented for the three pilots of the project.

Based on the decisions made during the preparatory stage, a plan for eye-tracking data collection was developed. The plan was presented at the consortium meeting in Paris (February 2020) and approved by the consortium. The key principles of the plan were:

- Test participants: application partners from the 3 pilots.
- For each pilot, we prepare a separate map using specific data of this pilot.
- We assume that the partners know the corresponding territories, phenomena, and relevant analysis tasks.

Respectively, the following schedule was approved at the consortium meeting:

- Selection of data and preparation of maps: March – April 2020. Responsible: FhG + application partners.
- Preliminary tests with FhG employees: April – May 2020. Responsible: FhG.

- We plan to use collected data for a workshop and/or poster presentation (a potentially suitable venue is ICA Workshop on Analytical Reasoning @ GIScience 2020 in Poznan).
- Main experiments: June 2020, during the consortium meeting at FhG.
- Analysis of collected data: June – September 2020.

While the first step of the schedule has been successfully performed according to the plan (see details in Section 3), the unexpected situation caused by the COVID'19 pandemics forced us to make substantial modifications to the remaining steps. During April-May 2020 all employees of Fraunhofer IAIS worked in home office settings, with serious restrictions on personal contacts. Respectively, the preliminary tests had been delayed until the situation would allow physical meetings. Moreover, it was decided that the Track&Know consortium meeting in June would be conducted as an online meeting, thus making eye-tracking data collection at the meeting absolutely impossible. All international business trips were prohibited.

Due to these unforeseen circumstances, we had to make the following modifications of the plan. It was decided to perform evaluation experiments with a small group of data analysis and visualization professionals who volunteer to take a risk of a personal meeting with the experimenters, which is essential for eye tracking experiments. To compensate for their lacking knowledge of the particular application domains, we explained them in detail the specifics of each pilot and corresponding data sets, analysis tasks, possible patterns that could be observed in the data, and visual representations. This change of the plan was approved at the online meeting of the consortium and agreed with the Project Officer in June 2020.

The experiments and data collection was performed during summer 2020, followed by analysis of the collected data, generalizing the findings, and writing this report. The analysis has been performed by two teams: FRHF team applied visual analytics approaches from WP5, and CNR used machine learning methods from WP4. We plan to write a follow-up research paper based on this study.

1.5 Methodology and structure of the deliverable

The project Track&Know in general focuses on analysing phenomena existing end developing in geographic space. Accordingly, all 3 application pilots of the project deal with geospatial phenomena. The use of maps is essential for studying and understanding geospatial phenomena by analysing geographically referenced data. On a map, data are shown in their spatial context, and a map enables perceiving spatial relationships between data items. Therefore, it was obvious that the processes of visual analysis of project-relevant types of data and phenomena need to be studied using maps as visual stimuli.

The next consideration taken into account was that amounts of data that need to be analysed in real applications are very large, which precludes visualisation of each individual data item and requires the use of data aggregation. In particular, spatial data are usually aggregated by spatial compartments. This is how geospatial data are represented on maps included in the dashboards. Therefore, it was logical to use maps representing aggregated spatial data for experiments involving eye tracking.

Concerning the map content, it was decided that each map should represent two or more phenomena distributed over the same trajectory. The rationale for this decision was that, generally, non-trivial analysis of geospatial data in real-world applications requires consideration of several phenomena co-existing on the same territory. A typical analysis task is to understand whether and how these phenomena are interrelated. This general task is relevant to each pilot in particular. For each individual pilot, it may have specific formulation referring to specific phenomena considered in this pilot and specific kinds of relationships that may be of interest.

Moreover, the general task may be instantiated in several pilot-specific tasks targeting different aspects of the phenomena or different kinds of relationships.

To summarise, the decision was to perform experiments with the use of maps representing aggregated data describing two or more pilot-specific phenomena.

Our ambition was to perform experiments so that results could be valid and valuable beyond a particular implementation in the Track&Know system, thus providing knowledge and guidance for future products (e.g. implementations in Track&Know components in industrial settings) and further research. In accord with these goals and wishes, we have constructed a set of analysis tasks and corresponding visual representations in close collaboration with domain experts from all 3 pilots.

We recruited a group of professionals skilled in data analysis and/or visualisation and asked them to use these visual representations for performing data analysis. The eye tracking technology was employed for recording the process of visual exploration of the maps. Additionally, the participants were asked to “think aloud” during the analysis, i.e., to comment on what they see, how they interpret it, and what conclusion they draw. For capturing what the participants say, audio recording was used. The same software was used for eye tracking and audio recording; hence, the eye tracking and audio data were synchronised and could be re-played together. In each analysis session, there was a facilitator answering arising questions of participants.

The collected data and voice narration of the analysis sessions were analysed. This analysis allowed us to identify different strategies used by analysts and understand how they are reflected in eye-tracking data. The results of our analysis provide guidance for designing map-based visualization components in future systems and, eventually, enable building intelligent guidance components for supporting data analysis.

In the following sections of the report, we present the state of the art in analysis of eye-tracking data, then consider an overall workflow of eye-tracking data analysis and systematically discuss the space of data analysis tasks. We shall present the analysis tasks selected for evaluation within the project and their instantiations for the 3 pilots. Results of the analysis will be discussed and conclusions are drawn.

2 Eye-tracking data analysis: State of the art

2.1 Visually-driven analysis of eye-tracking data

The application of eye tracking technology as a means of evaluating human behaviour has been established in many different research fields [8]. Due to the interdisciplinary constellation of researchers, the specific analysis tasks may also differ between the fields. While one researcher might be interested in the physiological measures (e.g., eye movement speed [15]), another wants to know in what order specific areas of interest on a visual stimulus were investigated [7]. Despite the differences between the research fields, it is possible to derive a high-level task categorization from a data perspective. Since the structure of the recorded data is usually identical in all eye tracking experiments, it has been proposed to categorize the analysis tasks according to three main data dimensions (Where? When? Who?) and three elementary analysis operations (compare, relate, and detect) [17].

Depending on the research question, a statistical analysis of established eye tracking metrics [14] can be sufficient. However, the more complex the analysis task becomes, the more visual aid is usually required to interpret the data. Regarding the increasing amount of eye-tracking data recorded during experiments [2], it is reasonable to incorporate visual analytics techniques that combine automatic data processing with interactive visualization [1] into the analysis process.

As a starting point, the analysis of eye-tracking data is usually supported by some basic visualization techniques. For statistical measures, the application of statistical plots depicting the changes of a variable over time can already be helpful to interpret the data. In these cases, the visual stimulus is neglected. If the visual stimulus is important for the analysis, additional visualization techniques are usually included in the software suites of the major eye tracking vendors such as the Tobii Pro X2[2] system that is available in Track&Know.

For many years, gaze plots and attention maps were (and still are) the most popular visualizations that include information about the underlying visual stimulus. However, not all analysis tasks are facilitated by these techniques. For example, it is hard to interpret changes over time by simply replaying the animation [21]. Therefore, many new techniques have been developed over the last years to address this and many other analysis tasks, summarized by Blascheck et al. [4][5]. Additionally, as a beneficial but also challenging aspect, apart from the pure eye movement data, a wealth of additional data sources can be integrated into an experiment [2]. Such a collection of heterogeneous data sources often impairs a combined analysis by statistical means and makes a visual approach indispensable.

2.2 The Eye-tracking data analysis pipeline

In [17], authors formulate the way from conducting an eye tracking experiment to gaining insight in the form of a pipeline (Fig. 1) that is an extended version of the generic visualization pipeline [9]. The acquired data consisting of eye movement data and complementary data sources is processed and optionally annotated before a visual mapping is performed. By interacting with the data and the visualization, two loop processes are started: a foraging loop to explore the data and a sense-making loop to interpret it [18], to confirm, reject, or build new hypotheses from where knowledge can be derived. Since the analysis task plays an important role in all steps of the pipeline, the authors first discuss the underlying data and how it is processed before introducing their categorization of analysis tasks.

[2] https://www.tobiipro.com/product-listing/tobii-pro-x2-30/

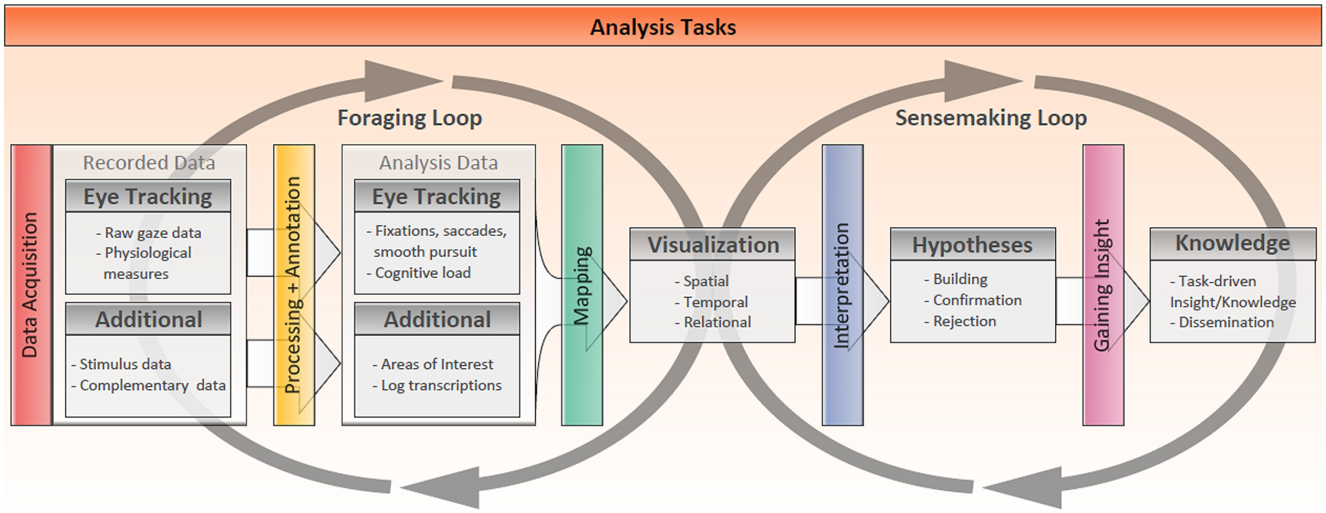


Figure 1: Eye-tracking data analysis pipeline. Source: [17]

2.2.1 Data acquisition

Eye movement data combines several data dimensions of spatio-temporal nature. Researchers distinguish between dimensions directly stemming from the recording of eye movements (raw gaze, physiological measures) and additional data sources serving as complementary data that can help achieve more reliable analysis results when combined with eye movement data. Typically, the displayed stimuli are an additional data source that can usually be included in the analysis process, since they are the foundation of most experiments anyway. Additional data sources provide complementary data such as verbal feedback, electroencephalography (EEG) data, and key press protocols.

The research question to be answered by means of eye tracking typically defines how the experiment is designed and which data will be recorded. Most scenarios predefine also the visual stimulus. Exceptions are, for example, “in-the-wild” experiments with mobile eye tracking where it becomes much more difficult to control the experiment parameters.

2.2.2 Processing and annotation

From the time-varying sequence of raw gaze points, more data constructs can be derived in a processing step. Fixations, saccades, smooth pursuits, and scanpaths are the most important data constructs [14]. In this processing step, automatic data-mining algorithms can be applied to filter and aggregate the data. Clustering and classification are prominent processing steps: For example, raw gaze points can be clustered into fixations and labelled. As another example, the convex hull of a subset of gaze points can be extracted to automatically identify areas of interest (AOIs). In general, the annotation of AOIs plays an important role in this step.

From the visual content of a stimulus (e.g., a picture or a video), AOIs can be annotated, providing semantic interpretation of the stimulus. With this information, additional data such as transition sequences between AOIs can be derived. Therefore, analysts can either rely on automatic, data-driven approaches to detect AOIs, or define them manually. Basically, there are two approaches: either defining areas or objects by bounding regions on the stimulus and calculating hits with the gaze data, or labelling each fixation individually based on the investigated content. Especially for video sequences, this annotation is a time-consuming step that often takes more effort than the rest of the analysis process.

From the additional data sources, recorded protocols and log files can typically be derived. It should be noted that each additional data source requires a synchronization with the recorded eye movement data, which can be difficult considering different sampling rates and irregularly sampled data (e.g., think-aloud comments) [3]. The processed data can finally be used for the mapping to a visual representation.

The analysis task influences what filters are applied to the data and what AOIs are annotated. For explorative scenarios in the context of visual analytics, visualization and processing are tightly coupled in a foraging loop, where the analyst can identify relevant data artefacts through interaction with the visualization.

2.2.3 Visualization

According to Blascheck et al. [4][5], the main categories of state-of-the-art visualization techniques for eye tracking are spatial, temporal, and relational data representations. Appropriate visualizations need to be selected according to the main data dimension that is required to perform an analysis task [17]. . In the foraging as well as the sense-making loop, the visualization has to convey the relevant information and should provide enough interaction supported by automatic processing to adjust the visualization to the specific needs of a certain analysis task.

It may be noted that only a few visualization techniques for eye movement data also take into account additional data sources for an enhanced visual design in order to explore the data. This is actually noteworthy for future work since those data sources may build meaningful input for sophisticated data analyses if they are combined with the traditional eye movement data.

2.2.4 Interpretation

For the interpretation of the visualization, we can distinguish between two strategies: (1) applying visualization to support statistical measures and (2) performing an explorative search. In the first case, hypotheses are typically defined before the data is even recorded. Therefore, inferential statistics are calculated on appropriate eye tracking metrics, providing p-values to either support or reject hypotheses. Here, visualization has the purpose to additionally support these calculations. In the second case, the explorative search, hypotheses might be built during the exploration process.

Filtering and re-clustering data, adjusting the visual mapping and reinterpreting the visualization can lead to new insights that were not considered during the data acquisition. This explorative approach is particularly useful to analyse data from pilot studies. Building new hypotheses, the experiment design can be adjusted and appropriate metrics can be defined for hypothesis testing in the final experiment.

The interpretation of the data strongly depends on the visualization. With a single visualization, only a subset of possible analysis tasks can be covered. For an explorative search where many possible data dimensions might be interesting, a visual analytics system providing multiple different views on the data can be beneficial. It allows one to investigate the data in general before the analysis task is specified.

2.2.5 Gaining insight

As a result of the analysis process, knowledge depending on the analysis task is extracted from the data. As discussed before, this knowledge could be insights that allow the researchers to refine a study design or conduct an entirely new experiment. In the cases where visualization has the main purpose to support statistical analysis, it often serves as dissemination of the findings in papers or presentations. In many eye tracking studies, this is typically the case when inferential statistics are performed on eye tracking metrics and attention maps are displayed to help the reader better understand the statistical results.

2.3 Categorization of Analysis Tasks

The visualization pipeline for eye-tracking data (Fig. 1) shows the steps in which analysis tasks play an important role. For the experienced eye tracking researcher, the first two steps, data acquisition and processing, are usually routine in the evaluation procedure. Mapping is a less trivial step in which the analysis task has to be considered. When the analysis task is clear, the chosen visualization has to show the relevant information. The authors of [17] propose a categorization of analysis tasks that aims at helping with choosing appropriate visualizations. First, they identify three independent data dimensions in eye-tracking data:

- Where? For these tasks, space is the most relevant data dimension. Typical questions in eye tracking experiments consider where a participant looked at.
- When? Tasks where time plays the most important role. A typical question for this dimension is: when was something investigated the first time?
- Who? Questions that investigate participants. Typical eye tracking experiments involve multiple participants and it is important to know who shows a certain viewing behaviour.

With these three independent dimensions, visualizations can be applied to display dependent data constructs (e.g., fixation durations). Many visualization techniques may not be restricted to just one of these dimensions but may facilitate different combinations of them.

Additionally to the three dimensions, there are general analytical operations, which can be also related to other taxonomies (e.g., the knowledge discovery in databases (KDD) process [11]):

- Compare: Questions that consider comparisons within one data dimension.
- Relate: Questions that consider the relations between data dimensions and data constructs.
- Detect: Questions about summarizations and deviations in the data.

This categorization is based on the survey by Blascheck et al. [4][5], the work of Andrienko et al. [1], and the work of Kurzhals et al. [16]. The cited publications provide a more in-depth overview of current state-of-the art visualization and visual analytics approaches for the analysis of eye-tracking data. In the following, we briefly recap those aspects. A more detailed discussion of all aspects can be found in [17].

2.3.1 Where? – Space-Based Tasks

Typical questions that consider the spatial component of the data are often concerned with the distribution of attention and saccade properties. Statistical measures such as standard deviations, nearest neighbour index, or the Kullback-Leibler divergence provide an aggregated value about the spatial dispersion of gaze or fixation points. If a saccade is defined as a vector from one fixation to another, typical where questions can also be formulated for saccade directions. If AOIs are available, measures such as the average dwell time on each AOI can be calculated and represented by numbers or in a histogram.

If the stimulus content is important for the analysis, attention maps [6] and gaze plots are typically the first visualizations that come to mind. Attention maps scale well with the number of participants and recorded data points, but totally neglect the sequential order of points. With an appropriate colour mapping and supportive statistical measures, an attention map can already be enough to answer many questions where participants looked at, if the investigated stimulus is static.

Space-based tasks for dynamic stimuli, such as videos and interactive user interfaces, require a visualization that takes the temporal dimension into account considering also the changes of the stimulus over time. If AOIs are available, we refer to the next section, because in this case, when and where are tightly coupled.

2.3.2 When? – Time-Based Tasks

Eye movement data has a spatio-temporal nature often demanding for a detailed analysis of changes in variables over time. Questions in this category typically have the focus on a certain event in the data (e.g., a fixation, smooth pursuit) and aim at answering when this event happened. Considering the detection of specific events over time, many automatic algorithms can be applied to identify these events. Automatic fixation filtering [20], for example, calculates when a fixation started and ended. For semantic interpretations, combining data dimensions to answer questions when was what investigated, the inclusion of AOIs is common. For statistical analysis, measures such as the time-to-first-hit in an AOI can be calculated.

Timeline visualizations are a good choice to answer questions related to this category. In general, timeline representations depict an additional data dimension or construct, allowing one to combine the data relevant for spatial analysis (e.g., gaze heatmaps) with its temporal progress.

2.3.3 Who? – Participant-Based Tasks

Typical questions raised when looking at recorded participants' data can be categorized into those concerning only a single individual or a larger group of people. Inspecting the viewing behaviour of participants can provide insights into the visual task solution strategies applied by them. For a single participant, a traditional gaze plot is useful to interpret the scanpath, assuming that the recorded trajectory is not too long nor located in just a small stimulus subregion. Generally, most visualization techniques for multiple participants work fine also for an individual participant. For the comparison of multiple participants, gaze plots are less scalable, because of the massive overplotting that occurs when many participants' scanpaths are displayed in one representation.

To ease the comparison of scanpaths, specific metrics to identify similarities between participants can be applied, such as the Levenshtein or Needleman-Wunsch distance [16, 46]. Based on visited AOIs, a string is derived that can be compared by the mentioned similarity measures. As a consequence, scanpaths from many participants can be compared automatically. To interpret the comparison results, a visual representation of the scanpaths that supports the similarity measure can be very helpful. A possible representation is so-called "scarf plot" in which sessions of different participants are represented by coloured bands oriented along a time axis and stacked one below another (see, for example, Fig. 16 in the following report). Coloured segments within the bands show which AOIs were visited in the corresponding time intervals.

2.3.4 Compare

Comparison in general can be seen as one of the elementary analysis operations performed during the evaluation of eye tracking experiments. In fact, statistical inference is typically calculated by comparing distributions of a dependent variable. For example, fixation durations between different stimulus conditions can be compared with an ANOVA to find out whether a significant difference between the two distributions exists. However, inferential statistics can only provide the information that a difference exists. To interpret the difference between the conditions, or, if the low number of available participants limits applicability of statistics, a visual comparison is usually a good supplement to the statistical calculations.

Comparison tasks are typically supported by placing several of the visualized data instances next to each other in a side-by-side representation, sometimes denoted as small multiples visualization. Each data instance is visually encoded in the same visual metaphor to facilitate the comparison.

2.3.5 Relate

In most analysis scenarios, not only a single dimension such as space, time, or participants is in the research focus. A combination of two, three, or even more dimensions and data constructs is included in the analysis to explore the data for correlations and relations between the data dimensions.

Investigating relations between AOIs across participants is an important aspect for analysis tasks in this category. Relations between AOIs are often investigated by transitions between them. They can show which AOIs have been looked at when and in what order. A standard statistical measure is the transition count. Given enough samples (participant sessions), transition matrices or Markov models can give valuable insight into search behaviour of a participant [14].

2.3.6 Detect

Detecting patterns of common viewing behaviour is important and often achieved by summarizations or aggregation of the data. Such summarizations can also be applied to find outliers in the data which might either result from a problem of the hardware or from unexpected and potentially interesting behaviour of a participant.

Descriptive statistics are often applied to achieve this goal. Calculating the average fixation duration, the variance of saccade amplitudes, or the mean scan path length are some examples. Box plots are typically used to represent these values and additionally depict outliers as a simple to understand graph. However, more sophisticated visualization techniques can be utilized to summarize the eye movement data and detect outliers visually. Summaries can be created for the raw data points, for aggregated data using AOIs, or for the participants. One possibility is to depict one dimension of the fixation position plotted against time [12]. This allows investigating the general scanning tendency of a participant with regard to specific, relevant stimuli.

An AOI view facilitates a simple summarizations of eye movement data on the basis of AOIs, and may also be used to find deviations in the data. For example, an AOI may not have been looked at during the complete experiment by one or multiple participants. This may be an indicator that the AOI was not needed to perform the experiment task or participants missed important information to perform the task. AOI time lines can help answer this question.

2.4 Specifics of our study

Our research question in the eye tracking experiments was: <u>What can eye-tracking data tell us about the process of visual exploratory analysis of spatial phenomena represented on maps?</u> Our aim was to derive knowledge of general relevance utilizable in designing visual displays and interaction techniques aimed to support analysis of various spatial phenomena.

The ambition to derive knowledge of high level of generality differentiates our study from the typical studies involving eye tracking. Accordingly, the common tasks of eye-tracking data analysis categorized and described by previous researchers need to be re-thought and re-formulated.

- Regarding the “where” dimension, identifying and characterizing specific AOIs at which participants looked can be just the first step in analysis providing material for reasoning and generalization. From the perspective of our research goal, AOIs with their characteristics and spatial arrangements are low-level information from which we need to extract general patterns invariant to particular spatial locations.
- Concerning the “when” dimension, we are not much interested in particular absolute or relative times when participants were looking at different AOIs, nor in the sequences of visiting the AOIs. As for the spatial

dimension, we want to derive general patterns of analysts' behaviour and general strategies analysts may take in exploratory tasks.
- Concerning the "who" dimension, we are interested in finding commonalities and differences between the analytical processes they apply. This focus of interest lies far beyond simple comparison of scanpaths. Moreover, we expect large differences and little or no similarity between scanpaths in terms of visits of particular AOIs and the times, durations, and sequence of these visits. Therefore, usual methods of scanpath comparison are not helpful to us.

Hence, for our analysis, we need to adapt state-of-the-art approaches and/or find new approaches that are appropriate to our research goals.

In the following, we describe the design of the experiments for the three pilots of project in Section 3. Section 4 describes the process of performing the experiments with professional subjects and reports about the analysis of the collected data using visual analytics approaches (Sections 5.2, 5.3, and 5.4). Section 5 reports about application of machine learning methods.

3 Application to pilots: Preparations

According to the types of data and analysis tasks addressed in Track&Know, it was decided to perform experiments with the use of maps representing aggregated data describing two or more pilot-specific phenomena. The rationale for this decision is explained in detail in Section 1.5. A map display showing aggregated data is a common component of the dashboards designed for all 3 pilots of the project. Our research task thus concerns the use of maps, i.e., it can be formulated as follows:

- Study how domain experts use maps to analyse and interpret spatial distributions and relationships between spatial phenomena represented on maps.

A common approach to representing spatial distributions of two phenomena on a single map was to use the choropleth map technique (i.e., painting of areas with varying degrees of darkness) for one phenomenon and proportionally sized geometric symbols (such as circles) or diagrams (such as pie charts) for other phenomena. The distributions are represented in an aggregated form as counts (or other summary statistics, e.g. average or median) by cells of a regular grid or units of territory division. In addition, locations of potentially relevant spatial objects can be represented by simple uniform symbols, such as dots. The representation of the data is superimposed on a background map showing the geographical context, including cities, roads, coastlines, and other kinds of geographic information.

The general tasks of studying spatial distributions and relating multiple spatial distributions have been instantiated for the three pilots. As the first step in preparation to the study, we designed and generated pilot-related maps we deemed suitable for the experiments. We sent these maps to the pilot-responsible project partners with explanations, formulations of the intended tasks for the study participants, and requests for partners' feedback. Here is the common part of the RFC (Request For Comments) documents that were sent to all partners:

Dear partners,

We are now making preparations to the experiments on the use of visual representations for exploratory analysis of complex data, in which we are going to employ eye tracking. As you know, we had to change our original plan of involving project partners into eye tracking experiments. The current intention is to involve instead colleagues from our institute.

Nevertheless, we want to conduct good experiments and obtain data that will be useful for our project and relevant to the application partners. For this purpose, we want to use visual stimuli created from the partners' data and related to their interests.

The general task for the participants of the experiment will be to detect possibly interesting and/or important patterns on a map representing spatial distributions of two supposedly related phenomena. The participants will be expected to compare the distributions and identify a few patterns they deem interesting.

Before using the maps in the eye tracking experiments, we want the application partners to evaluate the appropriateness of the maps according to the following criteria:

- Relevance of the represented distributions to the application domain;
- Understandability of the map content and the visual encoding;
- Presence of possibly interesting patterns.

Furthermore, we ask the partners to identify a couple of patterns that could be considered as important or interesting or, possibly, surprising and suspicious.

<… pilot-specific part …>

We kindly ask you to evaluate the suitability of this map with respect to the criteria of the relevance, understandability, and presence of interesting patterns. Critical comments and suggestions for improvement are welcome!

If you find the map potentially suitable, please, identify 2-3 spatial patterns deserving attention. Please, characterise briefly what is interesting; if possible, mark the patterns on the map.

Thanks in advance for your feedback.

The requests also contained pilot-specific parts, which will be presented in the following.

3.1 Pilot 1: Fleet management (VFI)

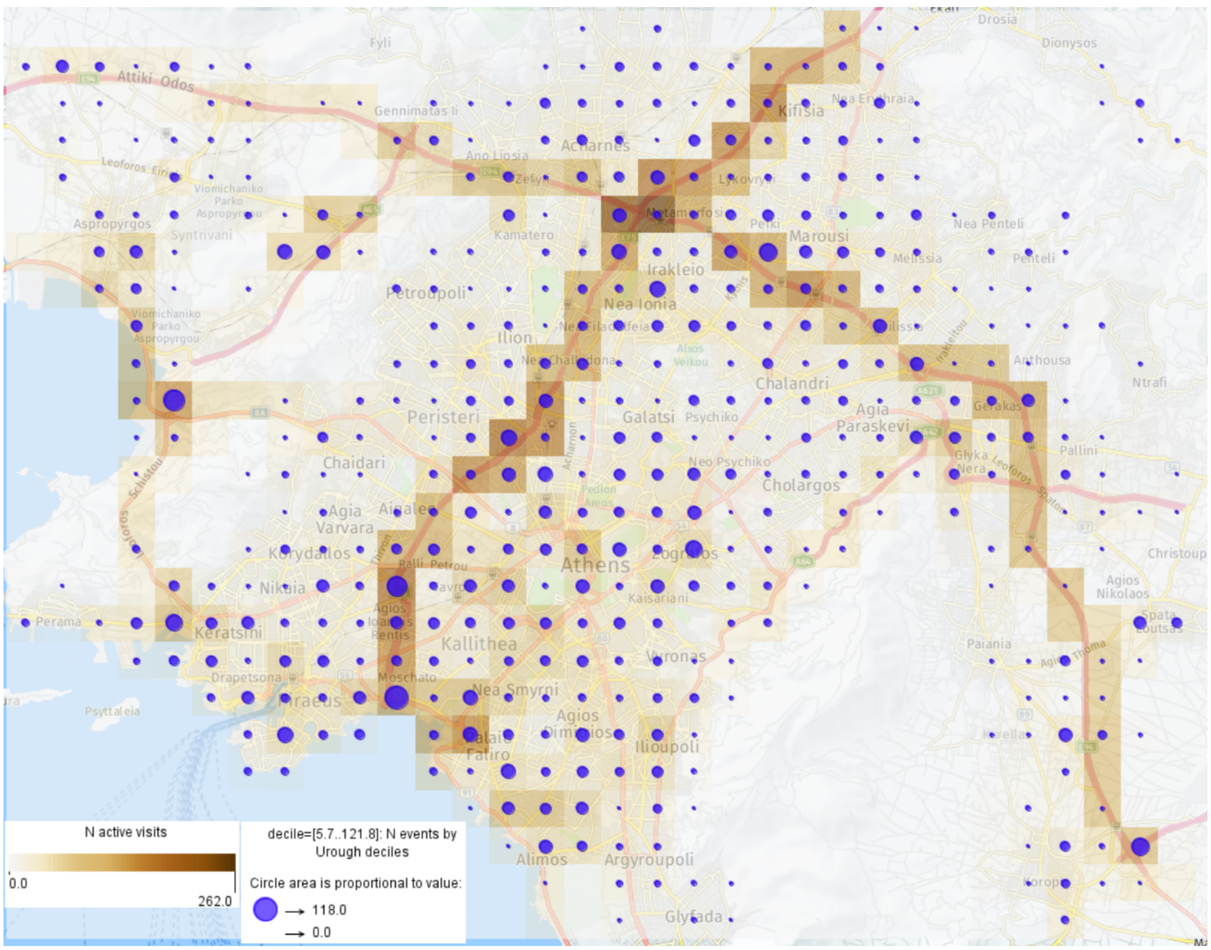


Figure 2: A map designed for the experiments based on the VFI pilot.

3.1.1 Map and task design

The general idea of the visualisation was to show distribution of some kind of driving events together with the distribution of the overall traffic intensity, so that test participants could analyse whether the distributions correlate and whether there are local disruptions of an overall pattern. The distribution of the overall traffic intensity was represented using the choropleth map technique. The distribution of the events was shown by means of proportionally sized symbols (see Fig. 2). The following text, which was sent to the Pilot 1 hosts as a part of the RFC document, explains in detail the data chosen for presenting on a map and the visual design:

In relation to the Vodafone pilot, we have prepared a map using the results of data processing performed by the partners from the University of Piraeus. The data consist of trajectories of 536 vehicles enriched by various position-associated features that were calculated during the data processing. In our example map, we use the feature called Urough, which characterizes the speed roughness. The values of the feature range from 0 to 121.8, but 80% of the positions have the values below 3.4 and 90% of the positions have the values below 5.7. We treat the top 10% of the values, i.e., 5.7 or higher, as events of rough speed.

The map that we have prepared shows the distribution of the rough speed events on top of the overall traffic distribution. The distributions are represented in an aggregated form. The aggregation has been done by a regular square grid with cell sizes 1x1 km. The overall traffic distribution is represented by the counts of the cell visits by the vehicles (computed after excluding stop points and large spatial or temporal gaps between records). The event distribution is represented by the per-cell counts of the rough speed events.

The background layer of the map consists of the cells painted in shades of brown. The shades encode the traffic amounts, more specifically, the counts of the cell visits. The foreground layer consists of proportionally sized circle symbols painted in dark blue. The sizes encode the counts of the rough speed events.

The project partners approved the design and analysis tasks. The following map demonstrates an example of spatial patterns marked on the map by the pilot-responsible partners:

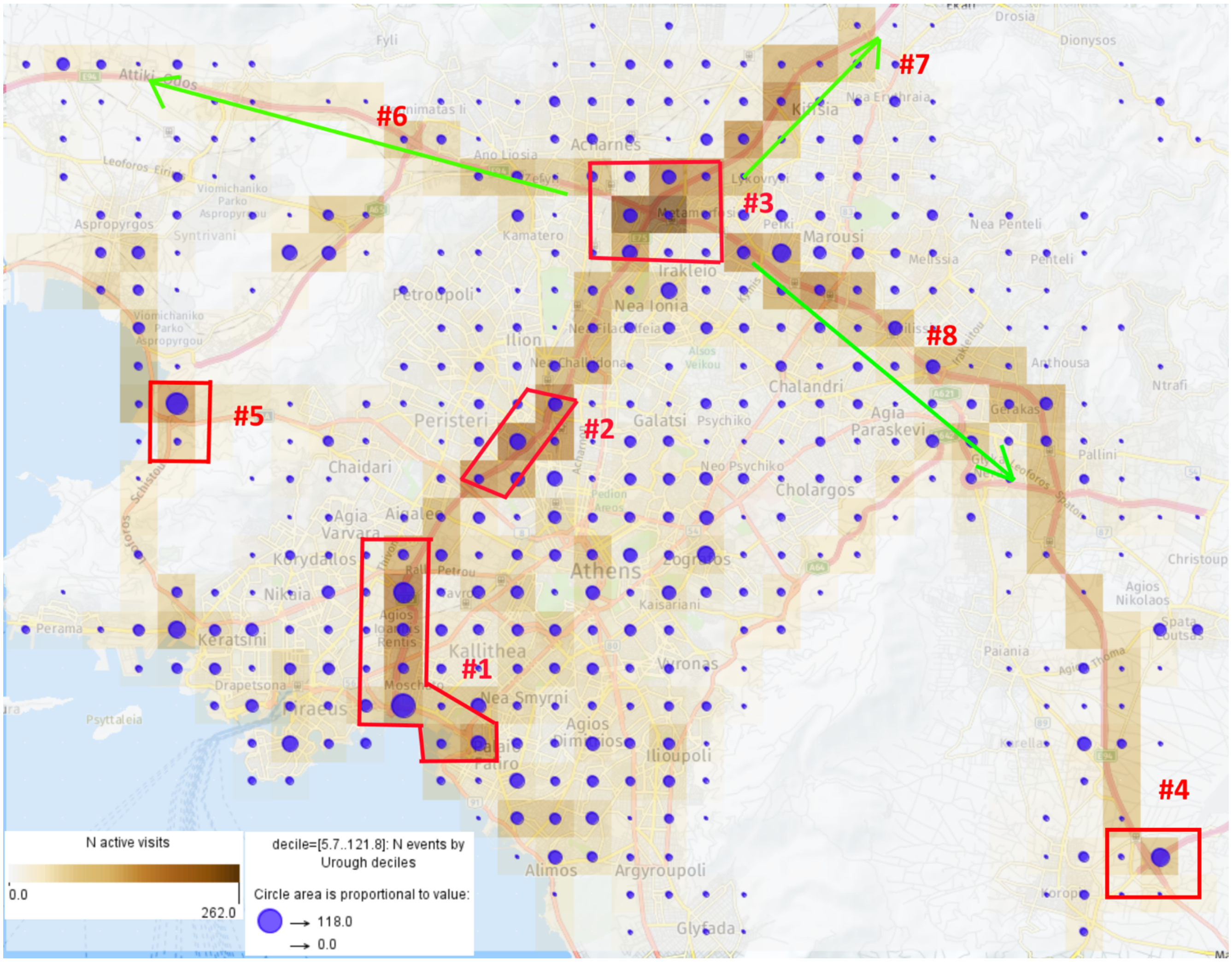


Figure 3: VFI: map with expert-marked locations of patterns

3.1.2 Instructions for participants.

After collecting the feedback, the following instructions were prepared for giving to test participants:

1. **Step 1: Task description**

Your task will be to explore the spatial distribution of a particular kind of driving events, namely, rough changes of the vehicle speed during driving.

The original data were obtained by tracking journeys of business vehicles. The companies utilising the vehicles are interested in an assessment of the driving behaviours of the vehicle drivers with regard to traffic safety and vehicle longevity. Researchers attempt to design informative and understandable metrics that could characterise relevant aspects of drivers' behaviours. The map you will use represents aggregated data obtained from trajectories of 536 vehicles. The background painting shows the overall distribution of the traffic over the territory. On top of it, proportionally sized symbols represent the amounts of the driving events when the speed was changing in a rough manner.

2. Step 2: Familiarization with the map

Please get familiar with this map!

The data in the map are aggregated by a grid with cell sizes 1x1 km. The background painting in shades of brown shows the overall traffic distribution, i.e., how many vehicles appeared in each cell of the grid. **Darker shades represent higher traffic amounts**.

The event distribution is represented by the per-cell **counts of the rough speed change events**. The counts are visually encoded by proportionally sized **circle symbols** painted in dark blue.

3. Step 3: Visual analysis task

Your task is to examine whether the distribution of the events of rough speed change differs from the distribution of the overall traffic, in particular, whether there are places where the amounts of the events are notably higher or lower than could be expected based on the traffic intensity. 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.

3.2 Pilot 2: Insurance (SIS)

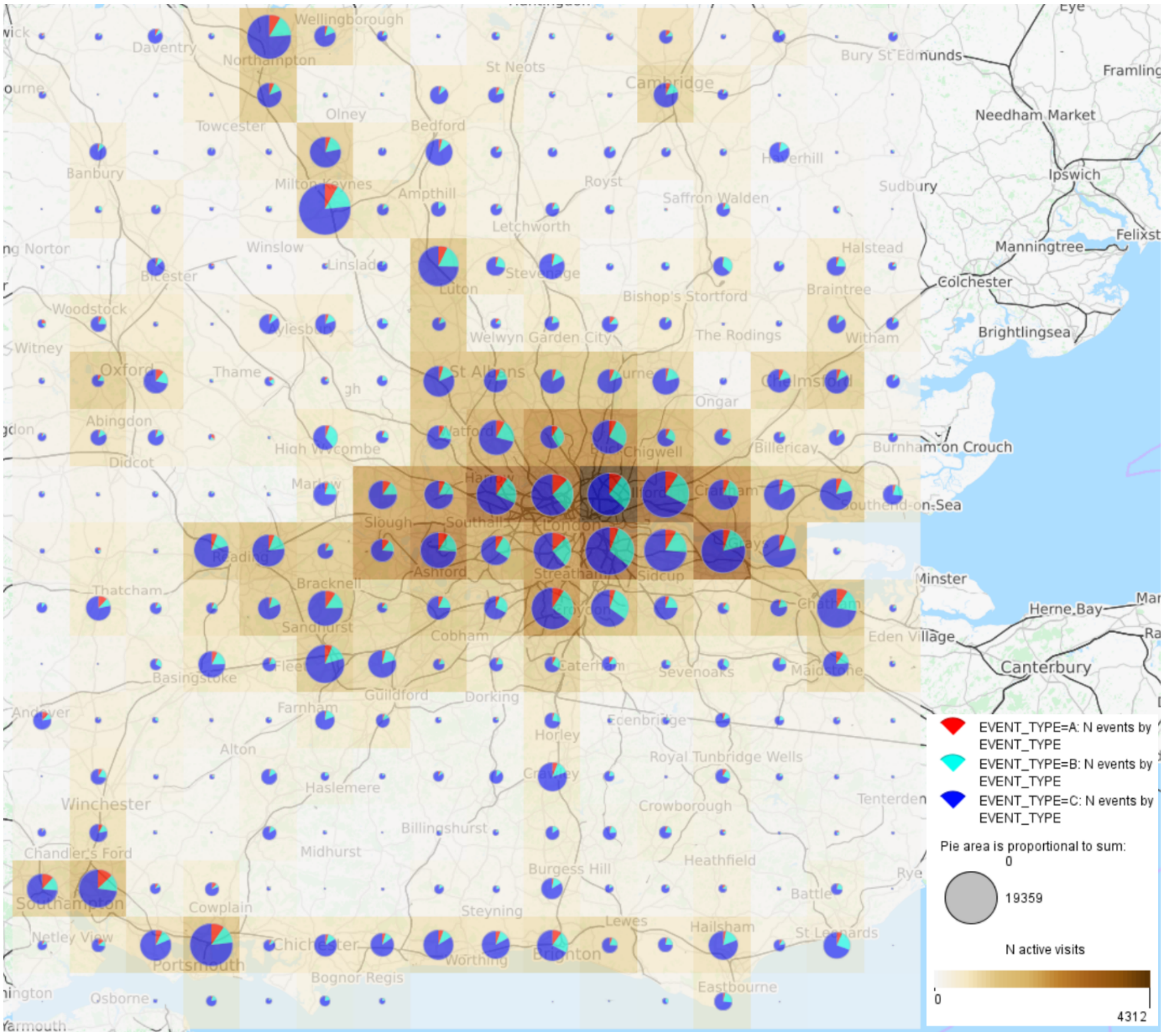


Figure 4: A map designed for the experiments based on the SIS pilot (initial version)

3.2.1 Initial design.

Initial idea for the visualisation was similar to the one used for the VFI pilot: show the distribution of driving events on top of the distribution of the overall traffic intensity. To make the participants' task different from the VFI-related task and also more challenging, we decided to represent the distributions of three types of driving events by means of the pie chart technique (see Fig. 4). The following description was sent to the Pilot 2 hosts as a pilot-specific part of the RFC document:

In relation to the Sistematica pilot, we have prepared a map using the data describing the vehicle trajectories and the driving events on the territory of London and surrounding areas for the time period from January 1 to January 13, 2017. The data were aggregated by a square grid with the cell size 10x10 km. Aggregated movements (obtained from the trajectories) are represented on the map by cell shading and counts of the driving events by proportionally sized pie charts.

The background layer of the map consists of the cells painted in shades of brown. The shades encode how many times the cars from the available dataset moved in or through the cells; darker shades correspond to higher counts of the visits. The foreground layer consists of pie charts drawn inside the cells. The pie sizes are proportional to the total counts of 3 types of events and the sector sizes show the counts by the types: red – acceleration events, cyan – braking events, and blue – cornering events.

3.2.2 Alteration of the original visualisation and task

In response to the proposal, the partners working on the SIS pilot suggested a different idea for the visualisation and the analysis task for the test participants. The map should present the distribution of the driving events together with the distribution of the vehicle crashes. The task should be to investigate whether crashes may be related to a particular type of driving events or to the overall frequency of all types of driving events. According to this suggestion, we created a map as shown in Fig. 5, which was approved by the SIS-responsible partners. For the new version of the map, we created the following instructions for the participants.

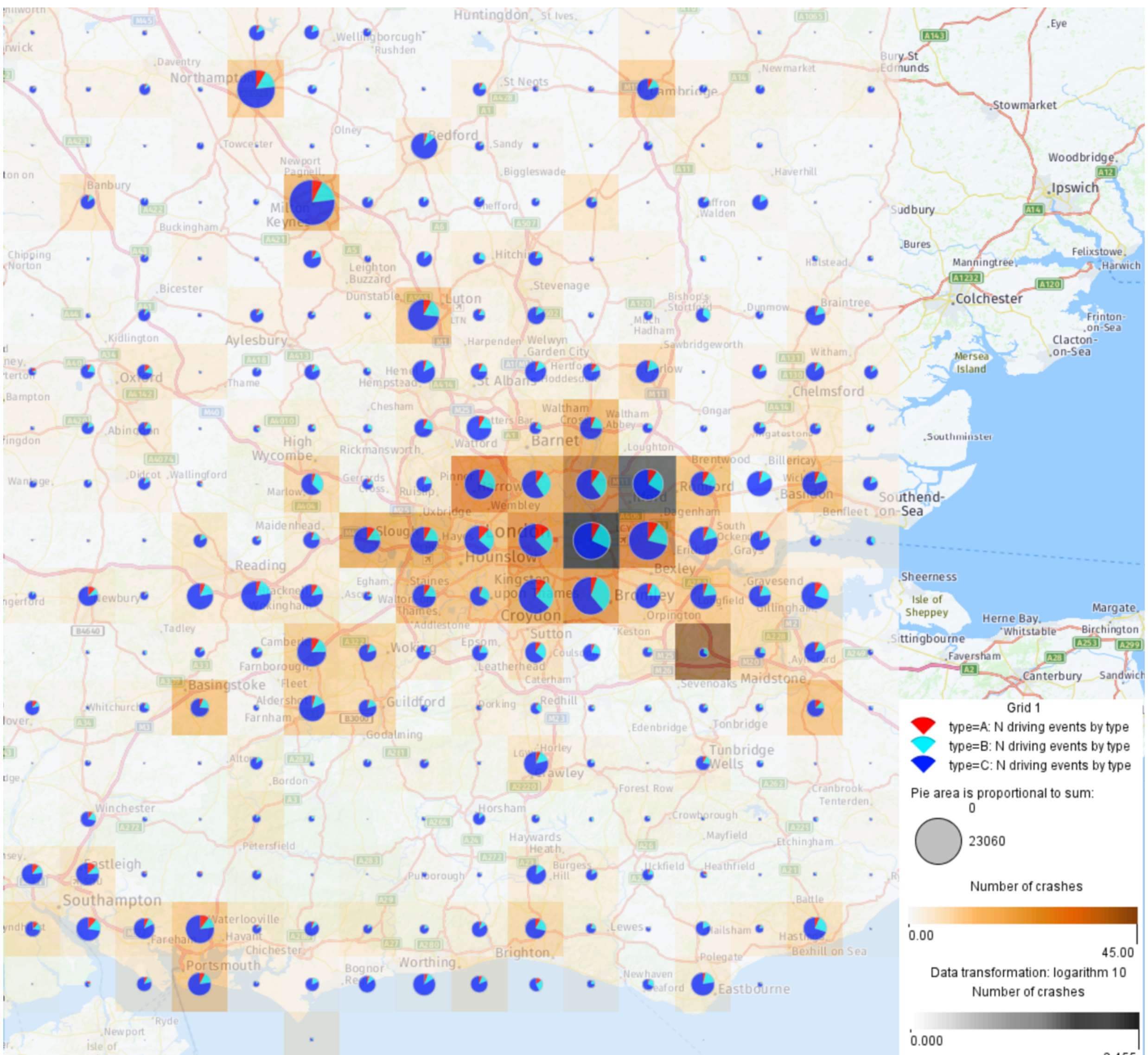


Figure 5: A revised version of the map designed for the experiments based on the SIS pilot.

3.2.3 Instructions for participants

After re-designing the visualisation and re-defining the analysis task, the following instructions were created for the test participants.

1. Step 1: Task description

Your task will be to explore the spatial distribution of three kinds of driving events in relation to the spatial distribution of vehicle crash events.
The original data were collected by a car insurance company by tracking journeys of insured vehicles. The company is interested in assessing the driving behaviours of the car drivers from the perspective of traffic safety. The assessment is supposed to affect the insurance fee paid by the vehicle owners. The insured cars are equipped with monitoring devices that register driving events of high acceleration, harsh braking, and sharp cornering.
The map you will use represents aggregated data about the driving events together with data about vehicle crashes that were also collected by the insurance company. The set of cars that were involved in the crashes is not the same as the set of cars that produced the three kinds of the driving events. Nevertheless, the insurance company wants to find out if there exist any global or local relationships between the risk of a crash and particular kinds of driving events.
In the map you will explore, the background painting shows the spatial distribution of the crashes. On top of it, proportionally sized diagrams represent the amounts of the driving events of three types.

2. Step 2: Familiarization with the map

Please get familiar with this map!
The data in the map are aggregated by a grid with cell sizes 10x10 km. The background painting in shades of brown shows the spatial distribution of the car crashes, i.e., how many crash events happened in each cell of the grid. **Darker shades represent higher numbers of crashes**. There are two cells with **exceptionally high numbers of crashes**, namely, 285 and 72 (while the next highest value is 45). These two cells are painted in **shades of grey** generated after applying a logarithmic transformation to the values.
The event distribution is represented by per-cell counts of three types of events: high acceleration (**A**), harsh braking (**B**), and sharp cornering (**C**). The **total event counts** are encoded in **proportional sizes** of the pie charts. The sizes of the pie sectors represent the fractions of the different types of events: **red** for high **acceleration**, **light blue** for harsh **braking**, and **dark blue** for sharp **cornering**.

3. Step 3: Visual analysis task

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.

3.3 Pilot 3: Healthcare (PAP)

According to the research conducted in the PAP pilot and the kind of the findings obtained, we decided that the map to use in the experiments should present the distribution of spatially aggregated results of testing patients for the presence and severity of the obstructive sleep apnoea (OSA) condition. These data should be presented together with information about the variation of the population deprivation over the underlying territory. The task is to check if the distribution of the patients referring for tests and the distributions of the different levels of the OSA severity may be related to the deprivation. Additionally to the aggregated patient data and the deprivation scores, the map should include the locations of the clinics distributing the devices for making the OSA tests. The initial map design is shown in Fig. 6.

3.3.1 Initial design.

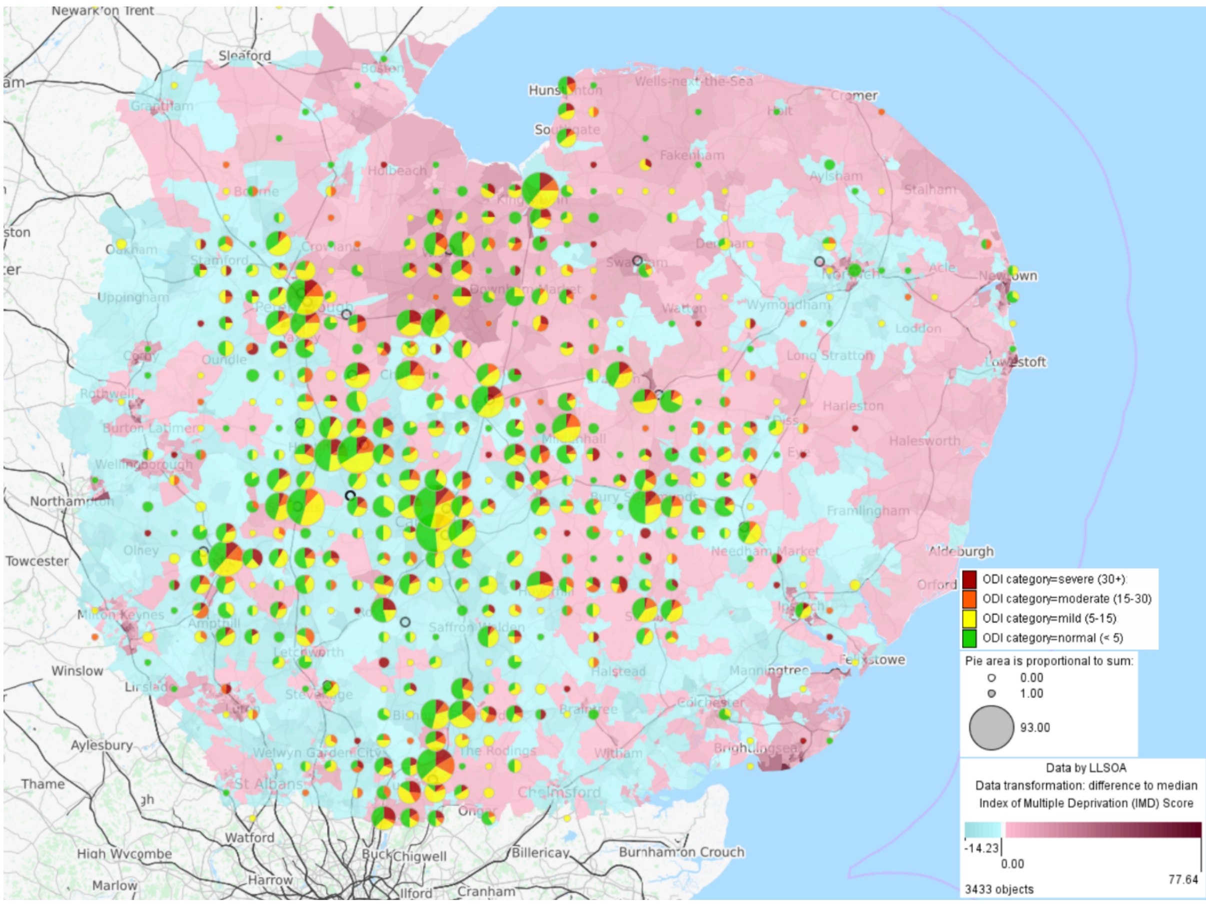


Figure 6: A map designed for the experiments based on the PAP pilot (initial version)

The following request was sent to the Pilot 3 hosts as a part of the RFC document:

In relation to the Papworth pilot, we have prepared a map using the data about the patients and the data about the population deprivation on the underlying territory.

The background layer in the map shows the variation of the Index of Multiple Deprivation (IMD) over the territory, based on the official data specified for the Lower Layer Super Output Areas (LSOA). The values of the

IMD are represented by shades of two colours, cyan and purplish red. The shades of cyan represent values below the territory's median and the shades of purplish red – the values above the median; the darker, the greater.

The foreground map layer consists of pie charts representing the patients' data aggregated by a square grid with the resolution 5x5 km. The grid lines are not shown for reducing the visual clutter and having a clearer view. The pie sizes are proportional to the total counts of the patients per grid cell. The coloured sectors of the pies represent the proportions of the patients grouped into four categories according to the severity of the ODI: normal (green), mild (yellow), moderate (orange), and dark red (severe). The visual encodings are explained in the legend shown in the lower right corner.

The project partners approved the design in general but recommended some modifications. The following map presents an example of interesting patterns that were identified by an expert and explained as "Large clusters of patients appear to be coming from 6 main high population areas (Cambridge, Harlow, Bedford, Huntingdon, Peterborough and Kings Lynn), and very few are coming from the Eastern region".

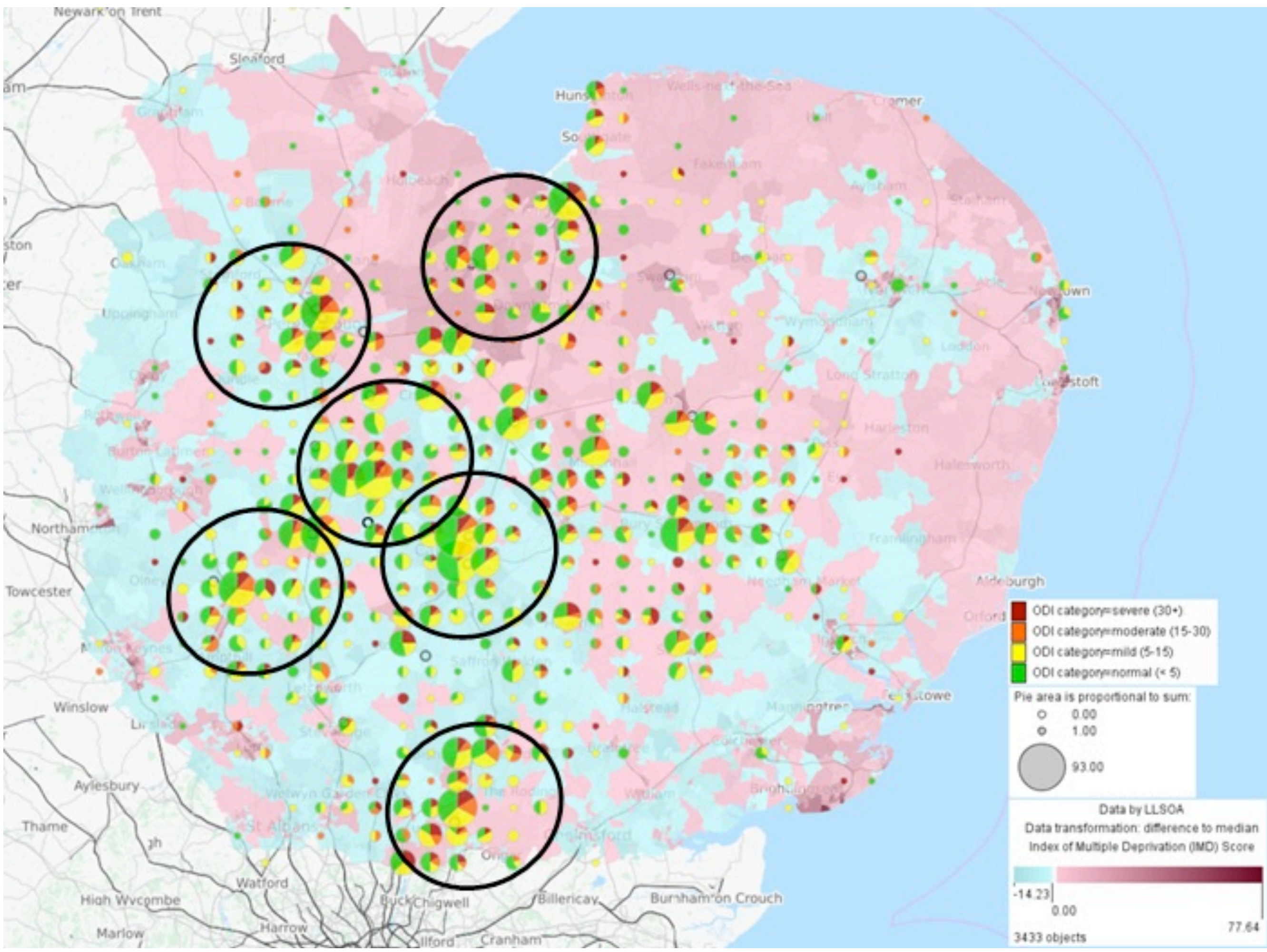


Figure 7: PAP: the map with expert-marked locations of patterns

The partners suggested that the visualisation needs to be simplified. Instead of presenting 4 levels of the OSA severity according to the test results, they suggested to aggregate the data into two larger categories: (1) condition that does not require medical treatment and (2) condition that requires medical treatment. Concerning the variation of the deprivation levels over the territory, the partners suggested that the information needs to be aggregated by the same spatial units as the patients data, i.e., by the square grid cells. Besides, the diverging colour scale with two colour hues should not be used as it complicates the representation and attracts too much attention, thus distracting from the patients data. It was also suggested to aggregate data by larger grid cells, to avoid overcrowding of the map with diagrams, which can be seen in the initial version of the map. After several steps of collecting feedback and redesigning the visualisation, the final map shown in Fig. 8 was created.

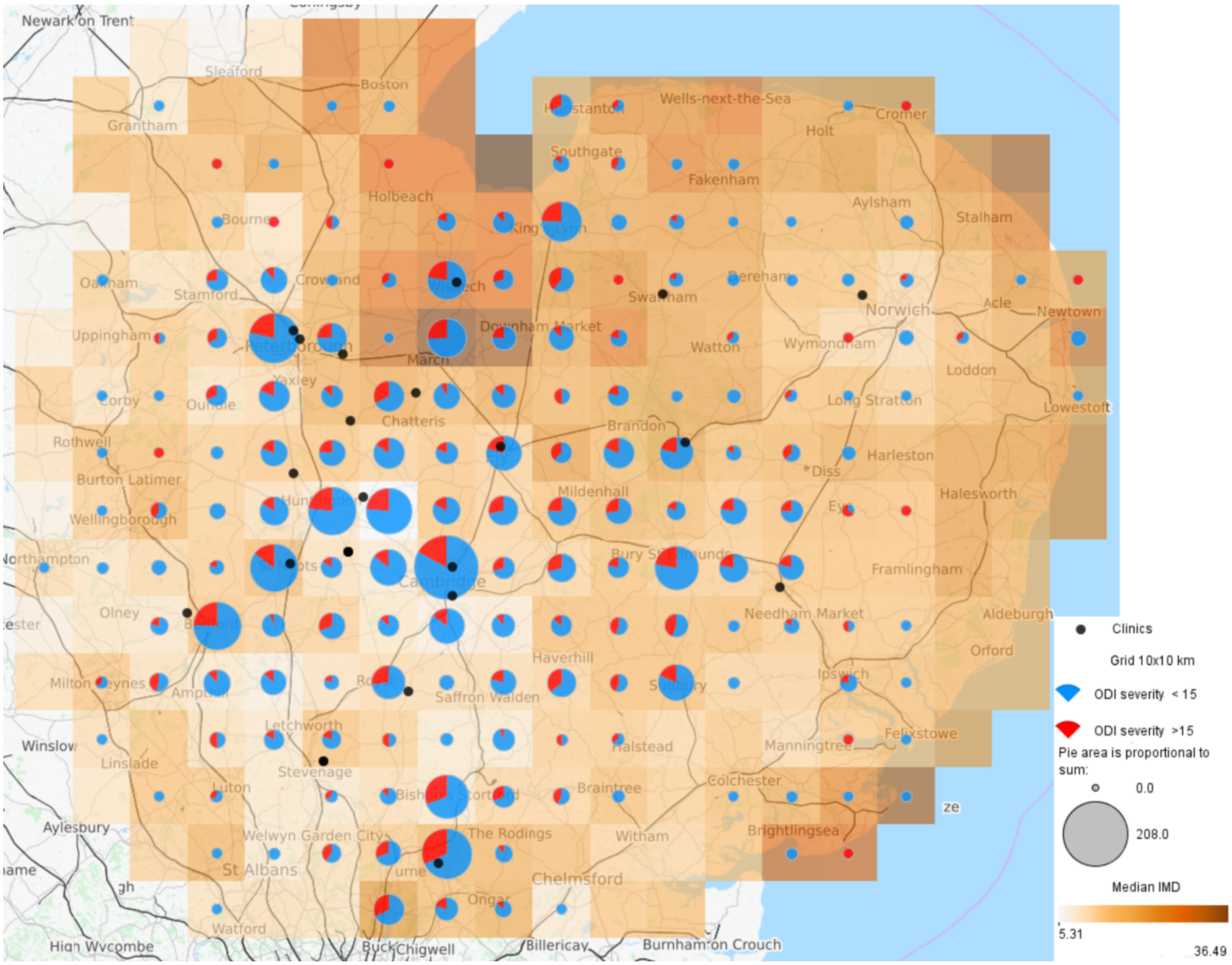


Figure 8: A map designed for the experiments based on the PAP pilot (revised version).

3.3.1 Instructions for participants.

The following instructions were provided to the test participants:

1. Step 1: Task description

Your task will be to explore the spatial distribution of people who came to clinics for testing the existence and severity of symptoms of the obstructive sleep apnoea condition (OSA), which is a potentially serious health disorder. Tests are done using special devices given to patients for taking home. A person who wishes to be

tested gets an appointment to one of multiple sites (clinics) distributing such devices. A test requires two journeys to a site, first for picking up a device and then for bringing it back.
Medical researchers had a hypothesis that the distance to a clinic might affect the inclination of people to undergo a test. People living far away from the distribution sites may be unwilling to spend their time and money travelling when they do not feel really serious problems. However, this hypothesis was not confirmed: no correlation was found between the distance to the clinic and the severity of the patient's condition according to both the objective result of the test and the patient's subjective estimation.
The map you will see on the next page is intended for investigating whether the patients' behaviours could be related to the conditions in which they live. The living conditions are represented by the values of the Index of Multiple Deprivation (IMD), which is the official measure of relative deprivation for small areas (or neighbourhoods) in England. This is an integrated measure accounting for the income, employment, education, health, crime, and other aspects of people's living conditions. **Higher IMD scores correspond to worse conditions.**
The spatial distribution of the IMD scores over the study area is shown by background painting. On top of it, diagrams represent in an aggregated form the spatial distribution of all people who were tested and, among them, those who had the **OSA severity level 15 or higher**, which requires medical treatment.

2. Step 2: Familiarization with the map

Please get familiar with this map!
The data in the map are aggregated by a grid with cell sizes 10x10 km. The background painting in shades of brown shows the deprivation score. **Darker shades represent higher deprivation**, that is, **worse living conditions**.
The black dots show the locations of the clinics where the OSA tests were performed.
The pie charts show the spatial distribution of the homes of the tested people. The **pie size** is proportional to the **number of tested individuals** living in a cell. The **red sector** in each pie represents the proportion of the individuals whose **OSA severity level** was determined as **15 or higher**, which requires medical treatment. The light blue part of a pie represents the count of the remaining people, who had the OSA level below 15.

3. Step 3: Visual analysis task

Your task is to examine 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. 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.

3.4 Commonalities and differences between the pilot-specific experiment designs

For all three series of experiments, we designed maps using similar visualization techniques, so that participants could spend less time on learning the visual encoding and focus more on the data exploration. The maps differ in the level of complexity of the information contained. For VFI, the map shows distribution of two phenomena: traffic (choropleth map) and driving events (proportionally sized circles). The map for SIS is more complex as it shows information about events of 3 different types; hence, there are 4 distributions to investigate (crashes and 3 event types) plus the joint distribution of all three event types. The choropleth map technique is used to show the distribution of the crashes, and the distribution of the events is represented by proportionally sized pie charts divided into sectors corresponding to the three event types. Hence, there are two visualization techniques, as in the VFI map, but one of them is more complex for SIS (pie charts) than for VFI (simple circles). For PAP, the map contains information about the distribution of a background phenomenon (deprivation level) and two levels of the OSA condition plus the total numbers of the patients. In addition, the map shows the distribution of the clinics. So, the PAP map involves more information layers (three) and uses more representation techniques (choropleth map, pie charts, and dots) than the other maps. The differences in the map complexity and, respectively, the task complexity suggest conducting the experiments in the sequence VFI – SIS – PAP, so that the complexity level increases as the participants get more familiar and experienced with the visual representations and the tasks.

The three analysis scenarios also differ regarding the relationships between the phenomena. In the VFI scenario, it is expected that the distribution of the driving events corresponds to the distribution of the traffic intensity, i.e., the two phenomena are correlated (null hypothesis). The map generally confirms this expectation, while there are local deviations. Accordingly, the participants are expected to note the global correlation and search for local deviations. In the SIS scenario, common sense also suggests that the phenomena (crashes and driving events) may be correlated. Hence, participants may try to find confirmations for this expected relationship. However, the map does not provide sufficiently strong evidence for this. The participants are expected to note "surprising" patterns contradicting to their initial hypothesis. In the PAP scenario, there may be no particular prior expectations concerning the relationships between the OSA data and the deprivation, or expectations may be of any kind. In the map, there is an obvious correlation between the distribution of the patients and the distribution of the clinics. This relationship is so prominent that it hinders seeing whether some relationships may exist between the OSA data and the deprivation. Accordingly, the participants can be expected to spend significant efforts for observing and comparing different areas of the map in order to conclude whether relationships exist or not. Hence, the analysis scenarios also differ in the difficulty of finding significant patterns and making inferences concerning the relationships between the phenomena. The VFI scenario appears to be the easiest, while the PAP scenario is the most challenging.

4 Eye-tracking data collection

Despite the unpredictable circumstances caused by the COVID'19 pandemics and corresponding restrictions of personal contacts, we have managed to recruit eight volunteers to participate in the experiments. Among them, six are professional data analysts (including four with PhD degrees), all having substantial experience in data science and visually driven data analysis. The remaining two are professional graphics designers working in industry and involved in a variety of projects that require data-driven visualizations. Three participants are females, five – males. The age of the participants ranges from 25 to 57, with the average around 40.

The data collection procedure was performed separately for each participant. Every participant was informed about the purpose of the study, read the experiment description and signed the informed consent form (see Appendices to this deliverable). The eye tracking equipment was demonstrated, and then 3 sessions were performed consecutively, in the order of increasing complexity: VFI – SIS – PAP. 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 a 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, in which the participants were supposed to characterize relationships between the phenomena shown on the map. 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.

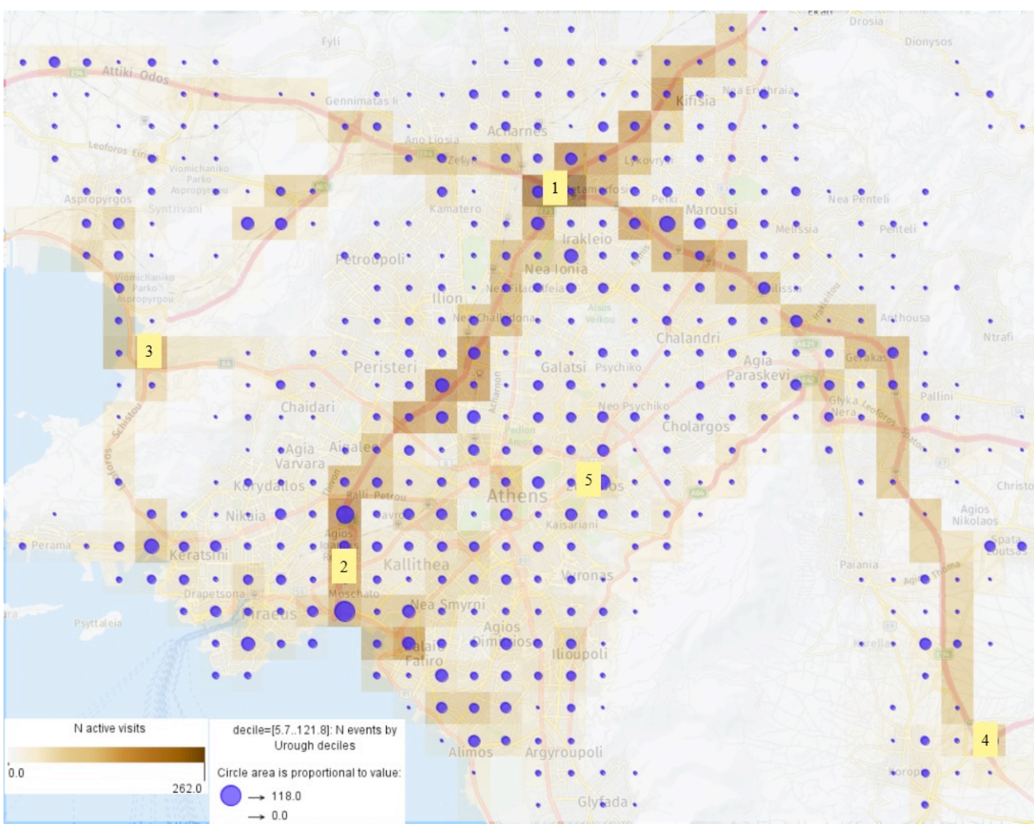


Figure 9: An example of a map with subject-marked locations of patterns (VFI pilot)

To provide a tangible result of the analysis, the participants were asked to mark on the map several patterns they deem important or interesting. For this purpose, a specially developed script displayed 5 numeric labels that could be dragged over the map and dropped on any location. The final states of the maps with labelled patterns were stored; an example is shown in Fig. 9.

The software suite that accompanies Tobii Pro X2 equipment contains tools for presenting scanpaths as trajectories on top of the stimulus screen (as, for example, in Fig. 10 and Fig. 11) and creating a density map of the scanpaths (as in Fig. 12), thus supporting “where” tasks (see Section 2.3.1). In addition, it is possible to extract fixations and present them as ordered positions on the map, thus supporting, to some extent, “detect” analysis tasks (section 2.3.6).

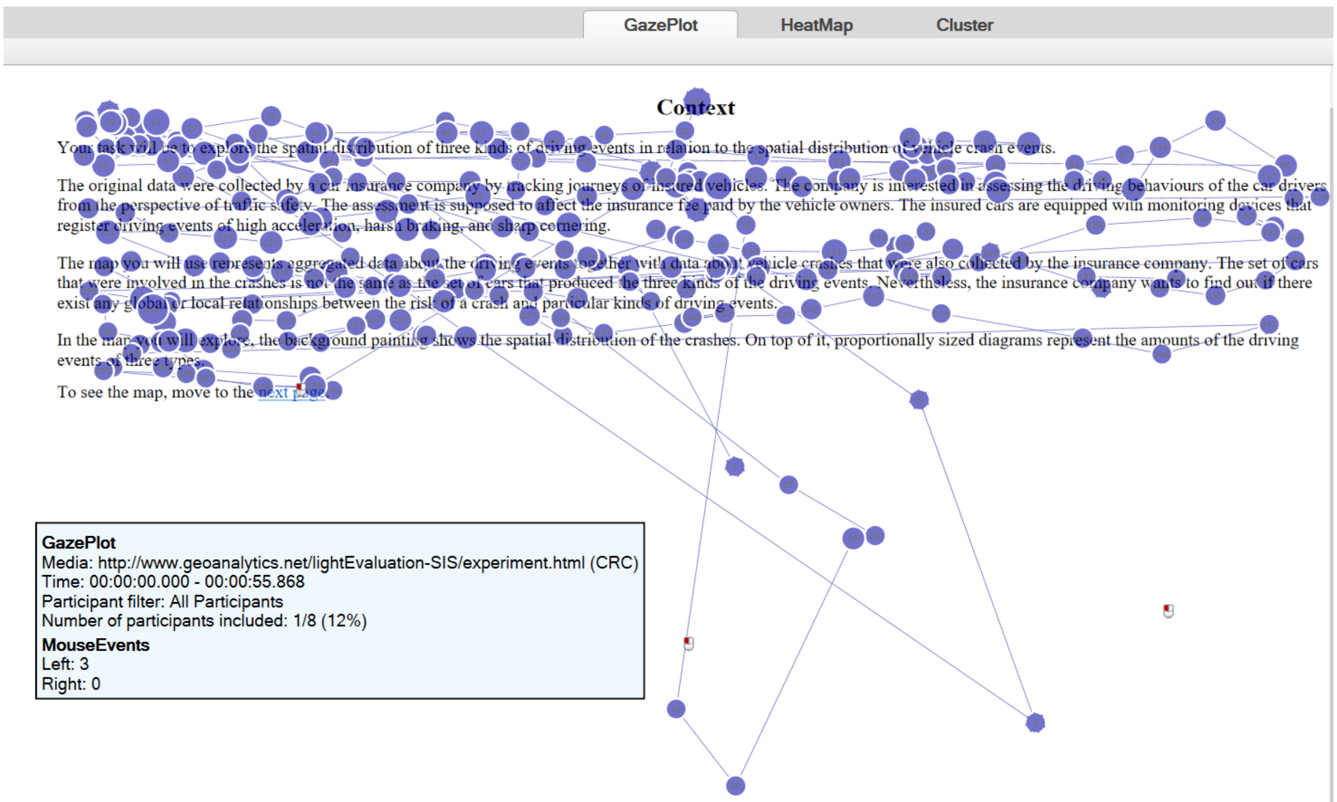


Figure 10: An example of a gaze plot illustrating how a subject reads a task description.

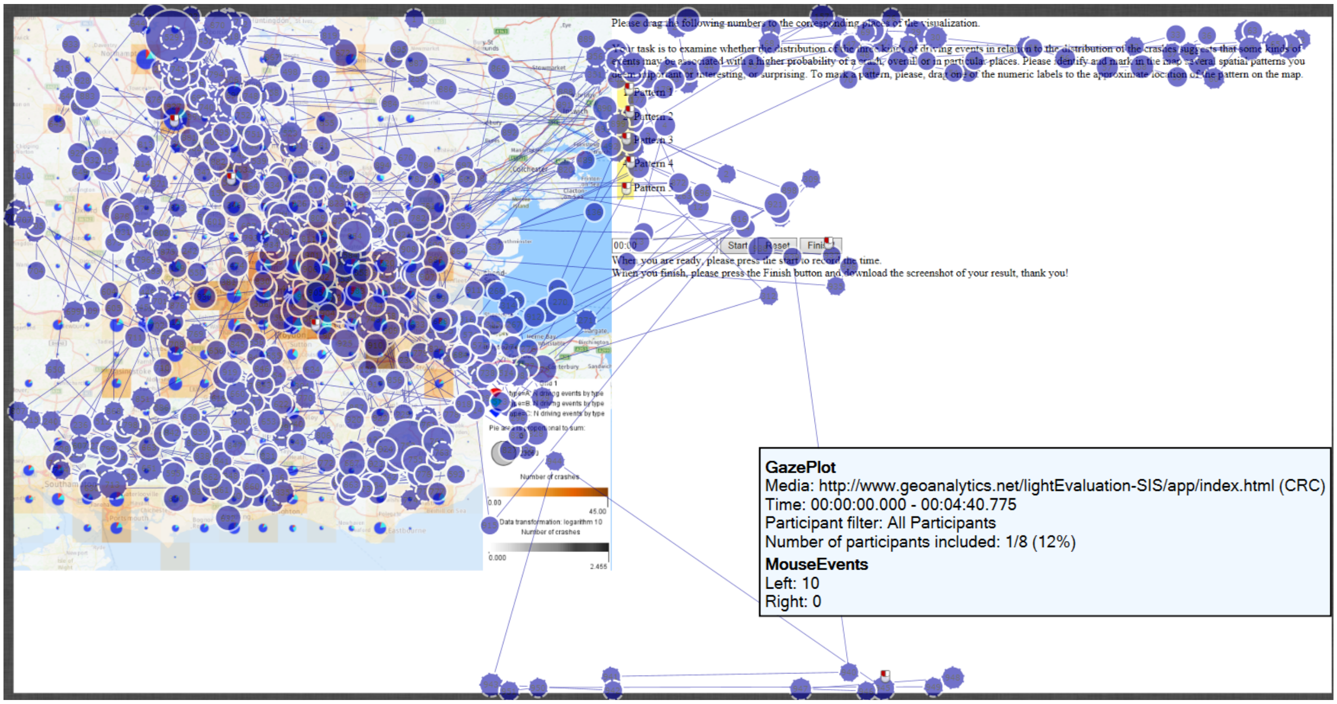


Figure 11: A gaze plot represents a sequence of fixations during a single data analysis session.

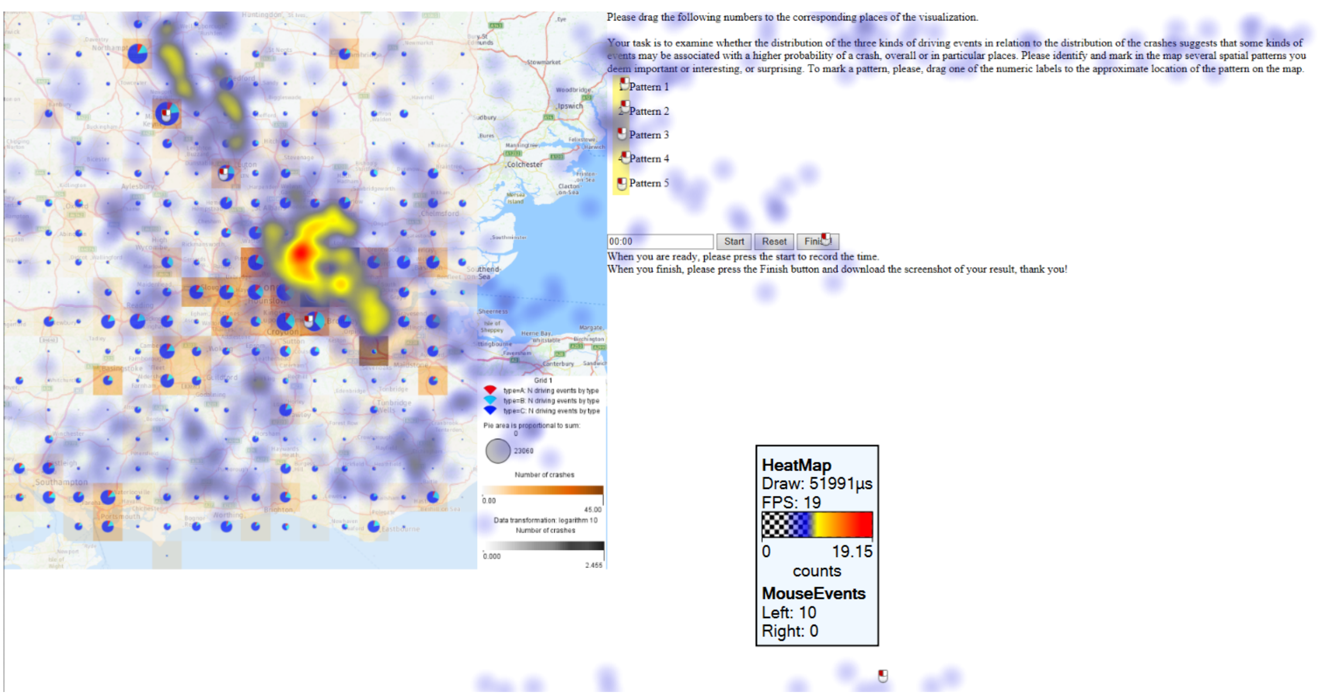


Figure 12: An example of an attention map for the same data analysis session.

Although these tools are useful for obtaining an overview of the collected data and checking the data quality, these techniques are not sufficient for a comprehensive analysis and understanding of analysts' behaviours and strategies.

To be able to analyse the data by means of more advanced methods, we exported the recorded data from the eye tracker software. The exported data for each scenario and participant consist of records containing time stamps, gaze positions at these time moments, and durations of the fixations. In essence, these are spatio-temporal data. Gaze fixations can be considered and analysed as spatial events, and gaze scanpaths can be considered and analysed as trajectories of moving entities. Such data can be analysed using, among others, some of the methods that have been developed in Track&Know (as described in the WP5 deliverable D5.1 and in the WP4 deliverables), or the methods applied in the project for analysis of vehicle trajectories and events occurring in the geographical space. In particular, V-Analytics software[3] that was used for data exploration and various analyses in all 3 pilots proved to be suitable for analysis of the eye-tracking data.

The visual analysis using the functions of V-Analytics is described in the next section. The application of machine learning approaches from WP4 is described in Section 6.

[3] V-Analytics research prototype: http://geoanalytics.net/V-Analytics/

5 Visual analysis of eye-tracking data

5.1 General approaches

A traditionally used technique in eye-tracking data analysis is to identify areas of interest (AOI) and then transform scanpaths into transitions between AOIs. However, this approach is not suitable in our settings, as every participant may have his/her own set of AOIs, and the AOIs of different participants may overlap. Moreover, the AOIs of an individual participants may change along the course of the analysis: extend, shrink, or shift. Defining AOIs as areas with fixed, crisp boundaries would hide such changes.

For uncovering and understanding participants' strategies, it is important to analyse the changes of the gaze positions over time. Due to very limited opportunities for analysis provided by animated displays, spatio-temporal information needs to be represented in other ways. The approach we found suitable and effective is a combination of a coloured map display (Fig. 13, top) and a temporal display called "scarf plot" (Fig. 13, bottom).

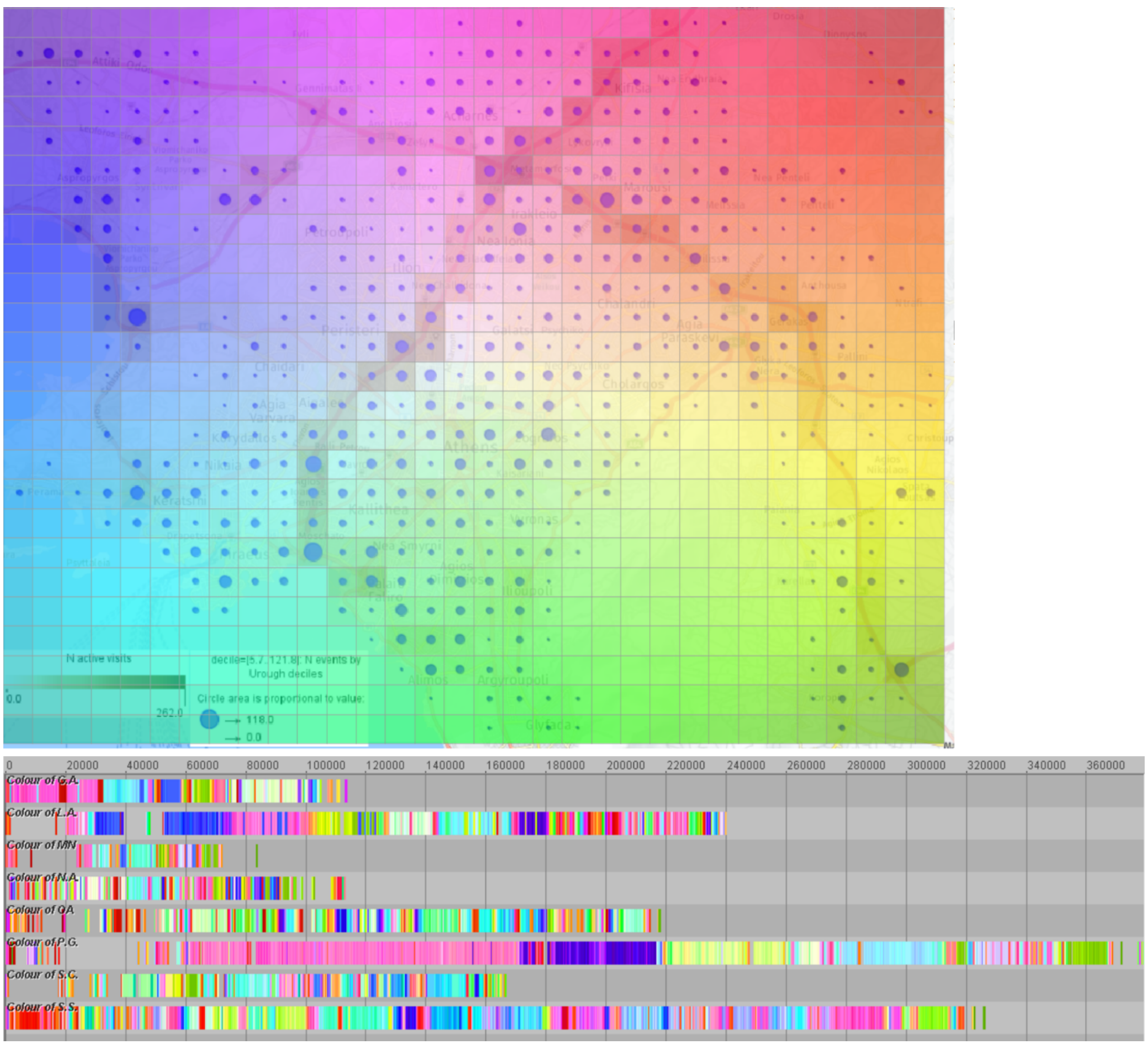
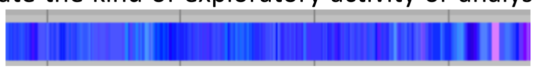
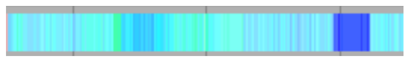
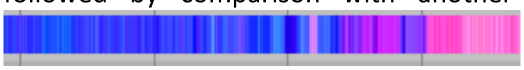
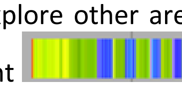


Figure 13: Colour-encoded (top) sequences of fixations (bottom) of all subjects (VFI pilot)

The whole map display is painted in gradually varying colour hues and shades, as shown in the upper part of Fig. 13. Hence, each position on the map gets its specific colour, and this colour can be used to represent this position in other displays, in particular, in a scarf plot display shown below the map in Fig. 13. The horizontal dimension of the display represents the session time. The gaze path of a participant is represented as a coloured “scarf”, i.e., a horizontal bar in which the colours correspond to the gaze positions at different times. Grey segments correspond to fixations outside the map. This happens, for example, when a participant reads the task description or takes one of the labels for dragging onto the map. Scarfs of multiple participants are arranged in the vertical dimension one below another.

Hence, the colours in the scarfs indicate, approximately, where the participants were looking and show how long they kept their attention around the same position on the map. Even more importantly, the variation of the colour along a scarf can indicate the kind of exploratory activity or analysis strategy of the participant. Thus, a scarf segment looking as this  may indicate focusing on some area and its surrounding, a segment like this  may tell about exploration of one area followed by comparison with another area distant from the first one, and a segment like this  may signify gradual shifting of the attention focus from one area to another. Since our major analysis goal is to find general patterns of exploratory behaviour and general, repeatedly applied strategies, the specific colours are not important but the character of the colour variation is the key to the information we are looking for.

To be sure that our interpretations of the colour variation along the scarfs are correct, we replayed the records of all sessions and all participants using the standard tools of the eye tracker. We watched animated representations of the gaze movements and fixation and listened what the participants were saying. We documented our observations and matched them to corresponding segments of the scarfs. For example, the segment  was matched to the following recorded observation of a participant’s activity: “Finds and marks an AOI; interprets the pattern (how the cars might behave in this area). Fixates on another area, compares it with the previous one, describes how it compares to the previous one. Does not place a marker, wants to explore other areas. Finds a place confirming the expected relationship, moves a marker”. The following segment  was matched to the observation record “Compares with place N1; says that the places have the same pattern”. The place N1 mentioned in this sentence is the place indicated by the blue colour shades.

In this way, we validated our interpretations of the different patterns of colour variation that we saw in the scarf plots. Based on these interpretations, it was possible to identify general behavioural patterns and strategies as reported at the end of this section. The following three subsections show some of the visualizations that were used in the visual analyses of the eye-tracking data concerning each of the pilot-oriented analysis scenarios.

5.2 Analysis of eye tracking sessions: Pilot 1 (VFI)

To check the consistency of the patters identified by multiple test subjects, we put all their markers on the same map:

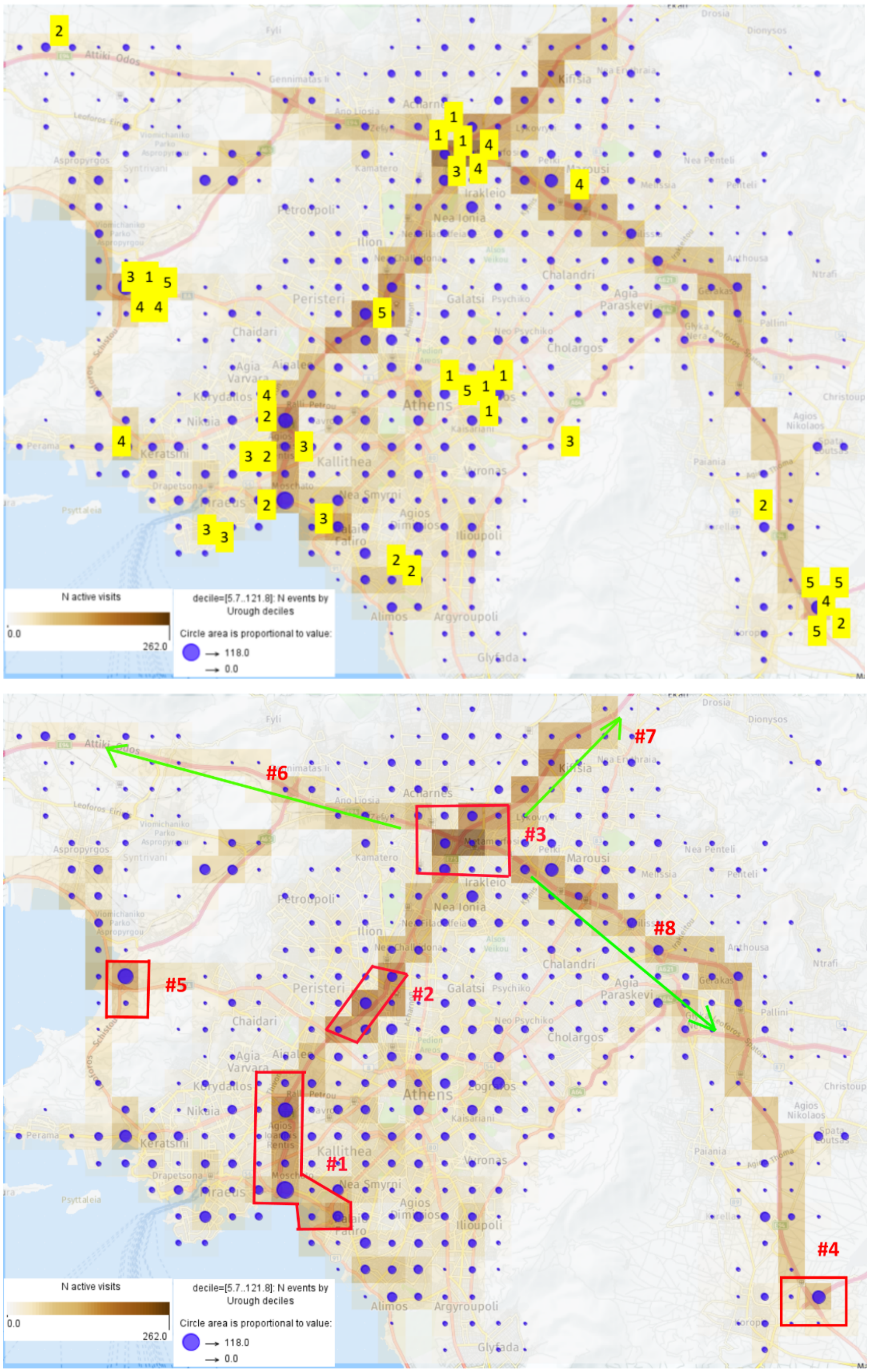


Figure 14: Top: Pattern markers of all test participants (VFI pilot). Bottom: Patterns marked by a domain expert prior to the experiments (copied from Fig. 3).

We can observe a strong consistency between the patterns identified by different participants (Fig. 14, top). These patterns also correspond to those marked by a domain expert at the stage of preparation to the experiments (Fig. 14, bottom and Fig. 3). The locations of the patterns correspond to dense concentrations of scanpaths and fixations, as shown in the next figures. It is evident that frequent very short scanpaths concentrate in areas of dense fixations, while medium range and long scanpaths are connecting these areas.

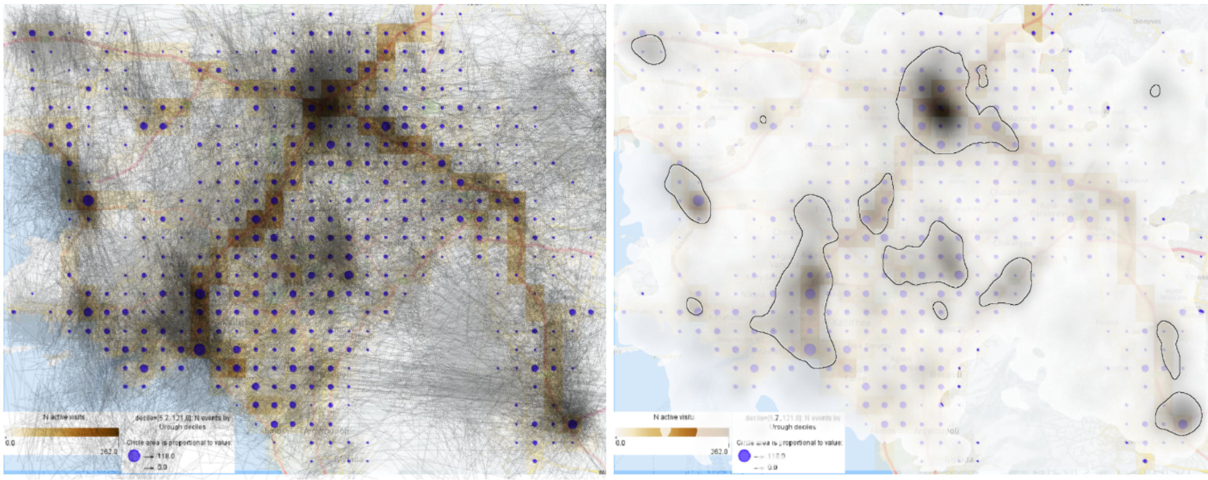


Figure 15: Scanpaths (left) and their density (right) for all subjects (VFI pilot)

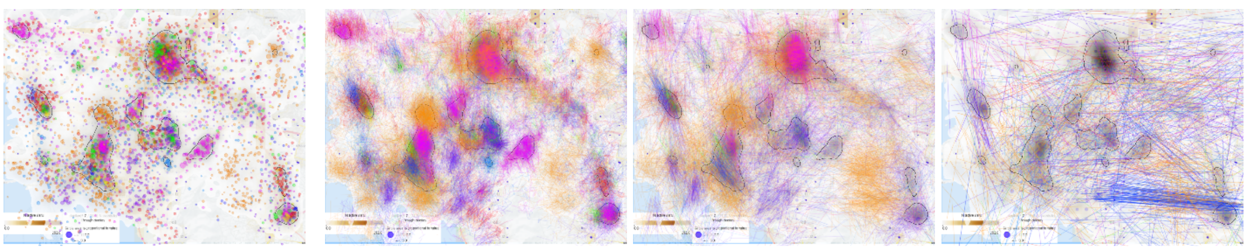
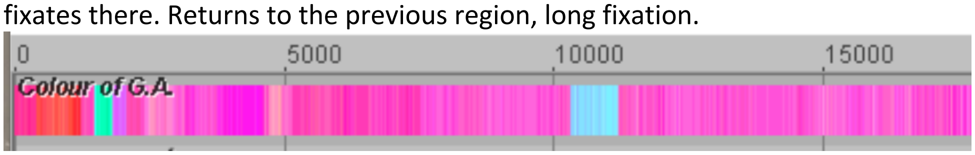


Figure 16: Fixations (left) and short, medium range and long scanpaths of all subjects (VFI pilot)

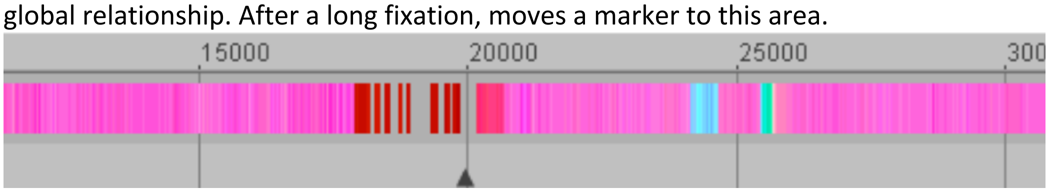
Generally, the patterns marked by the participants, as well as the main spots of their gaze fixations, are strongly associated with the street network of the underlying territory. This is logical, because the phenomena under analysis are related to traffic. As expected, participants noted the existence of a global correlation between the traffic intensity and the distribution of the driving events, and they also detected local deviations from the global patterns.

Here is an observation record of a representative analysis session of one of the participants abbreviated as GA:

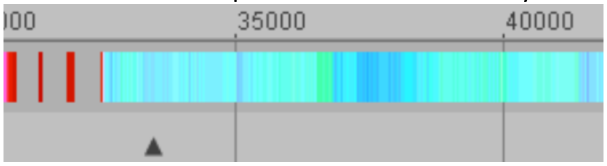
States the existence of a *global correlation*. Fixates within a certain region (mostly). Moves to another region, fixates there. Returns to the previous region, long fixation.



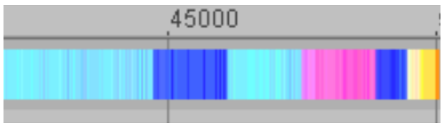
Says that the correlation is not straightforward: the AOI of the longest fixation contains a deviation from the global relationship. After a long fixation, moves a marker to this area.

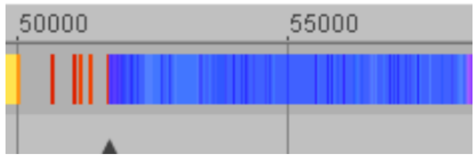


Fixates on another place and simultaneously moves a marker.

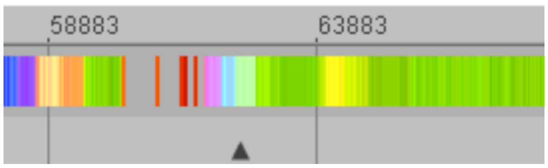


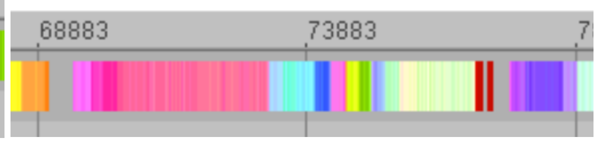
Says that this place differs from the previous one: "lower traffic but more events". Shortly fixates on a few other places, including the previously marked one. Fixates on the next place and simultaneously moves a marker. Comments it as a contrasting patterns disrupting the global correlation.



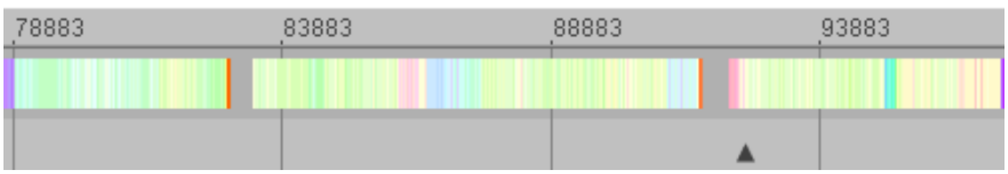


After short scanning, notes a place similar to the previous one. Says that there are other examples of this kind, looks at different areas.





Focuses on a relatively large region of low traffic.



<u>General observations</u>

- Pattern scales are diverse: global, local, regional
- Looked for deviations from the global pattern

5.3 Analysis of eye tracking sessions: Pilot 2 (SIS)

To check the consistency of the patters identified by multiple test subjects, we put all their markers on the same map:

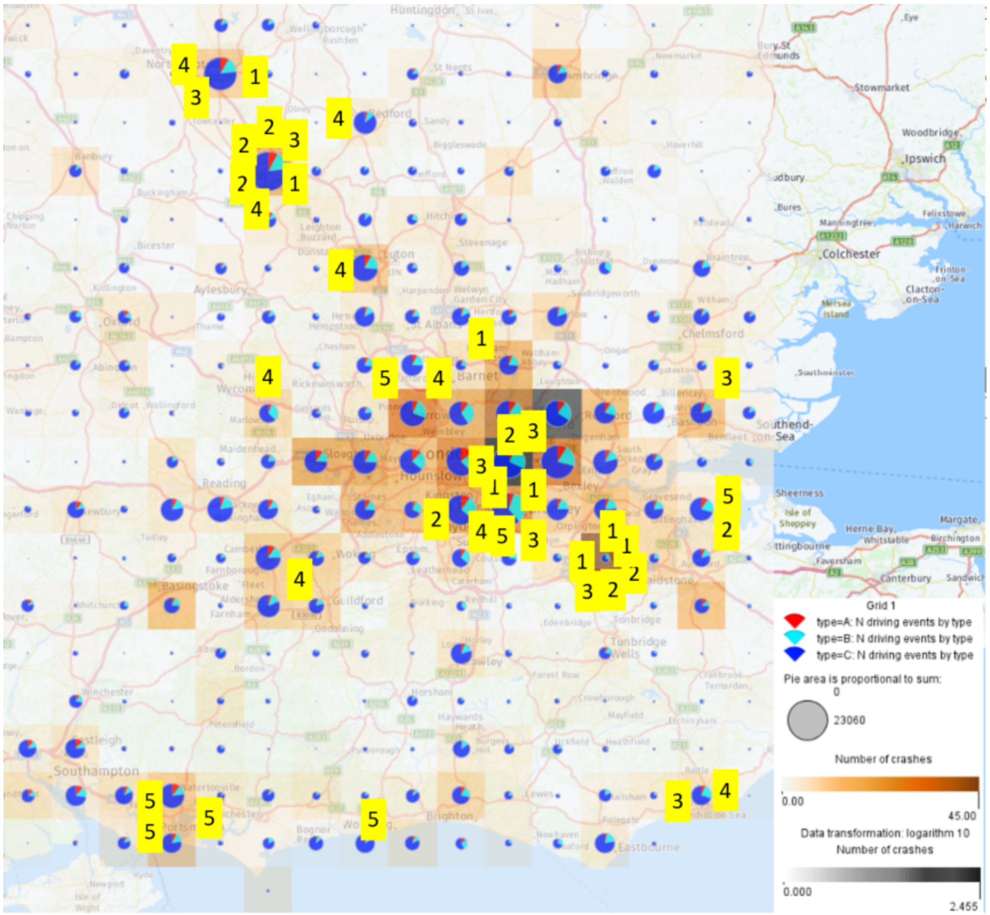


Figure 17: Pattern markers of all subjects (SIS pilot)

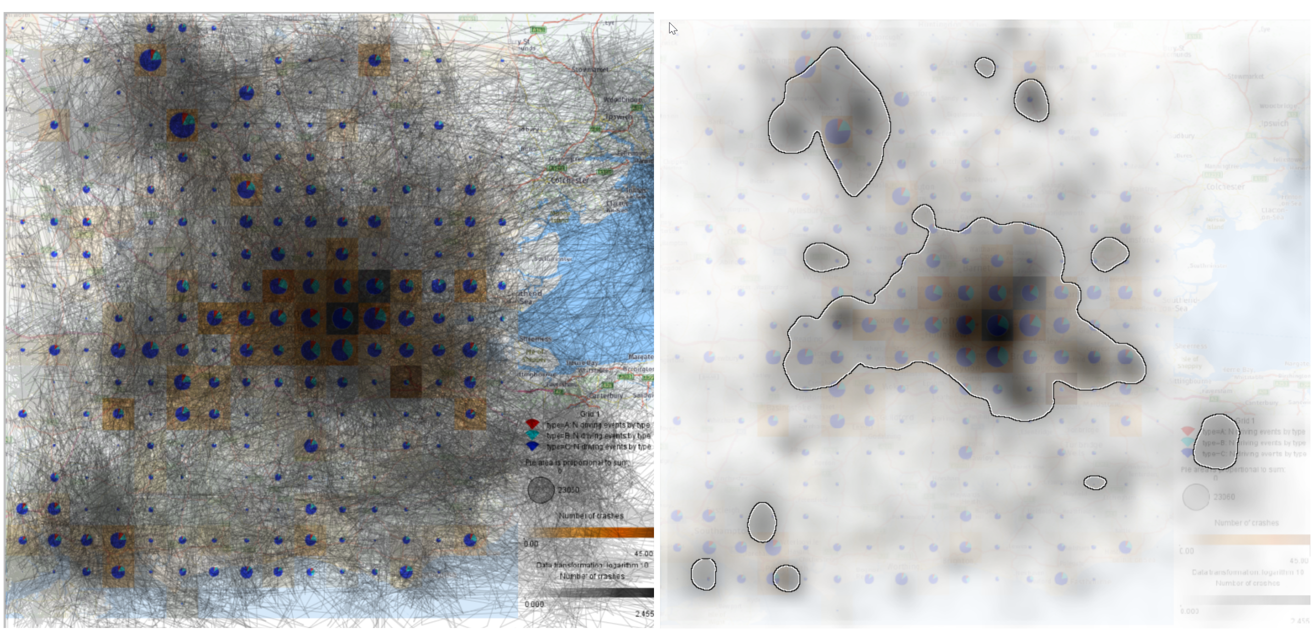


Figure 18: Scanpaths (left) and their density (right) for all subjects (SIS pilot)

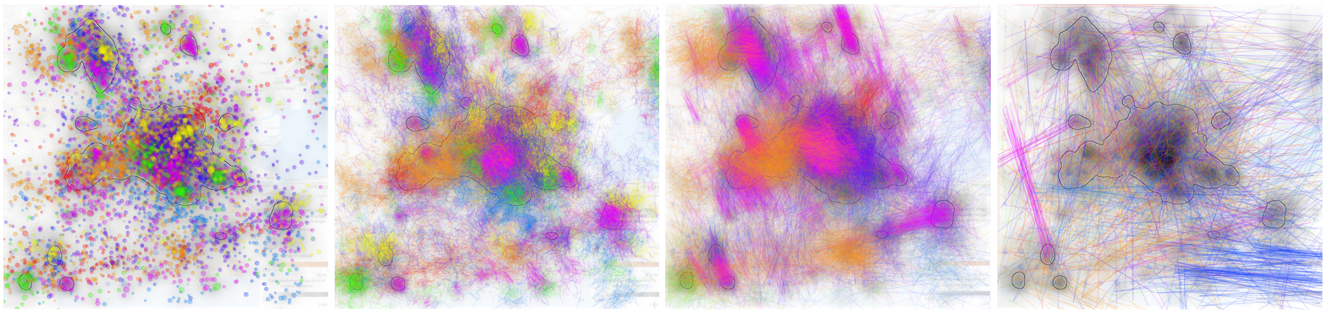
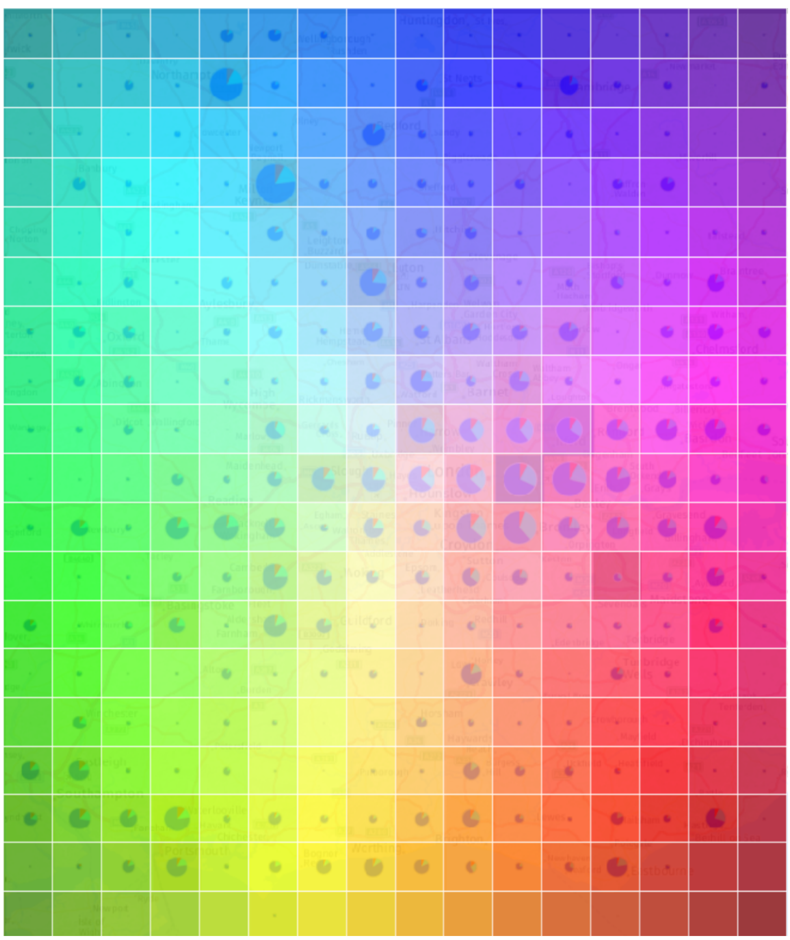


Figure 19: Fixations (left) and short, medium range and long scanpaths of all subjects (SIS pilot)

The consistency between the patterns identified by different participants is quite high. All participants put some of their pattern markers on the area of London, and almost all participants marked the places with high total numbers of events in the north-western part of the territory. In the other parts of the map, the consistency between the participants is lower. As in the VFI case, the placements of the pattern markers correspond to the density of gaze fixations and gaze movements of short and medium ranges.



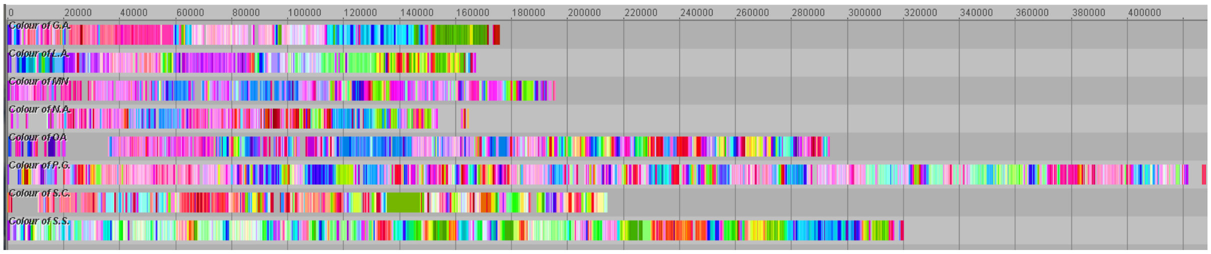
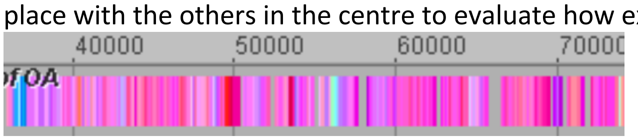


Figure 20: Colour-encoded (top) sequences of fixations (bottom) of all subjects (SIS pilot)

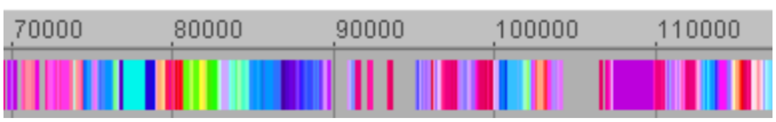
As mentioned at the end of Section 3.4, the relationship between the phenomena in this scenario is not as obvious as in the VFI scenario. The participants need to inspect and compare many different places on the map to be able to conclude whether a global relationship exists or not. Since there is no conclusive evidence of a global correlation, several participants tended to identify predominantly local, small-scale patterns. Some participants concluded that an overall correlation exists, with some exceptions, others tended to conclude that there is no global correlation.

Here is an observation record of a representative analysis session of one of the participants abbreviated as OA:

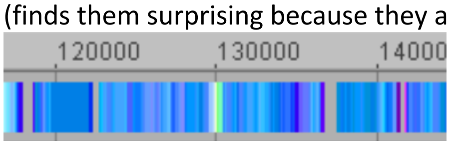
Explores at the central area: “as expected, there are more crashes and also more driving events, nothing surprising”. Notices a place near the centre with many crashes but almost no driving events, which is *surprising*. Marks the place. While talking about the exceptional place, continues looking around, apparently, comparing the place with the others in the centre to evaluate how exceptional it is.



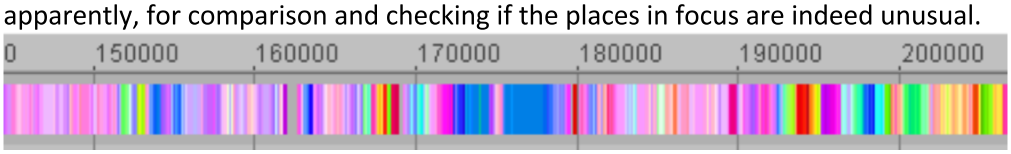
Scans the map (different parts), consults the legend, clarifies the colour encoding of the event types. Says that cornering seems to be the most popular event in London, acceleration is not that much. Braking is also popular. “Do they have so many sharp curves in London?” Long eye movements, often beyond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



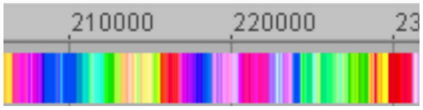
Focuses on the northern part of the map. Finds several places with much more events and accidents than around (finds them surprising because they are not in the centre), marks 3 places.



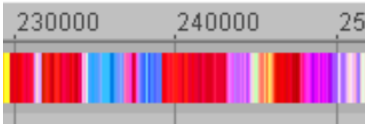
Investigates the central area again, looking for proportions of different event types. Finds a sub-area with more braking events than in other places, marks the area. While describing and marking, looks also at other places, apparently, for comparison and checking if the places in focus are indeed unusual.



Long map scanning. Apparently, OA tries to understand the distribution of the acceleration events, because at the end OA says: “It seems that there is not enough space to accelerate”. It seems that assessing different places during the scanning takes more time than usually during browsing. Apparently, OA investigates the structures of the diagrams and the proportions of the event types.



Notes a specific place with many braking events on the southeast and moves marker N4 from its previous location to the new place. Says that it has higher proportion of braking than anywhere else but not many accidents. Eye movements indicate comparison of the POI with other places.



Scans the map, says "In general, high N of events correlates with the N of accidents" (*global relationship*) "except for the coastal area, where more events do not produce many crashes". While talking about the coastal area, actually looks at other places in the map.

Lastly, scans the map but does not find other notable patterns.

The following general observations have been made about this analysis session:

- Finds regional patterns (London, sub-area of London, coastal area), local anomalies, re-occurring local patterns. Comes to a conclusion about the existence of a global relationship, with exceptions.
- Strives to interpret and explain patterns in terms of geo background and driving behaviours.

5.4 Analysis of eye tracking sessions: Pilot 3 (PAP)

To check the consistency of the patters identified by multiple test subjects, we put all their markers on the same map:

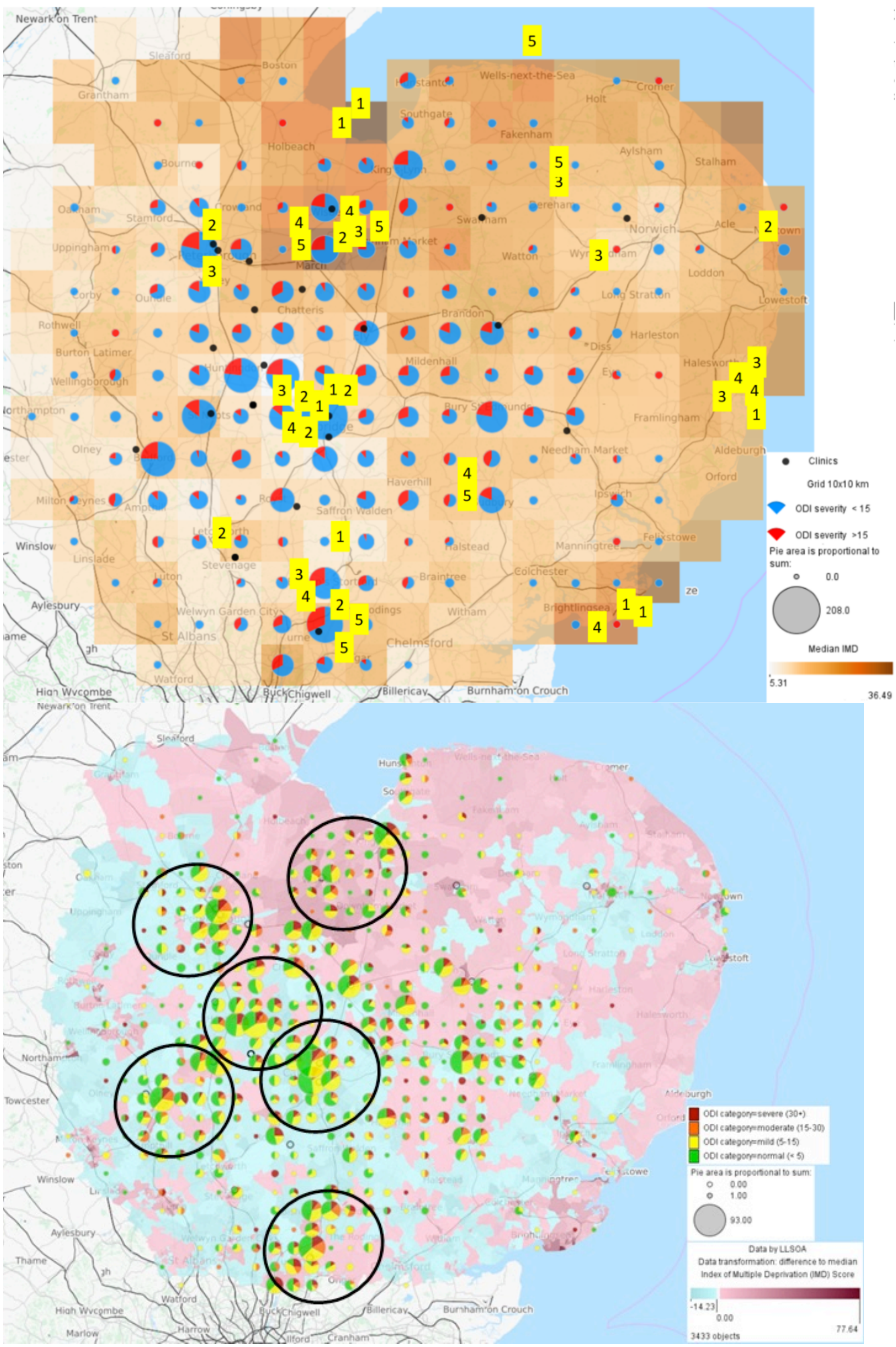


Figure 21: Top: Pattern markers of all test participants (PAP pilot). Bottom: patterns previously marked by domain experts based on a map with a different representation of the same data (see Fig. 7).

There is lower consistency between the marker placements of different participants than in the VFI and SIS scenarios. This corresponds to the lower certainty concerning the existence of relationships between the OSA data and the population deprivation, as mentioned at the end of Section 3.4. All participants placed their markers on the areas around Cambridge and to the north and to the south of it (Fig. 21, top). This corresponds to the patterns previously identified by the domain experts (Fig. 21, bottom). Some participants also noted patterns of absence or low number of patients in the eastern part of the territory, particularly, in coastal areas with high deprivation (Fig. 21, top). As expected, the obvious correlation of the distribution of the patients with the distribution of the clinics interfered with the analysis of the relationships between the patient data and the deprivation.

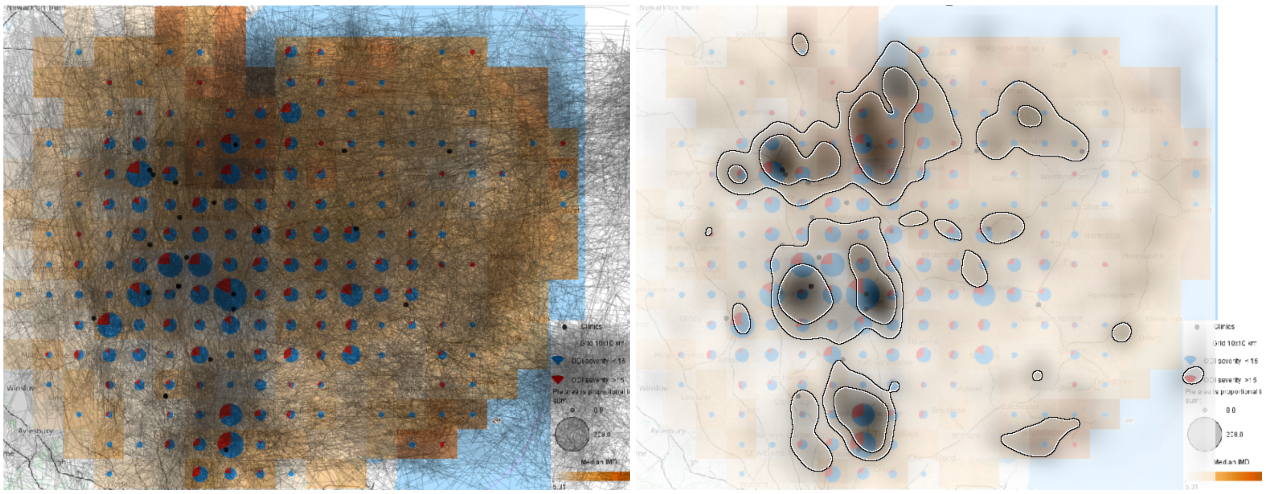


Figure 22: Scanpaths (left) and their density (right) for all subjects (PAP pilot)

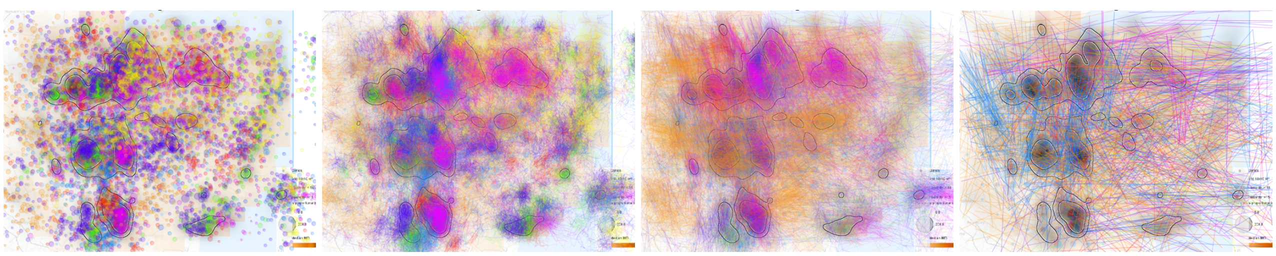


Figure 23: Fixations (left) and short, medium range and long scanpaths of all subjects (PAP pilot)

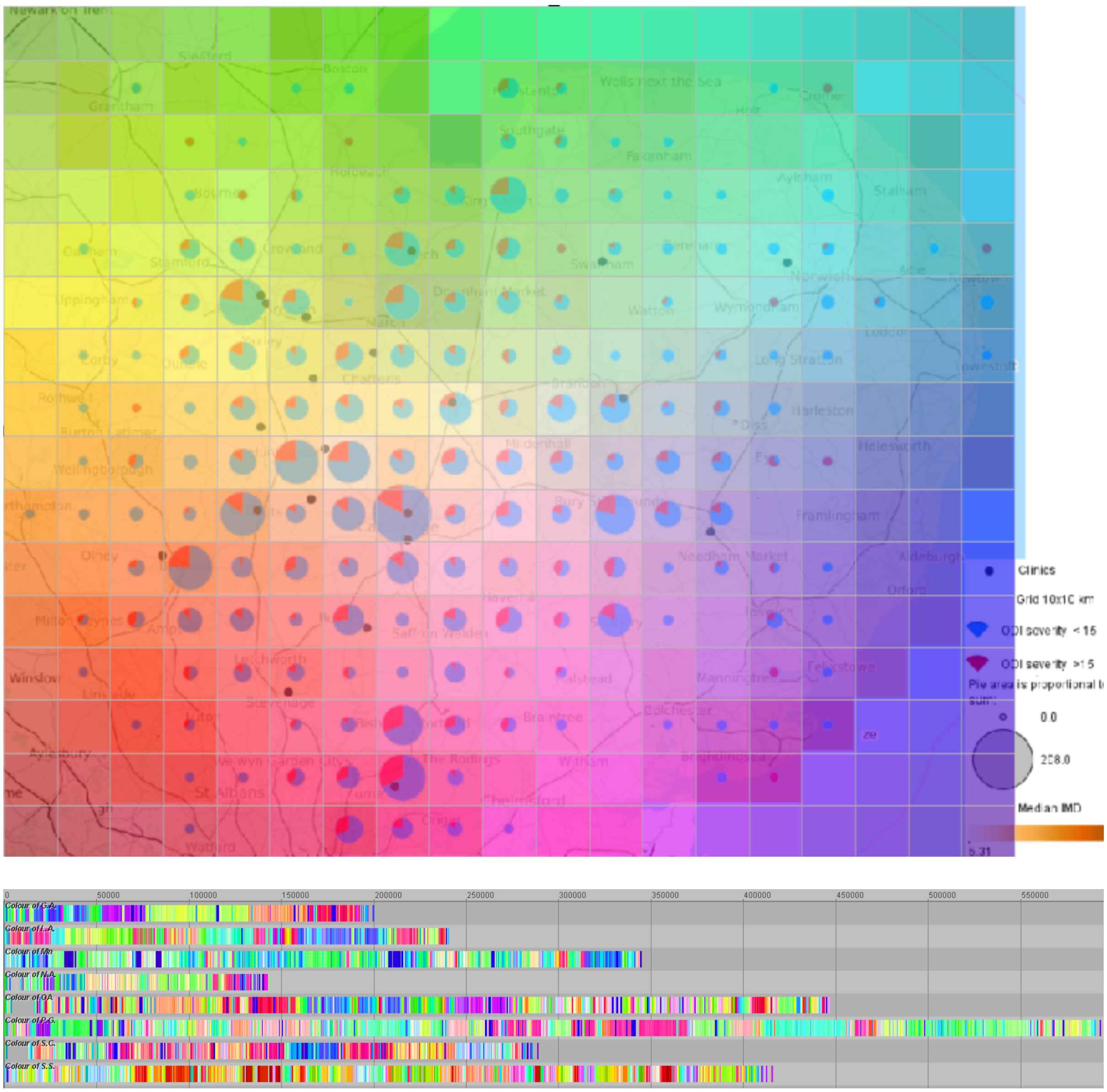
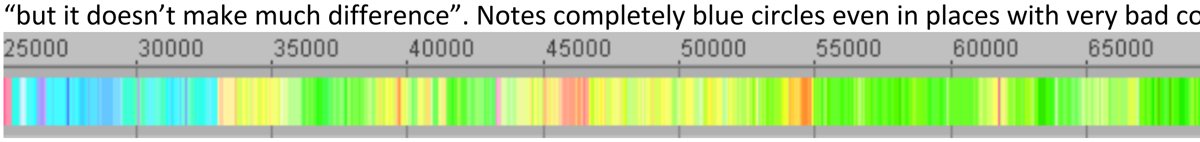


Figure 24: Colour-encoded (top) sequences of fixations (bottom) of all subjects (PAP pilot)

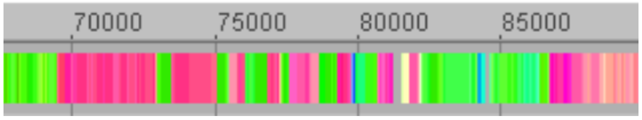
Similarly to the previous pilots, we analysed each session separately, aiming at detection of repeatedly occurring patterns. Here is a representative session of a single participant abbreviated as L.A.:

After quite short scanning, notes a place with low deprivation and absence of severe condition, interprets that people come for check just because they can spend time and money. Marks the place N1.

Explores the north-eastern part, then north-central, notes high N of clinics around an area of high deprivation "but it doesn't make much difference". Notes completely blue circles even in places with very bad conditions.

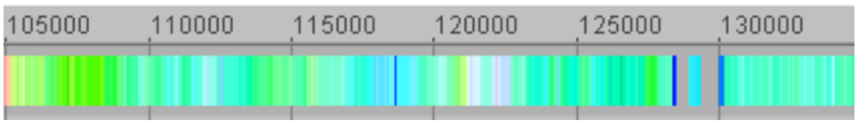


Suddenly moves from the north to the south. Notes a similarity between the symbols on the north with high deprivation and on the south with low deprivation.



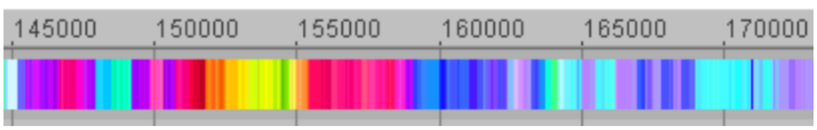
Focuses on the central area. “It is a particular thing: when you are living next to a hospital, you would rather go to check your health condition”. Points at several places in the centre, puts a marker (N2) close to them.

Explores the northern and north-eastern part. “I am still doubting about these small blue dots. People do not feel that bad, and they live in rather poor areas, but they still go for check”. Does not understand the reason, deems that additional data might clarify this. Marks an area with small blue dots on the northeast (N3).

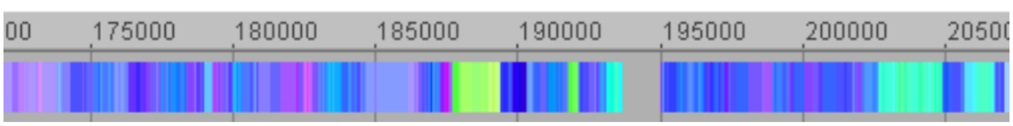


Says that the pattern in area N3 reminds that in area N1: also blue dots, but the living conditions are better than in N3.

Explores the eastern (coastal) part with mostly high deprivation, then moves through the south to the western part, then returns back to the south. Then again to the east and northeast.



Focuses on an area where there is a clinic but people from some places close to it do not come at all. Puts marker N4 in the area with no patients, says that further investigation with additional data is needed. Notes a similarity to area N3.



Revisits the areas on the north and on the south. Says that it is logical that in the wealthy area (on the south) people would go for a test even feeling not too bad. However, in the area on the north, where the deprivation is high, the N of people coming for a test is nevertheless quite high. Marks the area on the north N5, then continues comparing it to wealthier areas.

The following general observations were made about this session:

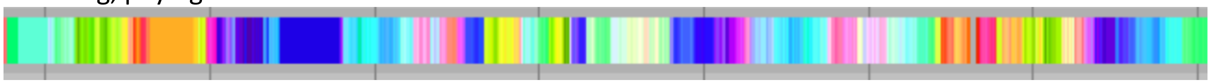
- Many comparisons of wealthy and poor areas.
- Many revisits of previously explored areas.

5.5 Frequent patterns and their interpretations

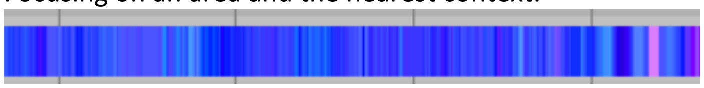
Here are examples of interpreted patterns that occur repeatedly in multiple sessions for different pilots:

Scanning, intermittent looking in the task description:

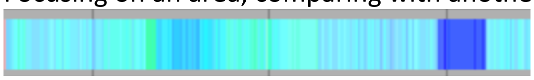
Scanning, paying more attention to some areas:



Focusing on an area and the nearest context:

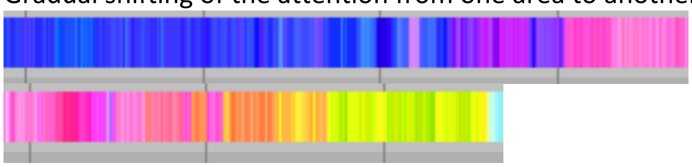


Focusing on an area, comparing with another (distant) area:

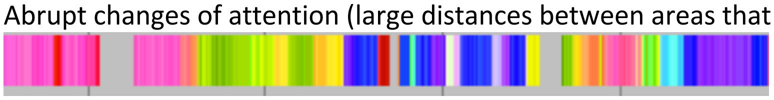


Changing the focus, assessing several candidate areas, focusing on one area and its nearest context:

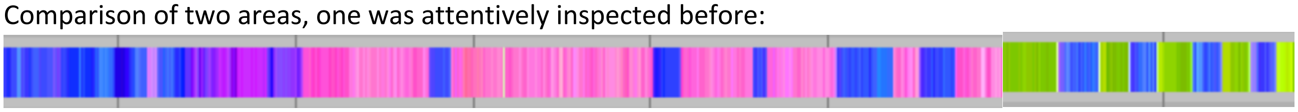
Gradual shifting of the attention from one area to another:



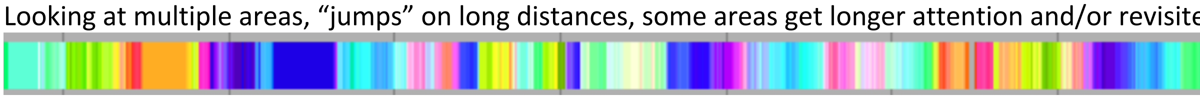
Abrupt changes of attention (large distances between areas that get attention):



Comparison of two areas, one was attentively inspected before:

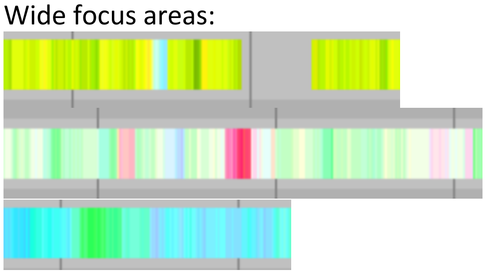


Looking at multiple areas, “jumps” on long distances, some areas get longer attention and/or revisited:

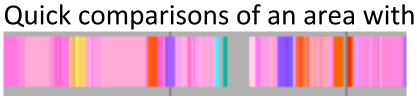


Narrow focus area:

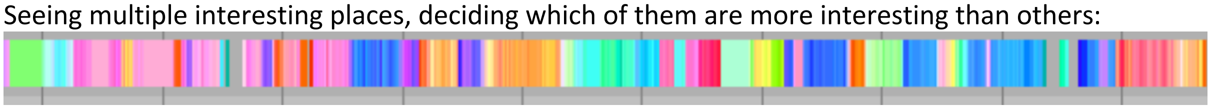
Wide focus areas:



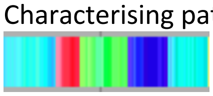
Quick comparisons of an area with many others:



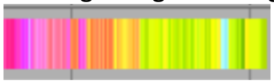
Seeing multiple interesting places, deciding which of them are more interesting than others:



Characterising patterns in several places:



Moving along a linear geo feature (a road):



5.6 Generalization of the observations and visual analysis results

After analysing systematically all 8x3=24 sessions, we have generalised and systemised our observations and findings concerning analysts' behaviours. The outcomes of this activity are presented in the following.

5.6.1 Characteristics of studied phenomena that affect analysis behaviours

Possible **scales** (**spatial extents**) of patterns or relationships that can exist in the visualized phenomena:

- 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).

In our experiments, analysts notice two major types of relationships:

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

We have identified that different characteristics of places and regions can motivate people to consider them as significant 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

5.6.2 Eye movement characteristics and patterns

Eye-movement saccades and sequences of fixations vary in their attention scope

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

We have observed the following types of movement patterns in the collected eye-tracking data:

- 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
- Assessing (e.g., assess potential interestingness of a location): relatively shot 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)

5.6.3 Cognitive tasks

The observed eye movement patterns may correspond to the following 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

Analysts need to translate these tasks into visual properties of locations in a display based on the way of data encoding. Some participants of our experiments (particularly, S.S.) verbalized this explicitly, e.g., “I search for big circles on light background or small circles on dark background”; “I search for places contrasting to the context”.

The following types of visual search targets have been observed:

- 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

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.

5.6.4 Exploration tasks

The following exploration tasks have been performed by the analysts at different scales:

- 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.

5.6.5 Strategies

In performing the exploration tasks, the analysts used different 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.

5.6.6 Individual tendencies and biases

In respect to pattern scales, we have observed 3 types of biases and tendencies:

- 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.

In respect to pattern diversity, we observed two different tendencies:

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

Data representation strongly affected the analysis strategies:

- 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

5.7 Conclusions

5.7.1 Implications for future intelligent implementations

Based on our findings, the following directions for potential support in future intelligent implementations can be proposed:

- Help to focus on larger regions
- Help to remember previously explored places/areas (mark them, index in terms of locations, durations and positions in a sequence, attributes etc., store their summary characteristics)
- Support efficient browsing of previously explored places according to their characteristics
- Help to compare new places/areas with previously explored
- Help to find similar places/areas
- Promote pattern diversity
- When there are diagrams or glyphs varying in size, support visibility of elements for small symbols.

5.7.2 Effectiveness of the visual analytics methods

Using the visual analytics methods, in particular, colour coding of positions in the visual stimuli and the use of the position colours in scarf plots, we were able to

- understand the exploratory activities of individual participants,
- involve abstraction in perception and interpretation of visual representations: consider the character of the changes of the attention focus over time irrespective of specific spatial positions;
- find repeated patterns of activities, which allowed us to make general inferences.

We can thus conclude that the visual analytics methods that we used proved to be very useful and can be recommended to other researchers analysing eye-tracking data.

6 Eye tracking individual mobility networks

In this section we describe an analysis of the eye-tracking dataset performed with a variant of Individual Mobility Networks (IMN), a tool developed in WP4 and discussed in deliverable D4.2.

6.1 General approach

IMNs are a model developed for representing mobility traces of vehicles over a territory. They are based on the key notion of "location", which is a place around which a vehicle stops one or more time. Finding such locations is inherently a spatial clustering task, which was originally implemented through an ad hoc procedure named Tosca. However, in experimenting with eye-tracking data, two main issues emerged: (i) the mobility of the eye is much more dynamic (sometimes even erratic) than vehicles on the map; (ii) the relevance of a location should take into consideration how long the eye rested on it, and not only the frequency of visiting. In order to take these two aspects into consideration, a variant of IMN, named "eye-IMN", was developed, which works as follows:

- Each point in the dataset, with its georeferenced position, is weighted based on the attribute "GazeDuration", representing the rest time on that location
- The Tosca location extraction process in IMN is replaced with a weighted K-means clustering
- The resulting clusters are associated with their cumulative GazeDuration, and those that do not reach a minimum value (2 seconds, in the experiments shown below) are discarded
- The locations are used as input for the normal IMN construction process, considering each pair of consecutive points as a (short) trip.

6.2 Individual mobility networks

In the following figures, we show the resulting networks obtained for each participant over the three Pilots. The size of a node on a map represents the overall gaze duration, and edges represent direct movement between the locations connected.

Figures 25, 26, and 27 demonstrate the IMNs derived from the participants' eye tracks in the scenarios VFI, SIS, and PAP, respectively.

Figure 26 shows the IMNs obtained over the SIS Pilot for all eight users. We can immediately observe that all have some activity in the centre of the city, as expected, and most of them involve the top-left and the bottom-left area. The complexity of the IMNs is rather variable, with P.G. covering many different locations with a very high connectivity. Others, like G.A., show some smaller "islands" (e.g. on the top-left) that are isolated from the others areas.

Figure 27 represents the same information for the PAP Pilot. Here we can observe that all users have a more distributed attention over several locations, and the movements between nodes are more complex. Compared to the previous case, it appears to require a more intensive search activity from the user. The subjects that had simpler IMNs in the SIS case tend to behave similarly also on the PAP Pilot.

The VFI case shown in Figure 25 presents a slightly simple structure in all users, with less links and with more linear paths. Probably, the simpler geographical structure of the territory has an impact on how the task is carried out. Singularly, several users spend significant time on reading the text, enough to emerge in the IMN (something that did not happen with SIS and PAP). This is explainable, as the VFI scenario was the first in the sequence. As the tasks in all three scenarios were similar, the participants did not need to spend as much time reading the task descriptions in the SIS and PAP scenarios as in the VFI scenario.

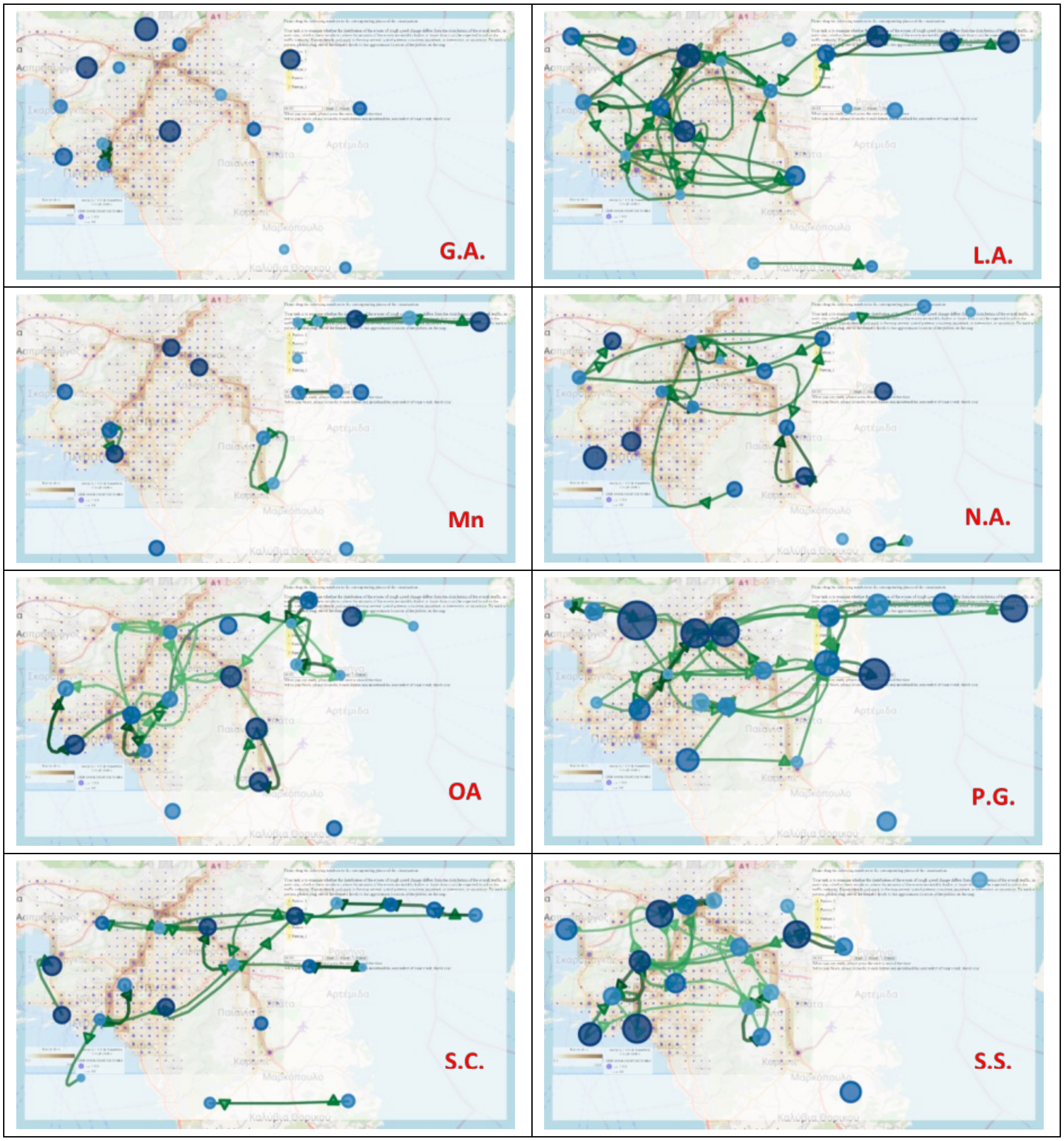


Figure 25: IMNs for the VFI Pilot.

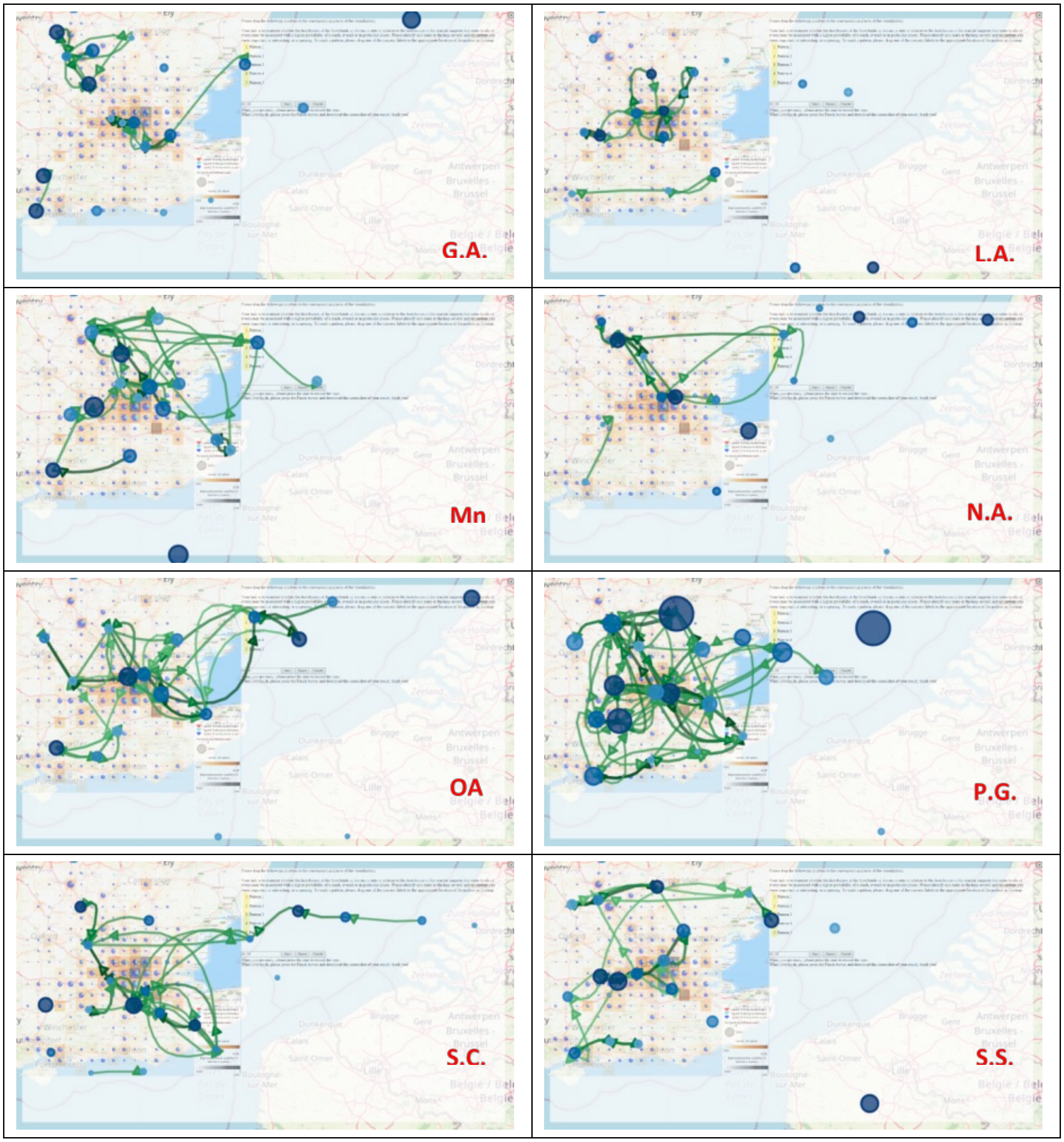


Figure 26: IMNs for the SIS Pilot.

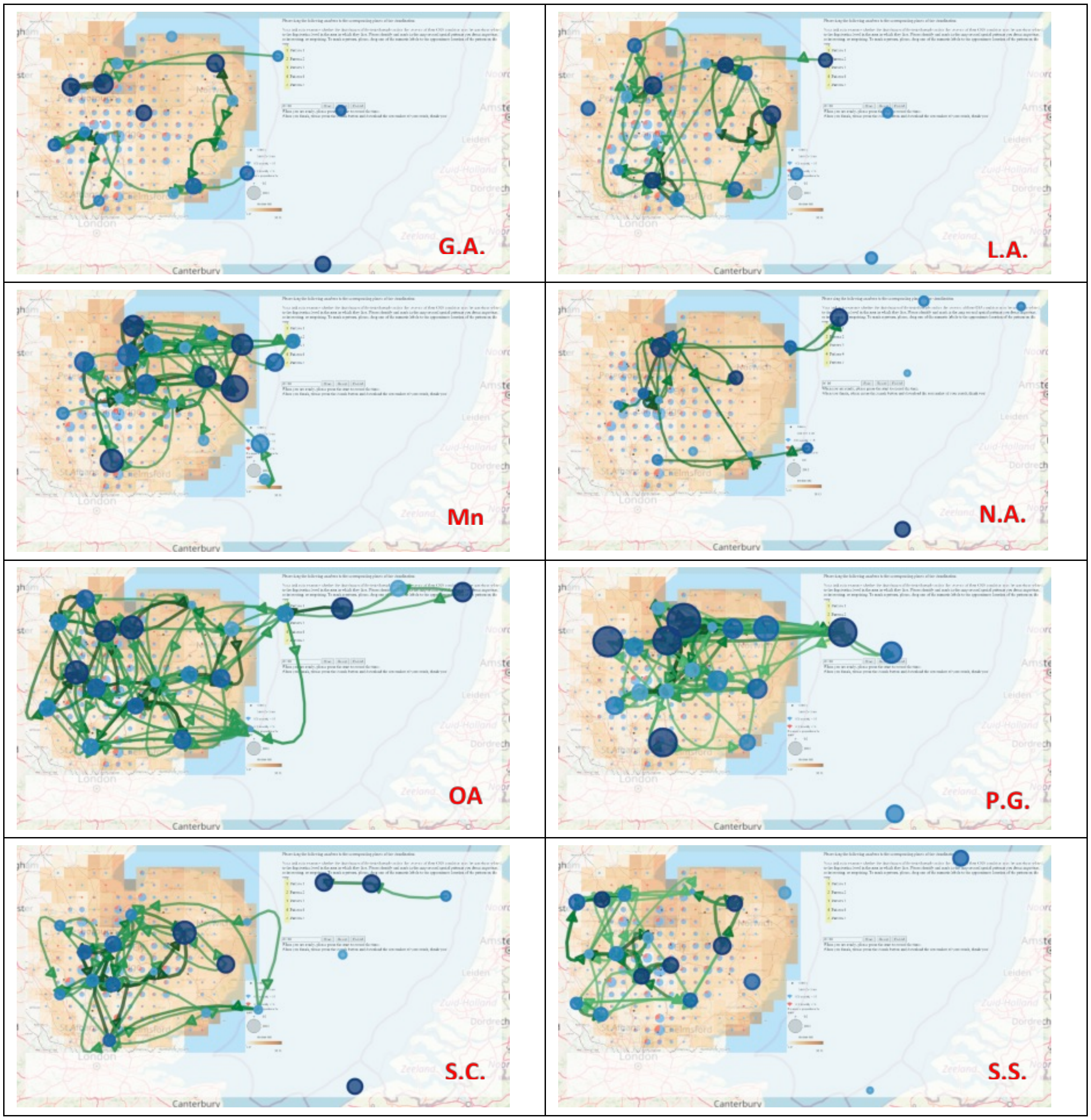


Figure 27: IMNs for the PAP Pilot.

6.3 Analysis of mobility graphs

This section presents a simple comparison of the IMNs obtained in terms of a few standard network indicators. In particular, we show a matrix of distributions for: degree centrality, betweenness centrality, closeness centrality, clustering coefficient and eigenvector centrality. The results for VFI, SIS, and PAP are shown, respectively, in Figure 29, Figure 30 and Figure 28.

A simple comparison on each separate Pilot suggests that each user tends to have distributions of values that are rather different from the others. The main exception is betweenness centrality, which appears to be rather similar for all users. Interestingly, a cross-Pilot comparison suggests instead that each user tends to create IMNs with quite similar characteristics, probably representing a general exploration strategy of the dashboards under analysis.

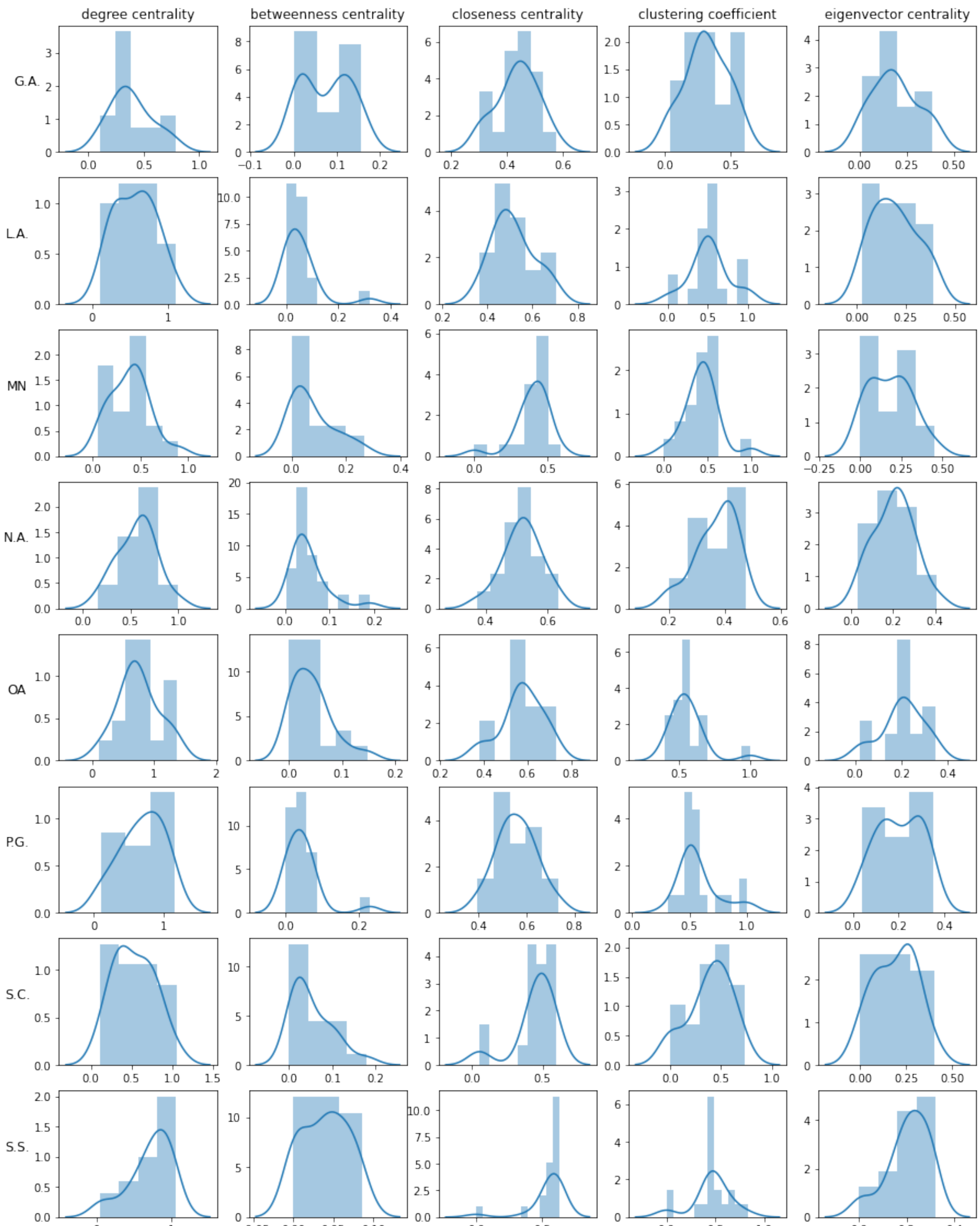


Figure 28: Distributions of IMN network indicators for VFI data.

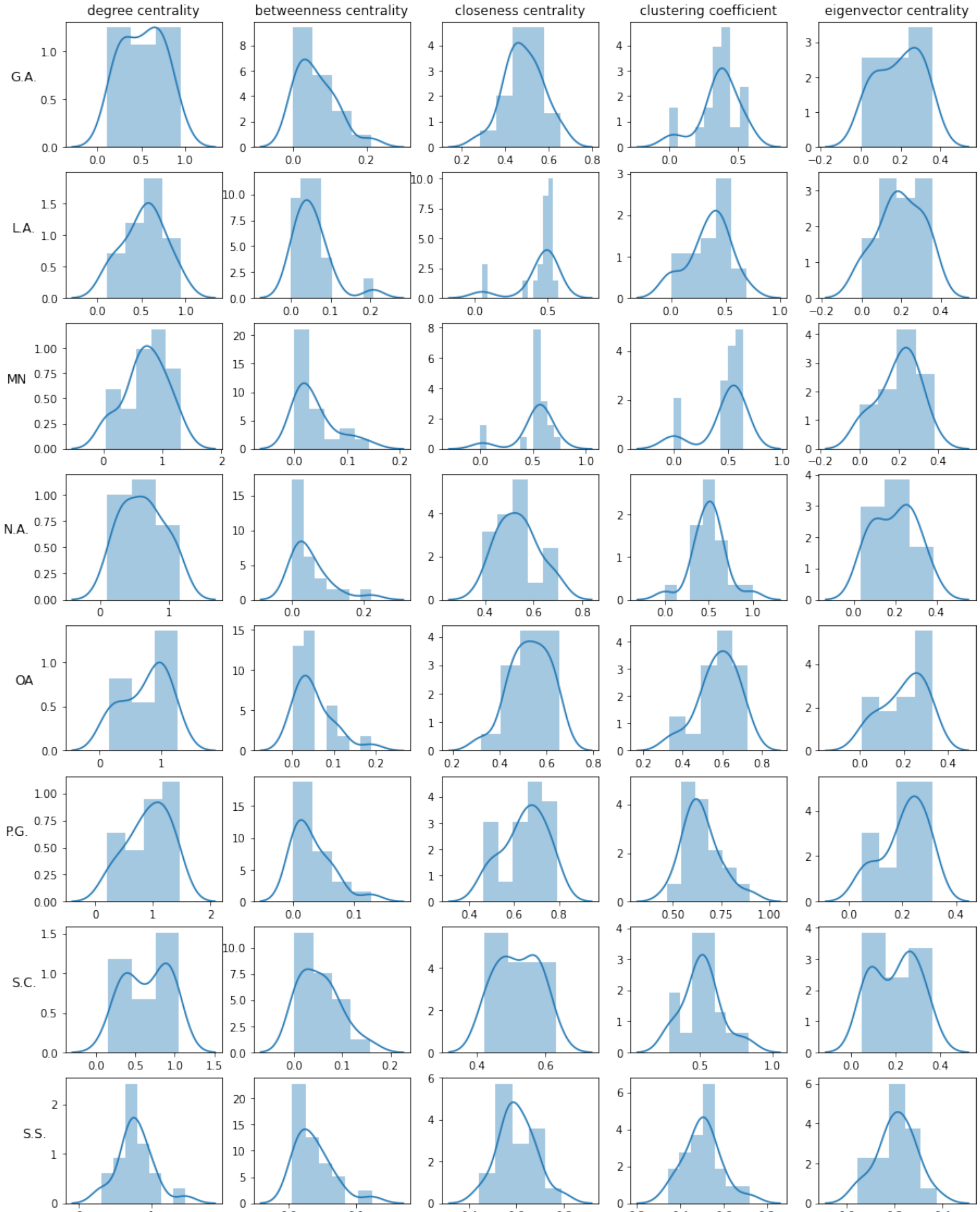


Figure 29: Distributions of IMN network indicators for SIS data.

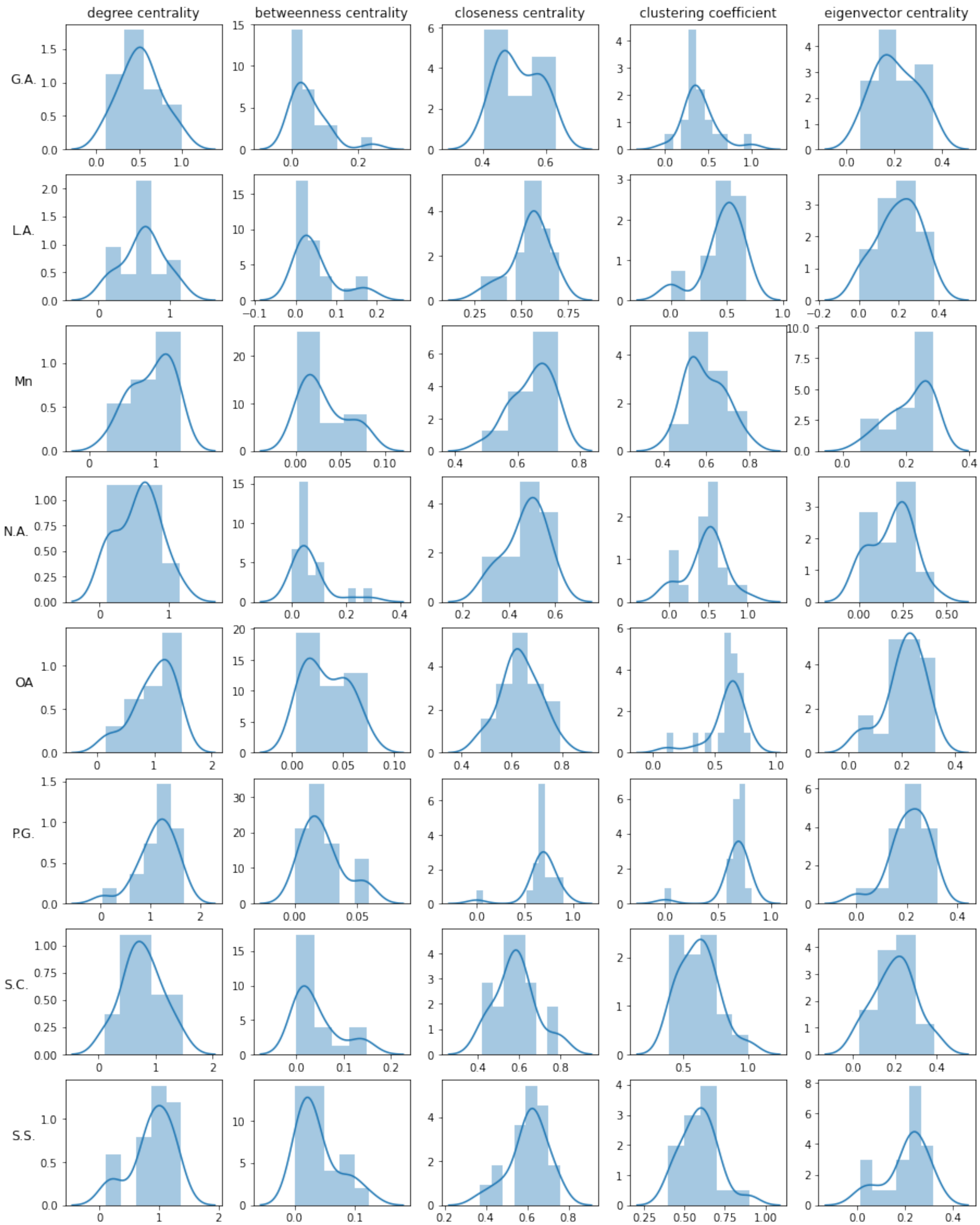


Figure 30: Distributions of IMN network indicators for PAP data.

6.4 Conclusion: potential of graph analysis methods

Application of graph analysis methods to eye tracking data is quite a novel approach, especially in settings when test participant are not supposed to have a common set of AOIs. In such settings, graphs (networks) derived from eye tracks of different participants cannot be straightforwardly compared, as each network has a different set of nodes. For comparison and generalization, graphs need to be characterized in terms of features that are not associated with specific nodes and links. Distributions of network indicators, as shown in Figs. 28 to 30, are potentially useful, but require further research to find out which of them are really relevant and how to interpret them in terms of data exploration behaviours and strategies. Our preliminary hypothesis is that the measures of degree centrality, eigenvector centrality, and clustering coefficient may be more useful and interpretable than the others. However, to verify this hypothesis and give more semantics to these measures, it is necessary to review the session records and try to understand how the observed behaviour of each participant translates into the corresponding distributions of the network indicators. This is a laborious and time-consuming kind of work, which, unfortunately, cannot be done within the remaining time of the project. Hence, the questions concerning the use of graph representations and graph analysis methods for eye-tracking data have to remain for future research.

We deem this research direction quite promising for the following reasons. In analysing the data by means of visual analytics techniques, the scarf plot approach was very useful due to enabling high level of abstraction, which allowed us to discover repeated behavioural patterns invariant to specific positions in the display. A graph representation has a potential to provide even a higher level of abstraction. First, it becomes possible to disregard specific sequences of visiting AOIs and investigate the connections and transitions between the AOIs. Second, it is possible to detach the networks from the display space and represent them visually using graph layout algorithms that minimize edge crossings and make the network structure more clearly visible than when the nodes are drawn in their original spatial positions, as it is in Figs. 25 to 27. This may enable discovery of meaningful structural components of the networks, i.e., subnetworks having particular structure. Examples of possible components are shown in Fig. 31. It is also reasonable to test the applicability and utility of network motif discovery methods (e.g., [13]), which search for recurrent subgraphs in large graphs or graph collections.

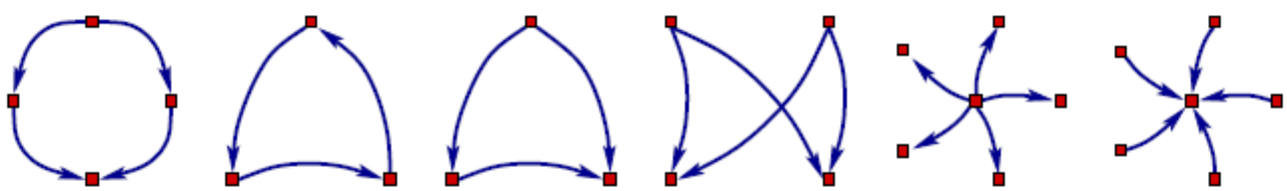


Figure 31: Possible structural components that may exist in IMNs.

To conclude, we regard the results that we obtained using the approach of IMN derivation and analysis as preliminary, requiring further research work, but we deem this approach as having high potential. We would be happy to have opportunities for further research in this direction in our future projects.

7 Discussion and conclusions

In Track&Know, we conducted experiments involving eye tracking for quite an unusual purpose. Instead of evaluating a specific display design, user interface, or data visualisation technique in terms of usability and task-effectiveness, we pursued a much more general and ambitious goal: investigate how eye tracking can be employed for understanding the process of visual data exploration. In accord with this goal, we designed challenging and practically meaningful analysis scenarios involving real-world data. We gave the experiment participants non-trivial and interesting tasks, in which they needed to find answers to open-ended questions concerning relationships between spatial phenomena. This differentiates our experiments from the usual practice of using toy examples, simple tasks, and close-ended questions with a few predefined variants of possible answers.

Due to these fundamental differences, the usual methods of analysing and interpreting eye-tracking data turned to be mostly useless for our purposes. Thus, it was not appropriate to define specific AOIs assuming that they are valid for all participants. Instead, we assumed that each participant may have his/her own AOIs. Hence, comparison of participants' gaze scanpaths in terms of similar sequences of AOI visits or searching for recurrent subsequences are neither possible nor useful. It was also not appropriate to compare the scanpaths with some hypothetical "optimal" scanpath defined using some measure of optimality. In our settings, there can be no optimal scanpath, as this very concept has no meaning in the context of exploratory visual data analysis. Moreover, in usual applications of eye tracking experiments, where the participants' tasks are to find specific pieces of information, re-visits of previously visited AOIs are considered undesirable and interpreted as evidence of display or UI problems. In exploratory analysis, re-visits should be considered as a normal and even necessary activity, as exploration requires performing multiple comparisons. Hence, instead of relying on usual, established approaches to analysing eye-tracking data, we had to search for other approaches that could allow us to achieve our goals.

It should be noted that raw eye-tracking data by themselves are very basic: they consist of mere display coordinates and time stamps. As we wanted to derive general, high-level knowledge, it was necessary to find approaches enabling an appropriate level of abstraction, to be able to find general patterns irrespective of specific display positions and times. In analysing the data by means of visual analytics techniques, we found the scarf plot approach very useful, as it allowed us to raise the level of abstraction in considering the data. We disregarded the specific positions in the display and investigated the character of the changes of the analysts' attention focus over time. This allowed us to discover and interpret repeated position-invariant patterns of changing the gaze focus. These patterns were quite easy to interpret as particular exploratory activities or strategies. We verified our interpretations by careful observation of the recorded participant's analysis sessions.

We systemised the patterns that we have discovered and further generalised them into analysts' exploration tasks and strategies, as described in Sections 5.6.4 and 5.6.5. We have been also able to draw general implications that are valid for design and implementation of software tools intended to support exploratory visual analyses (Section 5.7.1).

Besides visual analytics methods, we tested the possibility of using the approach involving derivation and analysis of individual mobility networks developed in WP4. The approach was originally intended for analysing human mobility in the geographic space. It was necessary to adapt the methods to the specifics of eye tracking data. Networks derived from eye tracking data provide an opportunity to analyse the data at even a higher level of abstraction than with the use of scarf plots. It becomes possible to disregard the sequence of visiting AOIs and reveal structural patterns of connections and transitions between AOIs. We assessed the first results obtained using this approach and found them potentially useful but requiring further investigation, which falls beyond the scope and time period of the project. We hope to have opportunities to further pursue this research direction in our future projects.

Hence, our research has shown that eye tracking can provide very valuable information, from which it is possible to extract pieces of general knowledge about data analysis processes. For this purpose, it is necessary to use techniques supporting high level of abstraction in considering the data. We see high potential in the approach of deriving individual mobility networks from eye tracks and analysing these networks by means of graph visualisation and analysis methods. The preliminary results obtained within the project Track&Know appear promising, but further investigation is required.

8 References

[1] G. Andrienko, N. Andrienko, M. Burch, nd D. Weiskopf. Visual Analytics Methodology for Eye Movement Studies. *IEEE Transactions on Visualization and Computer Graphics (Proceedings IEEE VAST 2012,* best paper award*),* 18(12): 2889-2898, Dec. 2012. doi: http://dx.doi.org/10.1109/TVCG.2012.276

[2] T. Blascheck, M. Burch, M. Raschke, and D. Weiskopf. Challenges and perspectives in big eye movement data visual analytics. In *Proceedings of the 1st IEEE International Symposium on Big Data Visual Analytics*. IEEE, Piscataway, 2015.

[3] T. Blascheck, M. John, K. Kurzhals, S. Koch, and T. Ertl. VA^2: a visual analytics approach for evaluating visual analytics applications. *IEEE Transactions on Visualization and Computer Graphics*, 22(01):61–70, 2016.

[4] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye-tracking data. In *Proceedings of EuroVis 2014*, pages 63–82. The Eurographics Association, 2014.

[5] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of eye-tracking data: A taxonomy and survey. *Computer Graphics Forum,* 36 (8): 260-284.

[6] A. Bojko. Human-Computer Interaction. New Trends; chapter "Informative or misleading? Heatmaps deconstructed", pages 30–39. LNCS. Springer, Berlin, 2009.

[7] M. Burch, G. L. Andrienko, N.V. Andrienko, M. Höferlin, M. Raschke, and D. Weiskopf. Visual task solution strategies in tree diagrams. In *Proceedings of 2013 IEEE Pacific Visualization Symposium (PacificVis),* Sydney, NSW, 2013, pp. 169-176, doi: 10.1109/PacificVis.2013.6596142.

[8] M. Burch, A. Kull, and D. Weiskopf. AOI rivers for visualizing dynamic eye gaze frequencies. *Computer Graphics Forum*, 32(3):281–290, 2013.

[9] E.H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Proceedings of IEEE Symposium on Information Visualization*, pages 69–75. IEEE Computer Society, Los Alamitos, 2000.

[10] A. Duchowski. Eye Tracking Methodology: Theory and Practice. Science+Business Media. Springer, 2nd, 2007.

[11] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.

[12] J.H. Goldberg and J.I. Helfman. Visual scanpath representation. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 203–210, New York, 2010. ACM.

[13] J.A. Grochow and M. Kellis. Network Motif Discovery Using Subgraph Enumeration and Symmetry-Breaking. In: Speed T., Huang H. (eds) Research in Computational Molecular Biology. RECOMB 2007. Lecture Notes in Computer Science, vol 4453. Springer, Berlin, Heidelberg, 2007.

[14] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press, 2011.

[15] H. Kirchner and S.J. Thorpe. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vision Research, 46(11):1762–1776, 2006.

[16] K. Kurzhals, B.D. Fisher, M. Burch, and D. Weiskopf. Evaluating visual analytics with eye tracking. In *Proceedings of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 61–69, New York, 2014. ACM.

[17] K. Kurzhals, M. Burch, T. Blascheck, G. Andrienko, N.Andrienko, and D. Weiskopf. A task-based view on the visual analysis of eye-tracking data. In M. Burch, L. Chuang, B. Fisher, A. Schmidt, and D. Weiskopf, editors, *Eye Tracking and Visualization*. *Foundations, Techniques, and Applications, Mathematics and Visualization,* pages 3–22. Springer, 2017.

[18] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, volume 5, pages 2–4, 2005.

[19] M. Raschke, D. Herr, T. Blascheck, M. Burch, M. Schrauf, S. Willmann, and T. Ertl. A visual approach for scan path comparison. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 135–142, New York, 2014. ACM.

[20]D.D. Salvucci and J.H. Goldberg. Identifying fixations and saccades in eye tracking protocols. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, pages 71–78, New York, 2000. ACM.
[21]B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *Int. J. Human-Computer Studies*, 57(4):247–262, 2002. doi: https://doi.org/10.1006/ijhc.2002.1017

9 Annex I: Ethics report

The following participant information sheet, participant consent form, and withdrawal forms have been approved by the ethics advisor. After reading the information sheet, every participant signed the consent form.

9.1 PARTICIPANT INFORMATION SHEET

Title of Research: Eye Tracking for Mobility Analytics

Track & Know is a Horizon2020 project, with a focus on Big Data. More specifically, Track & Know, which is an acronym for “Big Data for Mobility Tracking Knowledge Extraction in Urban Areas”, is researching, developing and exploiting a new software framework that aims at increasing the efficiency of Big Mobility Data. This framework is being applied in the transport, mobility, motor insurance and health sectors.
Track & Know aims to introduce innovative software stacks and toolboxes addressing new emerging cross-sector markets related to automotive transportations and urban mobility in general: commercial IoT services; car insurance; and, healthcare management. The addressed markets have significant industrial and commercial impacts for EU enterprises. The project will actively run between January 2018 and December 2020.

One of the goals of the project is to create analytical dashboards for presenting relevant information to domain experts. The goal of our study is to investigate the understandability, interpretability, and analytical effectiveness of specific techniques for cartographic representation of spatial distributions for domain experts. We do not intend to test the perceptual and cognitive capabilities of people; our focus is on evaluating the representations, which need to be understandable to and useful for people with varied professional backgrounds and experiences.

Your participation in ‘Eye Tracking for Mobility Analytics’ will consist of two phases. On the 1st stage, you will be presented a data map for familiarization. On the 2nd stage, you will be asked to explore how the events of interest are distributed in relation to the background distribution; detect “hot spots” and/or “cold spots” in the event distribution, and report your observations with sketching the patterns you have noted on a contour map. Your activity will be recorded on video (only sketching on the map, which will not include any parts of your body except the hands) and tracked using either eye tracking technology (if the social distancing regulations permit that at the moment of the study) or by tracking your mouse movements on the screen otherwise. A very general professional information will be collected via a questionnaire: profession, career duration, and self-estimation of your experience in respect to using maps for data analysis. The data will be stored on a protected server and used in anonymous and aggregated form for the purposes of the project and following scientific publications. The data will be deleted after publishing results of the study, latest 1 year after the end of the project.

Participation is voluntary and will not be used to assess your personal performance or affect your employment. You can withdraw from the experiment at any time if you change your mind; you do not have to give a reason.

We will not collect any names so please be as honest and specific as possible when answering the questions. If any of the video recording is shown to non-project members, we will ensure you cannot be identified (e.g. by blurring faces).

Data Protection By participating in this study and submitting this questionnaire we assume that you give us permission to use your questionnaire data in this project and refer to it in further projects. Data will go directly to the researchers who will store it securely and analyse it. The overall findings from the research will be shared with the academic research community and the European Commission to ensure that the knowledge gained can be put to use widely, thus benefitting EU citizens.

Research Contact Any questions about this research can be sent to:

Experiment Leader:	Project Manager:
Gennady Andrienko	Ibad Kureshi
Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS	Inlecom Systems, BE
gennady.andrienko@iais.fraunhofer.de	ibad.kureshi@inlecomsystems.com
+49 2241 142486	+32 4669 00962

9.2 PARTICIPANT CONSENT FORM

Title of Research: 'Eye Tracking for Mobility Analytics'

To be completed by the participant (delete or tick as necessary)	
1. Has the researcher explained the study to you?	YES / NO
2. Have you had an opportunity to ask questions and discuss this study?	YES / NO
3. Have you received satisfactory answers to all your questions?	YES / NO
4. Have you received enough information about this study?	YES / NO
5. Do you understand that you are free to withdraw from this study: ▪ at any time ▪ without giving a reason	YES / NO
6. Do you agree to take part in this study?	YES / NO

Name of participant in block letters	
Signature of participant	
Date	DD / MM / YYYY

This study is supervised by: Gennady Andrienko
Contact details:
Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
gennady.andrienko@iais.fraunhofer.de
+49 2241 142486

9.3 PARTICIPANT WITHDRAWAL FORM

Title of Research: 'Eye Tracking for Mobility Analytics'

<table>
<tr><td colspan="2">To be completed by the participant (delete or tick as necessary)</td></tr>
<tr><td colspan="2">I hereby withdraw from participation in the trial. In doing so, I acknowledge that I will now no longer receive payment for my participation given that I will now not complete the research experiment. YES / NO</td></tr>
<tr><td colspan="2">If possible, can you state why you have withdrawn?</td></tr>
<tr><td>Name of participant in block letters</td><td></td></tr>
<tr><td>Signature of participant</td><td></td></tr>
<tr><td>Date</td><td>DD / MM / YYYY</td></tr>
</table>

10 Annex 2: Ethics Proforma

Track & Know

Ethics Check-List and Questionnaire

THIS FORM NEEDS TO BE FILLED-IN BY THE DELIVERABLE LEADER BEFORE DURING AND AFTER THE WORK LEADING TO THE RELEVANT DELIVERABLE. DELIVERABLE LEADERS ARE ENCOURAGE TO DISCUSS EACH ACTIVE QUESTIONAIRE WITH THE ETHICS COMMITTEE. EACH OPEN QUESTIONNAIRE SHOULD EITHER BE STORED IN THE PROJECT TEAMWORK SPACE OR A LINK MADE AVAILABLE TO THE RELAVANT GOOGLE DOC. ALL OPEN QUESTIONNAIRES WILL BE REVIEWED AS PART OF PROJECT FACE TO FACE MEETINGS. COMPLETED FORMS NEED TO BE SUBMITTED AS PART OF THE DELIVERABLE Q/A PROCESS TO GET AGREEMENT WITH THE EtC. IN THE EVENT OF COMMENTS AND/OR QUESTIONS BY THE ETHICS COMMITTEE, THE DELIVERABLE LEADER HAS TO PROVIDE RELEVANT RESPONSES AND/OR CLARIFICATIONS IN A TIMELY MANNER

<u>A. PERSONAL DATA</u>

1. Has personal data going to be processed for the completion of this deliverable?

   Yes: one-screen eye-tracking during data exploration session

   1. If “yes”, do they refer only to individuals connected to project partners? Or to third parties as well?

      Yes, volunteers affiliated with the project partner coordinating this deliverable

2. Are “special categories of personal data” going to be processed for this deliverable? (whereby these include personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, and trade union membership, as well as, genetic data, biometric data, data concerning health or data concerning a natural person's sex life or sexual orientation)

   No

3. Has the consent of the individuals concerned been acquired prior to the processing of their personal data?

   Yes

   1. If “yes”, based on the Project’s Consent Form? On a different legal basis?

      Yes, based on the Project’s Consent Form

4. In the event of processing of personal data, is the processing:
   1. “Fair and lawful”, meaning executed in a fair manner and following consent of the individuals concerned?
      Yes

2. Performed for a specific (project-related) cause only?

   Yes

3. Executed on the basis of the principle of proportionality (meaning that only data that are necessary for the processing purposes are being processed)?

   Yes

4. Based on high-quality personal data?

   Only eye-tracking information was used

5. Are all lawful requirements for the processing of the data (for example, notification of the competent Data Protection Authority(s), if applicable) adhered to?

   Yes

6. Have individuals been made aware of their rights (particularly the rights to access, rectify and delete the data)?

   Yes

B. DATA SECURITY

1. Have proportionate security measures been undertaken for protection of the data, taking into account project requirements and the nature of the data?
   1. Brief description of such measures (including physical-world measures, if any)

      Data are stored only on protected computers with restricted access, with BitLocker protection

2. Is there a data breach notification policy in place within your organisation?

   Yes

C. DATA TRANSFERS

1. Are personal data transfers beyond project partners going to take place for this deliverable?
   1. If “yes”, do these include transfers to third (non-EU) countries?

      No

2. Are personal data transfers to public authorities going to take place for this deliverable?
   1. Do any state authorities have direct or indirect access to personal data processed for this deliverable?

   No

3. Taking into account that the Project Coordinator is the "controller" of the processing and that all other project partners involved in this deliverable are "processors" within the same contexts, are there any other personal data processing roles attributed to any third parties for this deliverable?

   No

D. ETHICS AND RELATED ISSUES

1. Are personal data of children going to be processed for this deliverable?
   No
2. Is profiling in any way enabled or facilitated for this deliverable?
   No
3. Are automated-decisions made or enabled for this deliverable?
   No
4. Have partners for this deliverable taken into consideration system architectures of privacy by design and/or privacy by default, as appropriate?
   N/A
5. Have partners for this deliverable taken into consideration gender equality policies?
   N/A
6. Have partners for this deliverable taken into consideration confidentiality of the data requirements?
   Yes