

**Palacký University Olomouc,
Faculty of Science, Department of Geoinformatics**

**Paris Lodron University Salzburg,
Faculty of Digital and Analytical Sciences, Department of Geoinformatics**

USER EVALUATION OF INTERACTIVE COVID-19 DASHBOARDS

Diploma thesis

Author

Anna PORTI SUÁREZ

Supervisor (Palacký University Olomouc)
RNDr. Stanislav POPELKA, Ph.D.

Co-supervisor (Paris Lodron University Salzburg)
Assoc. Prof. Dr. Barbara SCHERNTHANNER-HOFER

**Erasmus Mundus Joint Master Degree Programme
Copernicus Master in Digital Earth
Specialization Track Geovisualization & Geocommunication
Olomouc, Czech Republic, 2022**



Palacký University
Olomouc



ANOTATION

The outbreak of the COVID-19 pandemic brought dashboards to the fore and they became widely used by the public. Nevertheless, dashboard interfaces are related to business intelligence since their origins and the seek for the improvement of their design is not new. This thesis' objective is to conduct user evaluation of COVID-19 dashboards containing geospatial information through a formative study to identify problematic user interaction aspects. This is enhanced by comparing two self-developed dashboards with different appearances of their functionalities according to the results of the user testing performed previously. The development of two dashboards is part of the goal of familiarising with technologies to create dashboard interfaces building a spatial data infrastructure. The user evaluation is performed with mixed research combining objective (eye-tracking technologies) and subjective (a questionnaire and an interview) methods. The results formulate recommendations for better design of the elements and how the layout and content transfer the information correctly: the interactivity and availability to choose the metrics for the user are key elements to achieve this. Regarding the distribution of the elements in the interface, their size and position play a role in a more user-friendly and intuitive interaction with the user to transfer the content appropriately.

KEYWORDS

Eye-tracking, Qualitative Methods, Spatial Data Infrastructure, Usability Testing, User Interface

Number of pages: 58

Number of attachments: 3

This thesis has been composed by *Anna PORTI SUÁREZ* for the Erasmus Mundus Joint Master's Degree Program in Copernicus Master in Digital Earth for the academic year 2021/2022 at the Department of Geoinformatics, Faculty of Natural Sciences, Paris Lodron University Salzburg, and Department of Geoinformatics, Faculty of Science, Palacký University Olomouc.

Hereby, I declare that this piece of work is entirely my own, the references cited have been acknowledged and the thesis has not been previously submitted to the fulfilment of the higher degree.

A handwritten signature in black ink, consisting of a stylized 'A' followed by a wavy line that extends to the right.

May 19, 2022, Olomouc

Anna PORTI SUÁREZ

I would like to thank my supervisor, Dr. Stanislav Popelka, for all the time he has spent guiding me through the duration of this project, by teaching me the necessary background information, solving all my doubts and his constant assistance and support; together with my co-supervisor, Dr. Barbara Schernthanner-Hofer, and the insights she provided. Also, I want to thank all the people who took the time to participate in the experiments and helped me achieve this work.

ASSIGNMENT OF DIPLOMA THESIS

(project, art work, art performance)

Name and surname: Anna PORTI SUÁREZ
Personal number: R200694
Study programme: N0532A330010 Geoinformatics and Cartography
Work topic: User evaluation of interactive COVID-19 dashboards
Assigning department: Department of Geoinformatics

Theses guidelines

The aim of the thesis is to conduct user evaluation of different types of dashboards, including those dedicated to COVID-19 with an emphasis on spatial information. In the theoretical part of the thesis, the student will describe dashboards from different perspectives- historical, technological, and user. In the practical part, usability testing will be conducted on existing dashboards which will help identify problematic user interaction aspects. This formative study will be enhanced by the summative one comparing two (self-developed) dashboards with different appearances. A combination of objective (i.e. eye-tracking, screen-recording) and subjective (i.e. questionnaire, interview) methods will be used for the evaluation. The thesis results might help reveal how the users interact with the dashboards, whether they are able to use their advanced functionality (filtering, interaction) and whether the displayed information is correctly transferred.

Extent of work report: max. 50 stran
Extent of graphics content: as needed
Form processing of diploma thesis: printed
Language of elaboration: English

Recommended resources:

- Alhamadi, M. (2020, July). Challenges, Strategies and Adaptations on Interactive Dashboards. In Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (pp. 368-371).
- Bojko, A. (2013). Eye tracking the user experience: A practical guide to research. Rosenfeld Media.
- Dowding, D., & Merrill, J. A. (2018). The development of heuristics for evaluation of dashboard visualizations. *Applied clinical informatics*, 9(3), 511.
- Golebiowska, I., Opach, T., & Rod, J. K. (2017). For your eyes only? Evaluating a coordinated and multiple views tool with a map, a parallel coordinated plot and a table using an eye-tracking approach. *International Journal of Geographical Information Science*, 31(2), 237-252.
- il Yooa, K., & Kronenfelda, B. J. An Evaluation of COVID-19 Dashboards from Cartographic and Epidemiological Perspectives.
- Mukhtar, H., Ahmad, H. F., Khan, M. Z., & Ullah, N. (2020, December). Analysis and Evaluation of COVID-19 Web Applications for Health Professionals: Challenges and Opportunities. In *Healthcare* (Vol. 8, No. 4, p. 466). Multidisciplinary Digital Publishing Institute.
- Popelka, S., Herman, L., Řezník, T., Pařilová, M., Jedlicka, K., Bouchal, J., Kepka, M., Charvat, K. (2019) User Evaluation of Map-Based Visual Analytic Tools, *ISPRS International Journal of Geo-Information*, 8(8), 363; doi:10.3390/ijgi8080363
- Popelka, S., Vondrakova, A., & Hujnakova, P. (2019). Eye-Tracking Evaluation of Weather Web Maps. *ISPRS International Journal of Geo-Information*, 8(6), 29. doi:10.3390/ijgi8060256
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., ... & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30-41.

Supervisors of diploma thesis: **RNDr. Stanislav Popelka, Ph.D.**
Department of Geoinformatics

Date of assignment of diploma thesis: **December 9, 2021**

Submission deadline of diploma thesis: **May 5, 2022**

UNIVERZITA PALACKÉHO V OLOMOUČI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA GEOINFORMATIKY
17. listopadu 50, 771 46 Olomouc



L.S.

doc. RNDr. Martin Kubala, Ph.D.
Dean

prof. RNDr. Vít Voženílek, CSc.
Head of Department

Olomouc December 16, 2021

CONTENT

INTRODUCTION	8
1 OBJECTIVES	9
2 METHODOLOGY	10
2.1 User Evaluation.....	10
2.2 Dashboard Development.....	11
3 STATE OF ART	13
3.1 Dashboards.....	13
3.2 User Evaluation.....	16
4 EXPERIMENT I	21
4.1 Methods	21
4.2 Results.....	27
4.2.1 Correctness.....	27
4.2.2 Trial Duration	27
4.2.3 AOIs	29
4.2.4 Questionnaire and Interview	32
4.3 Conclusion	34
5 DASHBOARD DEVELOPMENT	36
5.1 Methods.....	36
5.1.1 Data Sources	36
5.1.2 Technologies	36
5.1.3 Implementation	36
5.2 Results.....	38
5.3 Conclusion	40
6 EXPERIMENT II	41
6.1 Methods	41
6.2 Results.....	45
6.2.1 Entry Time	45
6.2.2 Correctness.....	46
6.2.3 Trial Duration	47
6.2.4 AOIs	49
6.2.5 Questionnaire and Interview	51
6.3 Conclusion	52
7 RESULTS	54
7.1 Experiment I	54
7.2 Dashboard Development.....	54
7.3 Experiment II.....	55
8 DISCUSSION	56
9 CONCLUSION	58
REFERENCES AND INFORMATION SOURCES	
ATTACHMENTS	

INTRODUCTION

The term *dashboard* can be described as a graphical user interface displaying at-a-glance selected valuable information, retrieved from regularly updated data sources. Since the COVID-19 outbreak in 2020, dashboards have appeared to be widely used by the public to consult metrics related to the pandemic in a specific area.

The main objective of this thesis is to conduct user evaluation of dashboards containing geospatial information regarding the COVID-19 topic. From identifying the historical role of dashboards in the state of art of this work, finding out the technologies involved in their development during the research chapters, to evaluating the user interaction with this type of interface.

The research consists of a formative study that helps identify problematic user interaction aspects in existing COVID-19 dashboards and it is enhanced by comparing two self-developed dashboards with different appearances of their functionalities according to the results of the user experience performed previously. The dashboard development also allows the familiarisation with dashboard elaboration technologies.

The nature of this study requires the usage of mixed research. A combination of objective and subjective methods is used to reach these targets: eye-tracking together with an interview and a questionnaire, respectively. The thesis results should help reveal how the users interact with the dashboards and their functionalities in different elements arranged in various layouts, and if the displayed information is correctly transferred.

1 OBJECTIVES

The aim of the diploma thesis is to perform user evaluation on COVID-19 dashboards. For this, three goals are established and described as follows.

The first goal (partial goal 1) is to perform user evaluation on existing dashboards with the aim of obtaining insights about the user interaction with their functionalities and whether they correctly transfer the information to the user or not. This consists of a formative study that has three steps: the design of the experiment, the recording of the data and the processing and analysis of the obtained data. The results allow us to identify the problematic elements of these dashboards and formulate recommendations, which are used to obtain insights of which elements of a dashboard are user-friendly and communicate the information accordingly.

The second goal (partial goal 2) is to elaborate two self-developed dashboards according to the insights obtained in partial goal 1. Once the problematic elements and the recommendations for improving user interaction are identified, two dashboards are created to confirm or reject the recommendations formulated regarding the insights from the first goal within the third goal. Also, the aim is to get familiar with dashboard development technologies by creating a spatial data infrastructure (SDI) that contains a back-end system that retrieves and stores the data, and a user interface (front-end) to visualise it.

The third goal (partial goal 3) is to perform user evaluation on the two self-developed dashboards considering the insights of partial goal 1. It also consists of a formative study that identifies (and compares) the positive and negative elements of each dashboard to conclude with the previous studies. Like the first goal, it has three steps: the design of the experiment, the recording of the data and the processing and analysis of the obtained data, which lead to the results and conclusions. The goals are summarised in Figure 1.



Figure 1 Partial goals summarised.

As a result, two dashboards of a specific area for consulting COVID-19 metrics are developed as stimuli to perform user evaluation, together with identifying problematic interaction aspects and elaborate recommendations for designing user-friendly and communicative dashboard interfaces.

2 METHODOLOGY

The methodology is structured as an overview of the employed methods, used apparatus and software, and how they are integrated into the procedure. Since this thesis includes eye-tracking usage for user testing and creating dashboard interfaces, this chapter has two sections to cover them separately.

2.1 User Evaluation

Used Methods

To analyse what our visual-cognitive system perceives, user evaluation methods are used in all kinds of fields. Cartography is among them and can employ user testing to produce functional maps, or in this case, dashboards containing spatial information. This study uses mixed research by combining both qualitative and quantitative methods.

The following section is an overview of the methods employed for two experiments performed during the thesis: the first one with existing dashboards, and the second one with self-developed dashboards, taking into account the insights obtained in the first one. The details about the methodology of each are described in the subchapters 4.1 and 6.1.

Experiment I is a formative study to analyse four existing dashboards' problematic aspects and recommend their improvements. For this, eye-tracking is used for data recording of the eye movement of participants, while interacting with the dashboards during a task solving process. Qualitative methods employed include an interview and a questionnaire to obtain insights into the user's experience.

Experiment II is also a formative study that compares the previously developed two dashboards and uses the same methodology as experiment I, only changing in an added free viewing period of one minute for each dashboard for the quantitative methods, and subjective questions regarding comparative insights for the qualitative.

In both cases, the methods provide the following outcomes for analysis of the data: the correctness rate of the task solving to evaluate if the participants are able to reach the suggested targets; the trial duration to see the efficiency, or how long it takes the participants to solve the tasks; a sequence chart of the fixation time per area of interest (AOI) per task, that show which elements they focus more on when interacting with the interfaces; and the interview and questionnaire, which compile negative and positive assets to measure their satisfaction towards the interface, in a more comparative way in the case of the second experiment. Experiment II also includes the entry time regarding the free viewing part, which indicates the average time that it takes the participant to do the first fixation to a specific AOI.

Used Data

For experiment I, four existing dashboards with different appearances and functionalities are selected to carry out the formative study. These are the following: COVID-19 map by the John Hopkins University, Novel Coronavirus Incidence Map by the University of Washington, OCHA Coronavirus Data Explorer and Health Map, visible in Figure 8, Figure 9, Figure 10 and Figure 11 (pages 22 and 23), respectively.

As mentioned, experiment II uses two self-developed dashboards of the region of Catalonia. These are elaborated considering the insights of the first experiment and their interfaces are purposely designed to contain the same functionalities in diverse ways to compare them. They are visible in Figure 26 and Figure 27 (page 39).

Used Software

The following tools are used to conduct the experiments:

- SMI Experiment Center to design the experiment.
- iView X to record the data.
- BeGaze 3.7 for data recording, processing and analysis, once the experiments are performed.

The employed apparatus to record eye movement is the eye-tracker SMI RED 250, which records at a frequency of 250 Hz, in the eye-tracking laboratory of the Department of Geoinformatics of the Palacký University Olomouc, Czechia. A camera and a microphone (Logitech C920) are also employed, and the stimuli are displayed in a Google Chrome browser on the participants' 24-inch screen.

Processing Procedure

The obtained data for both experiments need to be processed. This involves detecting the saccades and fixations by applying the dispersion threshold identification (I-DT) algorithm, dividing the recorded data into custom trials to analyse the tasks in individual contexts and creating areas of interest (AOIs) for further analysis of the participant behaviour. Together with compiling all the qualitative data obtained in the questionnaire and interview.

2.2 Dashboard Development

Used Methods

Experiment II is performed with two self-developed dashboards of the COVID-19 in Catalonia. The target is to display the insights of experiment I and compare their interfaces, together with familiarising with dashboard development technologies. The method used for this is the building of an SDI by the usage of different standardised tools to communicate this information.

Used Data

The dashboards retrieve data from two Catalan public institutions: the Catalogue of Open Data of the Government of Catalonia¹ (*Dades Obertes Catalunya*) and the Statistical Institute of Catalonia² (*Institut d'Estadística de Catalunya*). The data refers to the number of COVID-19 cases from 01.03.2020 in CSV format, the population in 2020 (to calculate the incidence rate), also in CSV format, and the administrative boundaries of the displayed levels, counties and towns, in shapefile format.

¹ Data source available in the following link: <https://analisi.transparenciacatalunya.cat/>

² Data source available in the following link: <https://www.idescat.cat/>

Used Software

The SDI is the result of a compound of technologies: the back-end contains a Linux server with Python scripts running, and a spatial database that stores the data (PostGIS); the front-end is visualised using Tableau's desktop application, which retrieves the data and builds interactive dashboard interfaces, together with allowing the communication between the server and the front-end user interface. Moreover, Canva is used to design the background of the dashboards.

Processing Procedure

The implementation of these technologies results in a data flow where every element has a different role in the architecture. First, a Linux server is the back-end of the infrastructure that supports all the processes happening: a Python script manages the automatic data retrieval from the data source, the data cleaning, and the data manipulation, including the geometries of the administrative regions. Pandas is used to read the CSV data and perform all necessary manipulations and the geographic dimension of the data is handled with Geopandas.

The connection to the database is done in the server via SQLAlchemy and GeoAlchemy, which synchronises the resultant data frames into a PostGIS's database in the shape of tables containing all the relevant information to be displayed in the dashboard. It contains four tables: two with the geometry of each administrative level and two with the COVID-19 metrics for each administrative level, towns and counties.

Tableau Desktop, used to develop the dashboards, directly retrieves the data from the database, joins the tables containing the geometry with the tables containing the COVID-19 metrics in both administrative levels, as well as within them. It allows the construction of interactive dashboard interfaces with different functionalities, and Tableau Public enables publishing and sharing the link to the dashboards. Tableau Desktop, though, does not allow automatic refresh of the data, which needs to be done manually.

3 STATE OF ART

3.1 Dashboards

There is no exact definition of what a dashboard is. A way to define it by Few (2006) is as “a visual display of the most valuable information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at-a-glance”. Via a dashboard interface, the display of the information allows visual analysis of a specific phenomenon. This method permits to interact with visual representations, as well as, as Few (2006) points out, “change the nature of the display, filter out what’s not relevant, drill into lower levels of detail, and highlight subsets of data across multiple graphs simultaneously”.

It is important to understand the relationship between dashboards and the business intelligence field to contextualise their history, since their beginnings are very related to it. According to Few (2006), dashboard precursors are executive information systems (EISs), defined as “computer-based information systems designed to provide senior managers access to information relevant to their management activities” (Elliot, 2003), so these were limited to the offices of executives in terms of accessibility and understanding.

Data displayed integrated with other technologies and accessing information did not become widespread until the 90s (Few, 2006). Other precursors of dashboards, developed in this decade, were key performance indicators (KPIs) or “vital navigation instruments used by managers to understand whether their business is successful; the right set of indicators shines a light on performance and highlighted areas that need attention” (Marr, 2012). In other words, tools to monitor the performance of a firm, composed of indicators using diverse metrics. As Mattern (2017) says “data displays often mimic the dashboard instrumentation of cars or aeroplanes. Where in a car you would find indicators for speed, oil, and fuel levels, here you will find widgets representing your business’s key performance indicators: cash flow, stocks, inventory...”. Mattern also points out that the shapes of the displays that dashboards have nowadays are influenced by the classic appearance of the Bloomberg terminals, as displayed in Figure 2.

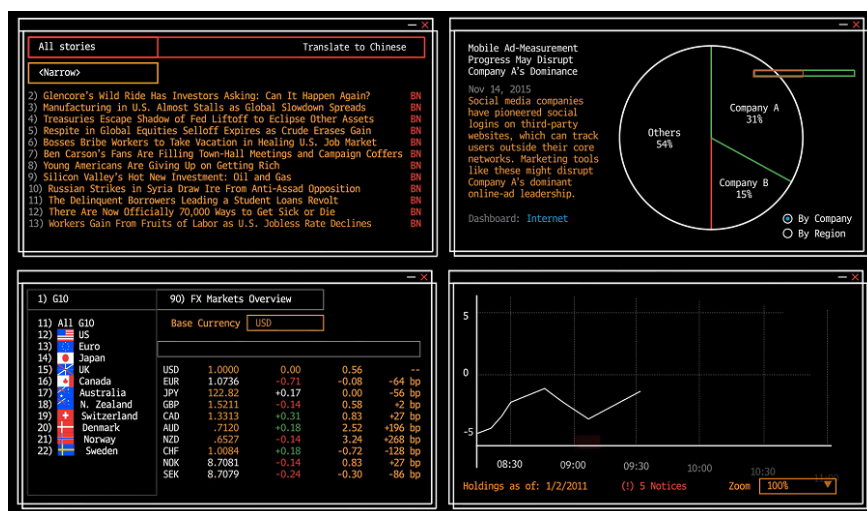


Figure 2 Bloomberg's dashboard terminal example (Bloomberg, 2016).

According to Mattern (2017), “the dashboard market now extends far beyond the corporate world”. Early examples of this trend date to the nineties when several American cities started combining metrics with mapping crime statistics, such as New York, Charlotte, or Michigan, among others. This evolved into the development of urban dashboards as a real-time data platform for communities to use and allow citizens to participate in urban planning decision making (Mattern, 2017). A first example was developed in London, a city dashboard including widgets that offer an overview of what is going on in the city related to weather, metro line status, air pollution, traffic, or even news and Twitter trends, as observed in Figure 3. Now, this trend is common, and even patterns can be observed in other urban dashboards (Mattern, 2017).

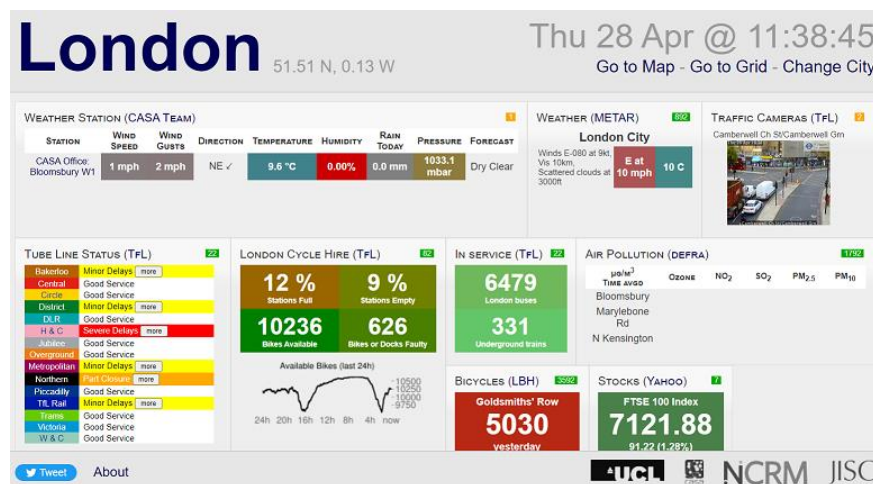


Figure 3 London City urban dashboard (City Dashboard, 2022).

This develops the idea of what a dashboard is today. As mentioned before, it is hard to define exactly the term but, as Few (2006) determines, “dashboards are visual displays: the information is presented visually, usually as a combination of text and graphics, but with an emphasis on graphics”. Few also highlights five important points regarding dashboard design and its relation to visual perception:

- Display the information necessary to achieve one or more specific objectives.
- Fit in a single computer screen so everything can be seen at once.
- They are used to monitor information at-a-glance.
- Their display mechanisms are small, concise, clear and intuitive.
- They are customised in order to serve their purpose.

When it comes to visualising geospatial information, dashboards are also a popular tool to be employed. Related to the previous definitions, geospatial phenomena can be monitored at-a-glance by data display using dashboard interfaces. Data can be visualised in several ways, combining maps with other elements, such as tables or graphs, according to Kitchin (2015), “with queries in the different panes replicated across them so that clicking on a data cell highlight the same data point on the graph and the area it refers to on the map”. Kitchin also defines the term *map-based dashboards*, therefore dashboards containing geospatial information, as “designed to present a collection of data, and also to support the visual learning and analytical reasoning of geospatial knowledge”.

Another mentioned point is that dashboards make possible that citizens can access online dashboards providing information about the city and trends and allow them to be able to get this information without knowing how to handle data and only by visualising it. When referring to urban dashboards, Mattern (2017) affirms that their accessibility is “changing the way we see our cities, since they do not merely seek to display information about a system, but to generate insights that human analysts use to change the system, to render it more efficient, sustainable, profitable...”.

Aside from the business intelligence field, dashboard interface data visualisations provide insights into one or more aspects in a panel displayed manner. When the information is combined with geospatial visualisation, it is in order to understand the spatial dimension of the selected phenomenon. Current trends show that dashboard popularity increased. The COVID-19 outbreak in 2020 brought dashboards to be widely used by the public for consulting metrics, in most cases including geospatial information. The Google Trends for searching the term *dashboard* show a considerable increase at the beginning of the year 2020. Figure 4 visualises this phenomenon from 2004 to current times (as of November 2021), by measuring the popularity of the search of the term: where a value of 100 is the highest point for the given time, 50 is half of the popularity.

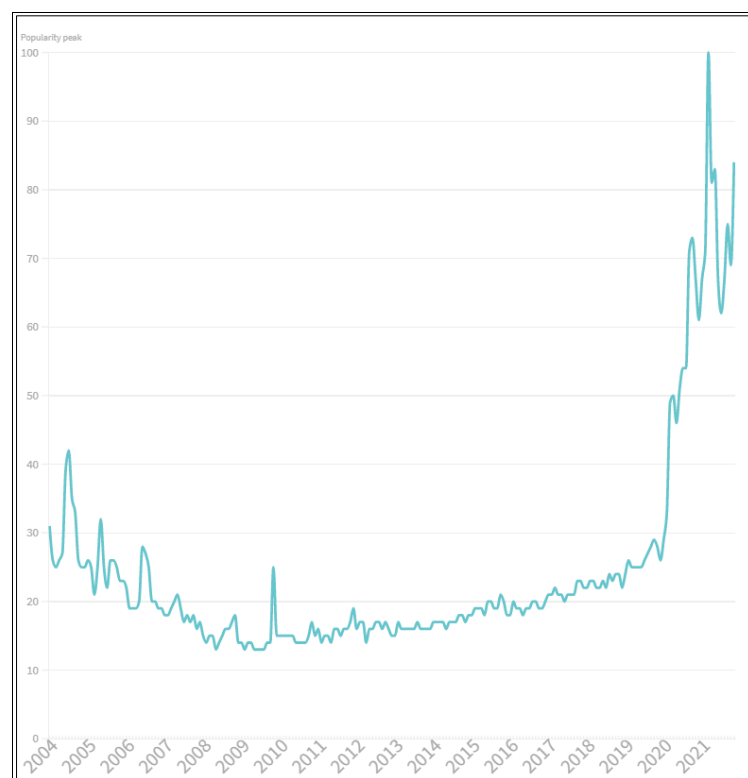


Figure 4 Google Trends for the interest in the search of the term *dashboard*.

One popular example from the initial stages of COVID-19 was the dashboard called *COVID-19 Map* elaborated by the Johns Hopkins University’s Coronavirus Resource Centre, displaying metrics about the COVID-19 cases in countries updated daily. Now it has evolved to the display of different regional-scale data, as well as other variables such as vaccination numbers, together with weekly or 28-day trends. Its interface is visible in Figure 5.

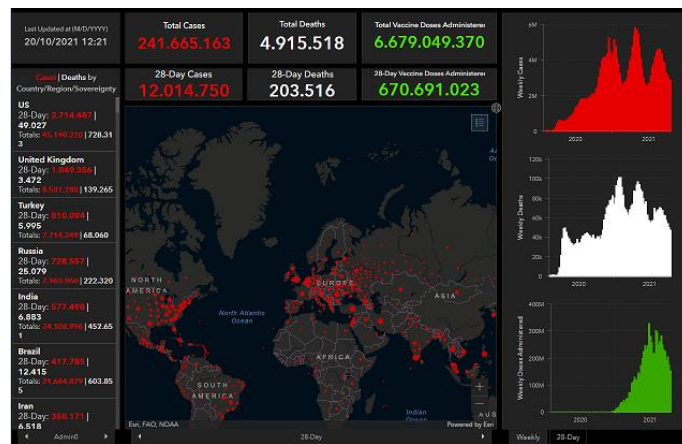


Figure 5 COVID-19 Map (Johns Hopkins Coronavirus Resource Centre, n.d.).

Nevertheless, for the purpose of communicating this information, the data and visualisation need to be appropriate. Even though dashboards have recently become a popular tool for data visualisation, not all the available displays are appropriately presenting the information. According to Few (2006), “visual communication involves semantics and syntax, like verbal language. You must know the rules to communicate effectively with graphs”. Also, Few states “it mostly involves science, a set of rules based on what we know about visual perception and cognition”.

Therefore, to make dashboards a useful visual tool to provide insights into specific phenomena, and in this case its spatial dimension, it is necessary to analyse how the information needs to be displayed, by using the proper elements. This is further developed in posterior sections, since the focus of this work is evaluating the user experience interacting with dashboards showing geospatial data related to the COVID-19.

From its origins in the business intelligence field to the short version of the definition and design properties of dashboards constructed by Few, it is stated that dashboards visually monitor one or more specific phenomena on a single screen and need to provide clearly and in an intuitive manner the information, so the user accessing it with a predefined objective fulfils its purpose.

3.2 User Evaluation

Related to Few’s rules of visual communication when producing dashboards, “to produce functional maps, we need to know about what our visual-cognitive system is designed to do and what is not, about the process by which vision and cognition allow us to derive meaning from visual scenes...” (MacEachren, 1995). Because of this reason, for a proper dashboard design containing geospatial information, user evaluation needs to be performed by conducting scientific studies of the perception of information by the human brain, related to the cognitive cartography field.

Cognitive cartography focuses on how information in a map is perceived by humans. From Robinson to MacEachren & Kraak, many works have been trying to focus on the features of the maps and their design related to cognitive elements. In the first case, Robinson, in *The Look of Maps* (2010, originally published in 1952), describes the elements a map contains and their features, in a detailed way, and how they should be employed for a better display of the cartographic information. In the second case, in *Research Challenges in Geovisualization*, MacEachren & Kraak (2001) focus on the importance of geovisualization to facilitate knowledge through visual exploration and analyses, and therefore how this visual outcome needs to be produced, “how to transform these data into information, and subsequently into knowledge”.

For a better understanding, the term *usability* needs to be explained, as to which aspects can be assessed regarding cartography in the cognitive field. According to Rubin & Chisnell (2008), usability can be tested when related to a service with which “the user can do what he or she wants to do the way he or she expects to be able to do it, without hindrance, hesitation, or questions”. They also state that usability has 6 attributes:

- Usefulness: the degree that enables the user to reach its targets.
- Efficiency: how fast the targets are achieved.
- Effectiveness: the degree of the product being as the user expects and they use it to do what they had in mind.
- Learnability: the ability to operate it with some defined level of competence.
- Satisfaction: the perception towards the product (if it meets their needs, or not).
- Accessibility: having access to a product, and what makes it usable.

Usability needs to be assessed in this case to design more effective and efficient maps or dashboards containing geospatial information, and therefore, the user experience needs to be evaluated. Beyond descriptive analysis regarding map design to facilitate the perception of the user, other methodologies can be employed in order to perform user evaluation and obtain the desired information regarding how the user perceives information when visualising a cartographic product. The usability testing will usually be the result of combining different methods.

Eye-tracking is, among others, a method used in cognitive cartography which records the eye movement and converts it into data that is derived and measured to obtain insights in cognitive processes (Holmqvist et al., 2011). The eye movement is recorded to know where a person is looking at any given time and how the eyes are going from one location to another (Poole & Ball, 2006). Technology (apparatus and software) needs to be employed to perform this method, as well as other parameters that need to be applied to the recorded data, that finally, can be displayed to be analysed in a range of ways. When talking about the technologies used, we refer to eye-trackers as the apparatus to record the eye movement data and software that needs to be combined with this hardware to store the data, as well as the posterior processing of the same.

The *eye movement* term is important to understand how the eye works. As explained by Holmqvist et al. (2011), the light enters through the pupil and projects it to the back of the eyeball where the retina is and converts the light into electrical signals that through the optic nerve are processed. Another element, the fovea, is a small part at the bottom of the eye that allows us to have a full acuity on a small specific area, and therefore allows us to focus on a specific object and see it sharply. The cornea is covering the eye and reflects the light. For eye movement recording, these elements compounding the eye are important: a reflection that the cornea produces is necessary since infrared sensors are part of the technology employed and high-speed infrared cameras are able to detect these corneal reflexions.

It is considered that tracking eye movements provides a dynamic trace of people's attention when interacting with a visual display (Poole & Ball, 2006). Measures can be related to fixation, the time that the eye remains attention to a specific element, such as a graph or a word; or the rapid motion of the eye from one fixation to another, which is called *saccade* (Holmqvist et al., 2011).

A long fixation could reveal a too high amount of processing time, which means the visual display probably needs to be redesigned (Poole & Ball, 2006), as well as a high number of fixations overall, indicate a search is less efficient (Goldberg & Kotval, 1999, from Poole & Ball, 2006). On the other hand, a high number of fixations in a particular area of interest, indicate that this is more noticeable than in other areas (Poole & Ball, 2006). These and other findings are used to evaluate the usability of the analysed visualisation.

The measurements of the eye movement, as in fixations and saccades, lead to the obtaining of analysis products such as AOIs, which consists of a group of fixations (Drusch et al., 2014) and are established according to specific map elements while the selection of the appropriate fixation detection parameters (Krassanakis & Cybulski, 2019). The number of fixations can be represented in heat maps (Golebiowska et al., 2020). These are used to compare users' fixation behaviour and describe visual stimuli since scanpaths are very unlikely to show similar results in two individuals (Drusch et al., 2014). The term *scanpath* refers to the description of a complete saccade-fixate-saccade sequence (Poole & Ball, 2006), or an ordered set of fixation points, which are represented with circles, connected with saccades, represented with lines (Drusch et al., 2014). In other words, trajectories of the eye movements which length show how the person behaves with the visualisation. On the other hand, using AOIs to analyse the recorded information, unlike scanpaths, the order of the fixations is not preserved (Drusch et al., 2014). The representation of the result of both types of measurements is compared in Figure 6.



Figure 6 Areas of interest heat map (left) and scanpath (right) representation.

Combined with other methods, such as using questionnaires or interviews with the users, these measurements offer both quantitative and qualitative data to analyse and obtain conclusions regarding user testing and experience.

Eye movement is more prone to be studied by other fields, but the interest in user experience when interacting with maps brought this method to be used in the cartographic field. A remarkable work related to the usage of eye-tracking in cartography was performed by Jenks in 1971, who was an enthusiast about “getting inside of the map reader’s head” (Steinke, 1987). Jenks used for the first-time eye movement recording techniques to analyse map users’ visual attention by observing regionalised patterns of dots in a dot map, which results were published in the *International Yearbook of Cartography* (Jenks, 1973).

A recent work by Krassanakis & Cybulski (2019) explains the current panorama of the existing eye-tracking studies in the cartographic research field as a literature review, concluding that eye movement analysis is part of the cartographic field and a remarkable number of research studies used eye-tracking technology methods to analyse map reading processes.

As stated, other methods can be employed to evaluate the usability of visual data, including geospatial information. Some studies mixing eye-tracking with other methodologies related to user testing when interacting with map-based visual interfaces have been conducted, also including the dashboard displayed manner. Golebiowska et al. (2020) explore the user interaction in a coordinated and multiple views (CVM) geovisualization tool with learning purposes, combining eye-tracking and the talk-aloud method. Participants are requested to talk about their impressions and opinions while performing the experiment. The conclusions relate to the elements that get more attention (e.g., the centre of the map) by analysing the AOIs and the fixation duration.

Similarly, Popelka et al. (2019) analyse graphical user interfaces (GUIs), which include similar visualisation elements to dashboards, such as interactive charts, tables and maps. The study focuses on the capability of the GUIs to be user-friendly by checking the user’s accessibility and finding their shortcomings. The methodology used is eye-tracking in two phases: one during which the user freely explores the interface; and another during which the user has to solve proposed tasks. The last task is combined with the possibility of the participants to verbally express their opinions. The results for the first phase focus on analysing the three first fixations and the results for the second allow obtaining insights about the shortcomings regarding the success/failure in the task solving.

These both lead to conclusions such as that the user is more prone to look to the left side, so important elements should be placed on that side; the point symbols should be interactive and mouse-drag and mouse-click should simultaneously be supported, among others.

The dashboards and user interaction are explored in studies such as *Eye-tracking Study of the Line Charts in Dashboards Design* by Orlov et al. (2015). It consists in analysing the user interaction related to the number of graphs displayed in a dashboard, using eye-tracking methods. It is focused on two research questions: if the number of graphs influences the fixation duration, the number of fixations and the time spent to solve the proposed tasks; and the same for the type of graphs. It concludes with no impact on the type of graphs, but the number of graphs alters the fixation duration: the more graphs are displayed, the lower is the fixation time. So, conclusions suggest that when designing dashboard interfaces, the designer should think about the number and the size of the graphs depending on the level of detail that is needed to display specific information to carry out specific tasks.

A Feasibility Study of Map-Based Dashboard for Spatiotemporal Knowledge Acquisition and Analysis, by Zuo et al. (2020), even includes the geospatial dimension in its research. They propose their own design model of a map-based dashboard with a methodology mixing eye-tracking and interviewing for the analysis of user experience. It is focused on five themes: a free-exploration stage to analyse the fixation position to identify the AOIs; solving tasks to evaluate the effectiveness of the dashboard, by the correctness of the answers, and the efficiency, by the task solving time; the return to AOIs; as well as the transition between the elements. The study leads to conclude which specific elements of the layout should be changed or improved, such as the font size, or the arrangement itself.

4 EXPERIMENT I

4.1 Methods

The first experiment is a formative study to obtain insights regarding the problematic aspects of the selected existing dashboards and elaborate recommendations for future self-developed ones. Thanks to user testing, the usability of these dashboard interfaces can be assessed in terms of effectiveness, efficiency, and satisfaction, according to Rubin & Chisnell's (2008) definitions.

Experiment Design

To begin with, the experiment needs to be designed to meet the requirements in order to achieve the desired goals. This is done with SMI Experiment Center 3.7 and it consists in the following steps:

- Calibration: to check that the deviation of the visual angle is under 1° .
- Introduction: including information about the experiment to the participant.
- Task solving: the participants must solve twelve tasks along four dashboards (three each) that appear in a randomised order.
- Interview: the participants express their opinion about the positive and negative assets of each dashboard.
- Questionnaire: to obtain information about the participants' field background and dashboard experience.
- Acknowledgement: thanking the participant and finalising the experiment.

The eye movement is recorded, as well as the participant's voice and image with a camera and a microphone, particularly important for collecting data during the interview. A summary of the procedure is displayed in Figure 7.

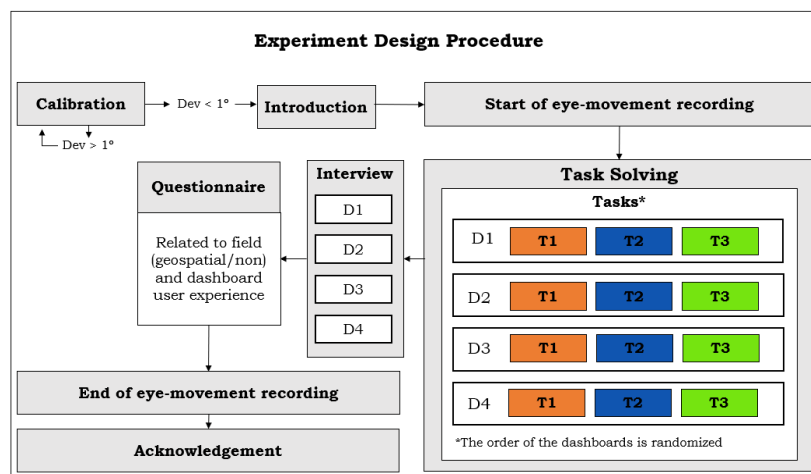


Figure 7 Experiment I design procedure.

Stimuli and Tasks

According to the definition of the term *dashboard* by Few (2006), these are interfaces displaying information at-a-glance. Because of this, the four existing dashboards selected for the stimuli are interfaces meeting these characteristics. Moreover, the dashboards must have different functionalities and geospatial visualisation methods.

Three tasks (T1, T2 and T3) that consist in finding an answer to a specific question and involve interacting with the dashboard interfaces are assigned to each of them, so in total there are twelve tasks. From task 1 to 3, the level of difficulty increases: T1 is the simplest and always refers to finding the total COVID-19 cases by last update in a specific region; T2 and T3 focus either on analysing different metrics through a specific period or region, or on using specific functionalities from the dashboard. The selected dashboards and the tasks for each of them are explained in the following lines.

The COVID-19 Map by the John Hopkins University³ (D1) displays a world map with graduated symbols and pop-ups, and other widgets such as: numeric metrics, a sidebar with a country list and graphs. These are compounded by tabs that lead to obtaining regional/temporal data. It is visible in Figure 8 and the proposed tasks are the following: total accumulated cases in Switzerland by last update (T1); daily cases in France on 25/10/21 (T2); and weekly cases in California (US) during the week of the 03/01/21 (T3).

The Novel Coronavirus Incidence Map by the University of Washington⁴ (D2) visualises country data on a choropleth map, as well as a sidebar with the dashboard's purpose information, and graphics with daily numeric metrics that change when selecting a country. There is the option to switch to a situational heat map, which depicts the trend of cases in countries by the last update, as visible in Figure 9. The tasks to be solved are: total aggregated confirmed cases in Germany by last update (T1); aggregated confirmed cases in Russia on 25/02/21 (T2); and check if the situation regarding the evolution of cases, from 07/06/21 to 27/06/21, in Norway, has improved or worsen (T3).

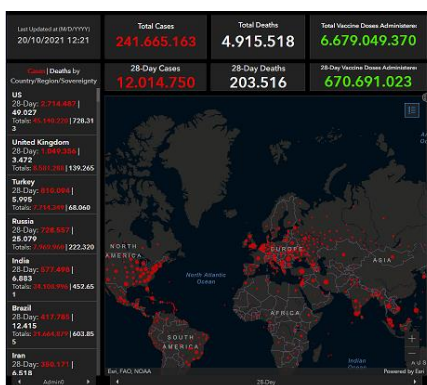


Figure 8 D1, COVID-19 Map (John Hopkins Coronavirus Resource Centre, n.d.).

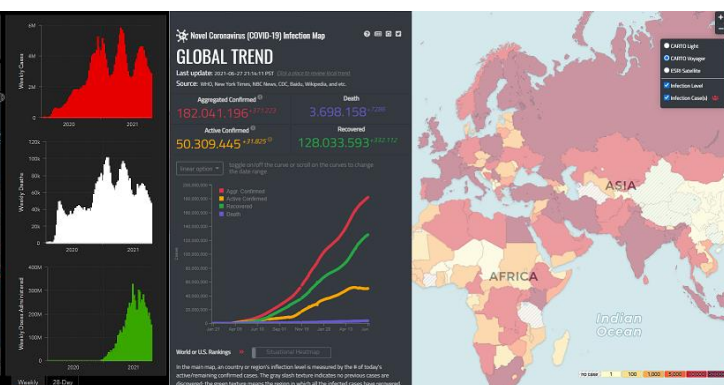


Figure 9 D2, Novel Coronavirus Incidence Map (University of Washington, n.d.).

³ Dashboard available in the following link: <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>
⁴ Dashboard available in the following link: <https://hgis.uw.edu/virus/>

The OCHA COVID-19 Data Explorer⁵ (D3) combines a map with graduated symbols and choropleths that displays pop-ups, together with a wide range of variables to choose on a sidebar. Displayed in Figure 10, other functionalities are viewing the metrics of a specific country by the *filter by country* option or choosing up to five countries to compare their metrics, including numeric metrics and graphs. The graphs are visible when changing the tab from *Map View* to *Chart View*. These tasks should be fulfilled: total accumulated cases in Cameroon by last update (T1); which country has more weekly new cases between Brazil, Argentina, and Bolivia (T2); and the number of accumulated cases in Ethiopia by 10/07/2021 (T3).

Finally, the Health Map⁶ (D4) includes the total number of cases per colour-coded dots (see Figure 11). There is a sidebar with a country list and a search option, and some symbols display pop-ups. All these are combined with a time-slider that allows us to see how the virus spreads through 2020. The proposed tasks are the following: accumulated cases in Italy by last update (T1); find out total accumulated cases in Nicaragua by last update (T2); and number of Cases in Melbourne (Australia) on the 05/03/20, and on the 05/04/20 (T3).

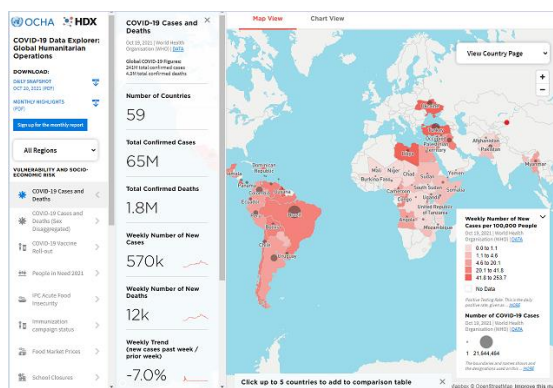


Figure 10 D3, COVID-19 Data Explorer (OCHA, n.d.).

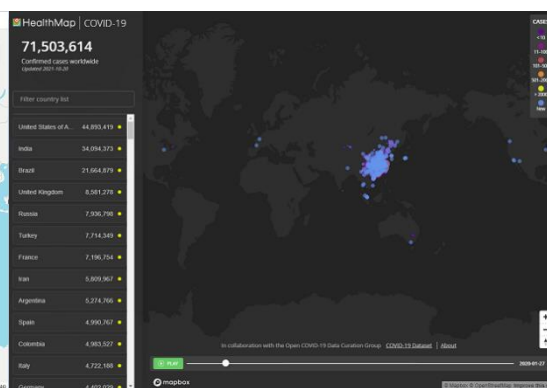


Figure 11 D4, Health Map (Health Map, n.d.).

Apparatus

The employed apparatus to record the eye movement is the eye-tracker SMI RED 250, which records at a frequency of 250 Hz, in the eye-tracking laboratory of the Department of Geoinformatics of the Palacký University Olomouc, Czechia. As mentioned, a camera and a microphone (Logitech C920) are also employed, and SMI Experiment Center 3.7 is the software used to design the experiment, as well as to record it, combined with iView X. The stimuli are displayed in a Google Chrome browser on the participants' 24-inch screen.

Participants

Between the 18th and 27th of December (2021), experiment I is performed with nineteen participants, university students or recent graduates.

⁵ Dashboard available in the following link: https://data.humdata.org/visualization/covid19-humanitarian-operations/?ga1&gclid=EAlaIqobChMI3aCi08Kk8wIVxQwGAB11fgTgEAAAYAiAAEgJl0fD_BwE

⁶ Dashboard available in the following link: https://www.healthmap.org/covid-19/?mod=article_inline

Slightly more than the half (58%) are members of the geospatial domain field; 79% of them use dashboards in their daily life, in different frequencies: mostly once a week (40%), but also once a month (33%), or hardly ever (27%). The 21% remaining does not use dashboards. As Google Trends indicates (see chapter 3) the terms *COVID-19* and *dashboard* have increased their popularity since the outbreak of the pandemic: 58% of the participants used dashboards before it and 42% did not. The participants' overview is summarised in Figure 12.

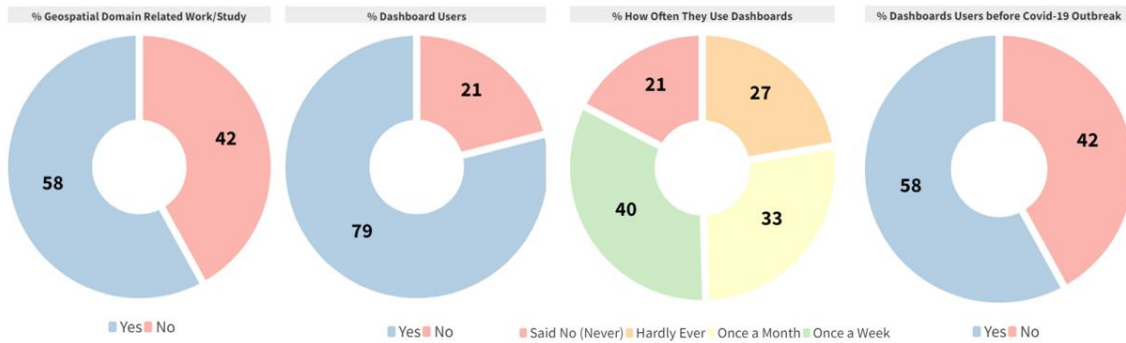


Figure 12 Experiment I participant's overview.

To identify the ideal number of participants for this study, the Sample Size Calculator for Discovering Problems in a User Interface for Measuring U⁷ is employed. From the sample data, it estimates the problem occurrence (p) using the Good-Turing and Normalisation procedure by Turner et al. (2006). For this study, it is estimated how many respondents would be appropriate to detect at least 85% of the problems encountered. Considering twenty-seven problems encountered, the result is thirteen. Nineteen participants are selected, which is higher than the recommended number. This information's summary is in Figure 13.

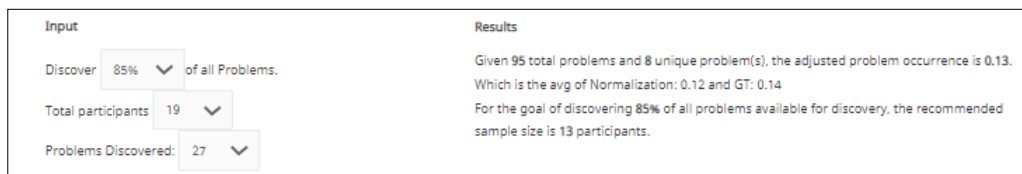


Figure 13 Participants number estimation (Sample Size Calculator for Discovering Problems in a User Interface for Measuring U, 2022).

As displayed, some of these problems are unique (eight), and others are more generalised. A lot of participants encounter problems to find specific functionalities in the dashboards that are required to solve a task. For D1, these include finding the graphic that refers to the weekly/daily cases, as well as the tabs to change from one administrative level to another one; even some participants have problems with the display of the pop-up when clicking the dot that refers to a specific region. For D2 the hovering option on the graph makes it hard to obtain the specific date, together with the fact that almost any participant uses the situational heat map to assess T3.

⁷ Tool available in the following link: https://measuringu.com/calculators/problem_discovery/

D3's functionalities are not intuitive either, many participants are not able to find the possibility to change from the view displaying the map to the view displaying the charts and see the evolution of the data, and neither the comparison table. Finally, D4's list of countries with a number referring to COVID-19 cases next to them is confusing to some participants and, even though the time-slider provides interesting information, it causes problems in its usage in most of the cases.

Data Pre-processing

Once the experiment is performed the data obtained needs to be processed. The metrics evaluated are not dependent on the tracking ratio and provide valuable information to obtain insights for the study, both quantitative and qualitative, so there is no need to exclude any data.

The fixations and the saccades need to be identified. For that, an algorithm is applied, in this case, the I-DT algorithm, used for data measured at 250 Hz and below (Holmqvist et al., 2011), by applying two thresholds: one for the minimum fixation duration and one for the maximum dispersion. The optimal fixation detection settings for cartographic research in SMI BeGaze 3.7 are 80 milliseconds for minimum fixation duration, and 50 pixels for maximum dispersion, according to Popelka (2014).

The recorded eye movement and screen monitoring need to be processed in a way that allows the data to be separated into different individual tasks. SMI BeGaze 3.7 allows the segmentation of video recordings into what is called *custom trials*. These are used to analyse eye-tracking metrics in the general context of the twelve tasks, but cannot be used to analyse each participant's behaviour, for which we establish AOIs, that are manually created by using the AOI editor in SMI BeGaze 3.7. Since each dashboard is composed of different elements, each has different AOIs layouts.

Methods of Analysis

Several methods are used to analyse the processed data. The correctness rate of the task solving evaluates the effectiveness, whether the participant can reach the targets (answer correctly) or not. The trial duration, or the time needed to solve the tasks, visualised in boxplots created in R studio, evaluates the efficiency, or how fast they solve the tasks (Rubin & Chisnell, 2008). A longer trial duration means the user interface has problems, or the complexity of the task is high (Popelka et al., 2019). The scanpath length of each task could also be analysed, because if it is long (a higher number of pixels) indicates it is not efficient to navigate and search elements through an interface (Goldberg et al., 2002), in this case, a dashboard. It is highly correlated to the trial duration, so its analysis is going to be omitted since it shows the same metrics and leads to obtaining the same insights.

The AOIs are analysed with a sequence chart per task (twelve), which shows the fixation time of each participant around different elements of the dashboard. This enables us to see how the participants behave when interacting with the interfaces. Each dashboard has different elements, so the AOIs are going to be different in each case. These are established the following way:

- For D1, the AOIs are divided into: the title, the date, the list of countries sidebar, the numeric metrics, the map field and the graphs. This is visible in Figure 14.

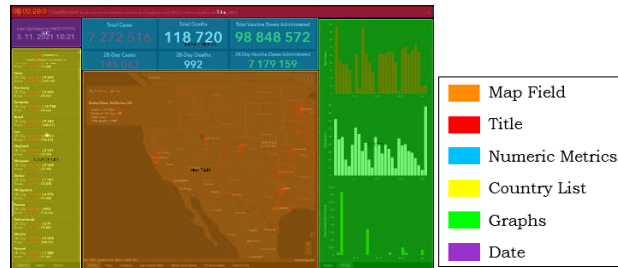


Figure 14 Areas of interest of D1.

- For D2, the AOIs show: the title, the numeric metrics, the graph, the sidebar with explanations and the map field (see Figure 15).



Figure 15 Areas of interest of D2.

- For D3, there are four AOIs: the title, the sidebar, the tabs to change from map field to graph field, and the map field by default, which can become the field displaying the graphs, as visible in Figure 16.

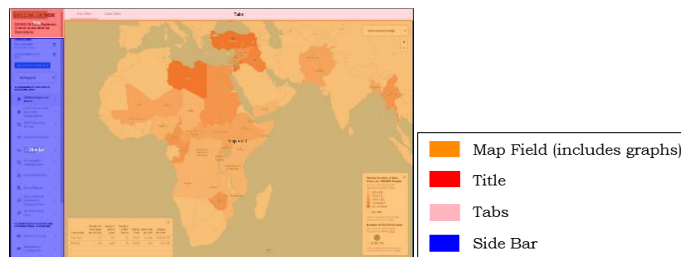


Figure 16 Areas of interest of D3.

- For D4, the AOIs are: the title, the numeric metrics, the search bar, the sidebar with the list of countries, the map field and the time-slider (see Figure 17).



Figure 17 Areas of interest of D4.

As explained, the experiment also contains an interview and a questionnaire. The interview compiles a table with the positive and the negative assets mentioned, in order to measure their satisfaction or perception towards the elements of the interface.

Finally, the questionnaire includes questions to also analyse the satisfaction, regarding the confidence of the user when solving the tasks to test their usability, as well as if they find them aesthetically nice; together with questions regarding their academic/professional background (geospatial or not) and their experience interacting with dashboards, which results are summarised in the participant's overview (Figure 12, Page 24).

4.2 Results

4.2.1 Correctness

The first results analyse the success rate of solving the tasks. In each dashboard, the questions go from easy to hard. To avoid that it does not negatively affect the internal validity of the experiment, the order of appearance of the dashboards is randomised. An example is that when the first stimulus appears, the participants might answer wrong because they are not used to the dynamics of the experiment, and vice versa, after a few questions the participants are familiar with it.

The results (Table 1) show that the first question of every dashboard is in most cases correctly answered. In D4's case, T2 is very similar to T1, so it is showing 100% of correctness, being the only dashboard that does not show a descending pattern on answering correctly the answers from T1 to T3. The last of the questions of each dashboard is always showing the lowest correctness rate since it is the hardest. When comparing dashboards, D1 is the one with less success rate, having even no one able to solve the third task. This relates to the necessity to access the tabs to change the administrative levels of the countries and the period (weekly or daily), which proves to be not intuitive for the participants, who are unable to find the right way to assess the task. D2 has the highest percentage in the number of correct answers, while for D3 more than half of the participants correctly answer each question; and D4T1 and D4T2 have an almost perfect success rate (89% and 100%), but D4T3 has only 37% of success rate since it refers to a functionality of the dashboard that is not intuitive at first sight, the time-slider.

Table 1 Correctness of the task solving.

	D1						D3					
	T1	%	T2	%	T3	%	T1	%	T2	%	T3	%
Correct	16	84%	2	11%	0	0%	16	84%	14	74%	10	53%
Incorrect	3	16%	17	89%	19	100%	3	16%	5	26%	9	47%
	D2						D4					
	T1	%	T2	%	T3	%	T1	%	T2	%	T3	%
Correct	19	100%	17	89%	15	79%	17	89%	19	100%	7	37%
Incorrect	0	0%	2	11%	4	21%	2	11%	0	0%	12	63%
TOTAL							19	100%	19	100%	19	100%

4.2.2 Trial Duration

From T1 to T3 in each dashboard, the difficulty tends to increase, and so does the number of incorrect answers in most of the cases, as explained before. Considering this, the trial duration's length should increase as well, but it is not always the case.

The median of time required in T2 of D1 and D2 is higher than the one in T3. In the first case because for D2, the participants need to look for a specific date by hovering on a graph, which is hard to maintain still; and in the second case because T3 is similar to T2: in both cases participants need to look for a graph date, and they are already familiar with the procedure.

D3 follows an increasing tendency of the trial duration time, but again D4 does not since the median of T1 is slightly higher than T2, but T3 is the highest. In the case of D3, the first question is about finding the accumulated COVID-19 cases of a specific country, a number that is displayed in a pop-up when hovering, but in small letters and not highlighted, which makes it difficult to spot. As previously, the case of the D4 can be influenced by the fact that T2 is very similar to T1, and participants already know the procedure to obtain the answer. In general, D1 is the dashboard that requires to invest more time to solve the tasks (total of 3:39 minutes), whereas D2 is the one that takes the least (1:50 minutes).

For the statistical analysis, with the usage of the Shapiro-Wilk test, the normal distribution of each value can be analysed. Most tasks' values show that they do not have a normal distribution. For this reason, a non-parametric test should be performed, in this case, `kwAllPairsNemenyiTest` (from R's package `PMCMRplus`). In the first dashboard, there is a statistically significant difference between T1 and the rest, which means the first task is easier than the others. The opposite happens in D3 and D4, T3 statistically differs from the rest, because it is the hardest task. The second dashboard shows the most statistically similar values, with a slight difference between T1 and T2, of $p = 0.003$. The boxplots in Figure 18 show this information.

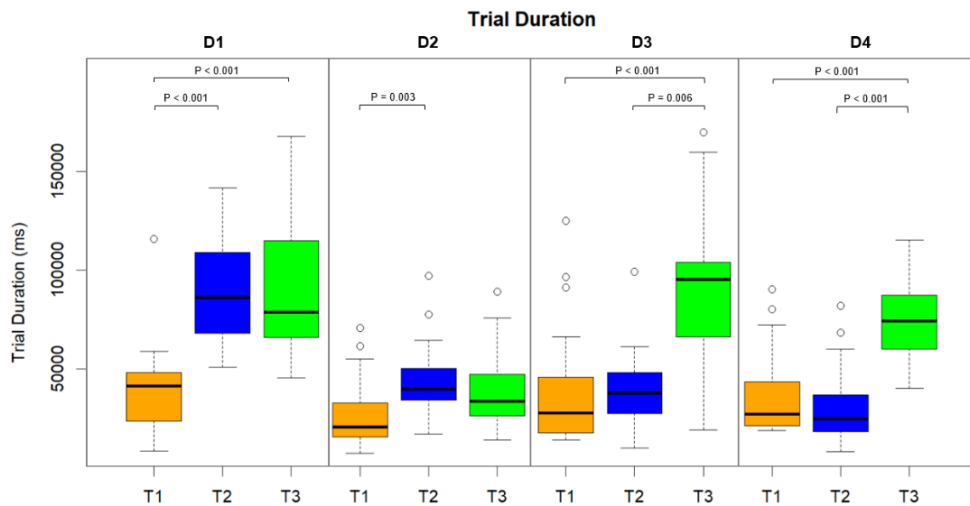


Figure 18 Trial duration of each task.

As mentioned, the trial duration is highly correlated with the scanpath length of each task, and therefore this metric is avoided for analysis. The usage of a correlation test provides more than 0.5 correlation values in each case.

To do a comparison between dashboards, it is used the same non-parametric test to analyse the statistically significant similarities/differences between T1, T2 and T3 along the dashboards.

For T1, it is visible that D1 has the highest median, while D2 has the lowest, but there is no statistically significant difference. For T2, there is a statistically significant difference between the first dashboard and all the rest, which took longer, and therefore might have been the hardest. Finally, T3 shows the opposite to T2, and D2 shows statistically significant differences from all the other dashboards, but has a very low value of trial duration, which indicates it might have been the easiest. This is visible in Figure 19.

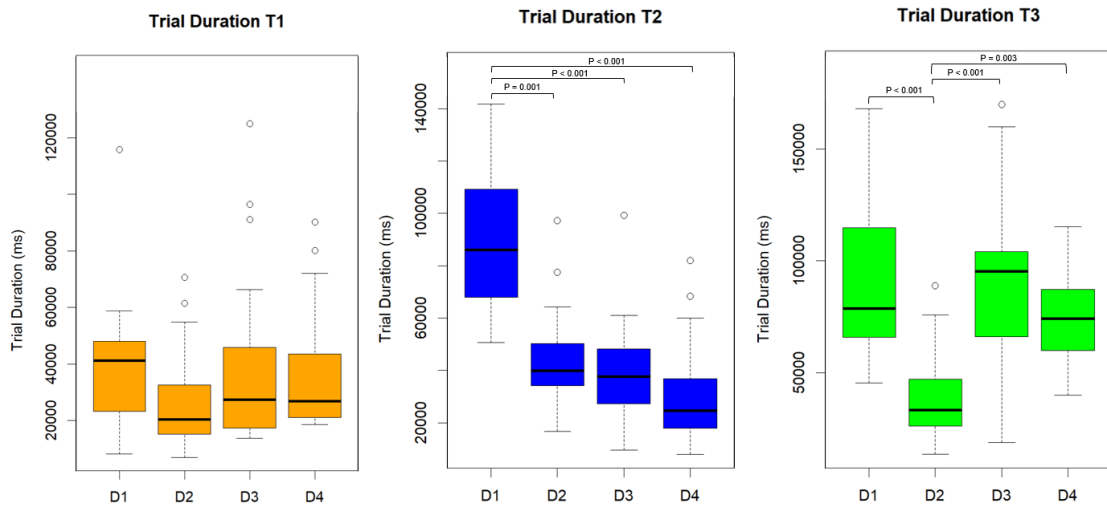


Figure 19 Trial duration of each task per dashboard.

4.2.3 AOIs

The sequence charts show the fixation time of the participants on each of the AOIs and enable to see their behaviour during the task solving process, in terms of which elements of the interface they focus more on. Depending on the proposed activity, the participant focuses more on one or another, but there are outstanding areas of interest no matter what. The way AOIs are established is detailed in the subchapter 4.1 and, as mentioned, they are different for each dashboard, since each dashboard's composition differs.

For the first dashboard, the map field is very predominant in the three tasks, but since T2 and T3 require using the charts, the time spent focusing on this element is very high, especially in T2 (see Figure 20), because in T3 they are slightly familiar with them. The country list is also visualised very often since the tasks include looking for a specific region. On the other hand, some participants focus on the numeric metrics during the first task, but these are not present in the others, together with the rest of the elements during the entire time: the date and the title.

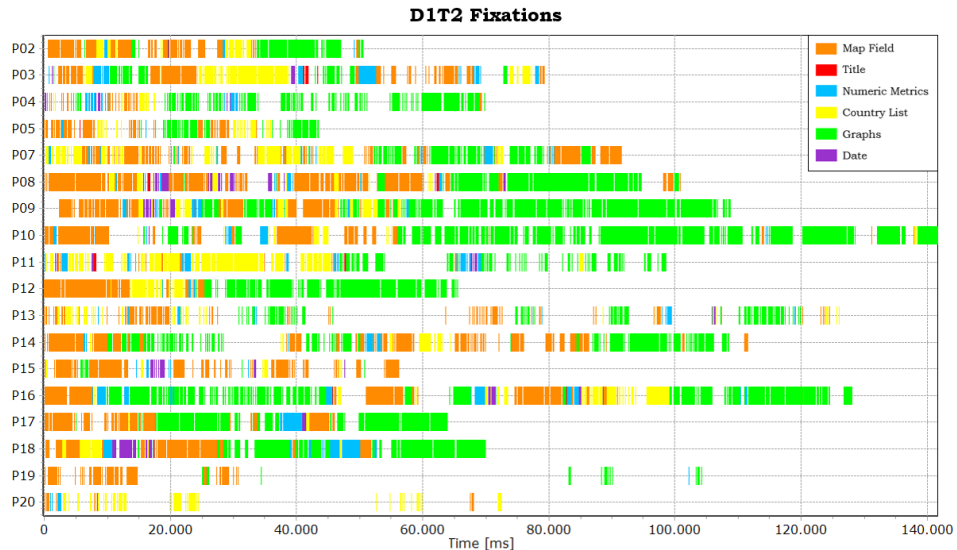


Figure 20 D1T2 Fixation time to AOIs per participant.

A similar pattern is observed in D2, the map field and the graphs take a lot of fixations from the participant, especially in T2, when the questions require checking the graphs for an answer; and in T3, even though the graph was not the element to consult to answer this question. The question asks if the situation improves or worsens in Norway through a specific period, and the correct way to do it is by clicking on the tab *Situational Heat Map*, but the majority use the graph.

Instead of the country list, the sidebar is often checked because it contains a big text with explanations and numeric metrics, because, unlike D1, there are no pop-ups, and the numeric metrics change when selecting different countries and provide the information regarding the COVID-19 cases. Again, the title is highlighted on a few occasions. As mentioned before, it is hard to find a specific date in the graph, because it uses a hovering system, and that might have an impact on the duration of the trial because it takes time to keep the mouse in an accurate position. Also, the fact that the sidebar needs a long time to be read does not necessarily have to be a positive element, being the big block of text not user-friendly.

D3 is the one with more predominance of fixation time towards the map field. That is explained by the fact that the graphs are also included in the same area (when switching to viewing the charts), together with the pop-ups, the legend, and the comparison table among countries. Therefore, all the tasks involve focusing on this area.

The sidebar also has some attention from the participants, since it has a lot of tabs to check different variables and change parameters. Nevertheless, the questions are simplified and it is not required to check these and a longer time spent in that area might be the result of checking unnecessary tabs. As in the other cases, the title does not receive much fixation time and neither does the added element in this dashboard, which is a tab to change from the map to the charts. Most of the participants are not even able to notice it and cannot interact with the graphs to answer T3, as visible in Figure 21.

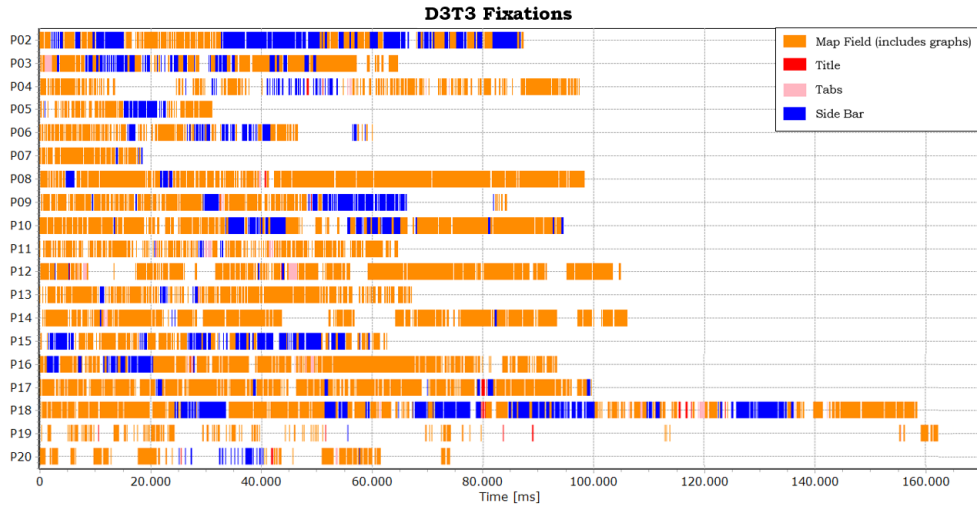


Figure 21 D3T3 Fixation time to AOIs per participant.

Finally, the way users interact with the fourth dashboard has its particularities. The map field plays a major role again, and so does the countries list, since it is a tool to look for the regions asked. An interesting tool in this dashboard is the search bar option, which also takes the participant's attention, in a very short time frame, but that they make use of to optimise their searches.

Another added feature in this dashboard is the time-slider. T3 involves comparing the COVID-19 cases between two dates and this element receives a lot of fixations during its solving, but also during T1 and T2 to a smaller extent. So are the numeric metrics, even though it consists of a static number of the cases worldwide. Since there are no informative pop-ups, perhaps the users with the experience of the numeric metrics changing when clicking regions (it is the case of D2) expect the same function. Finally, the title is only perceived in a very short time, like in the other cases, and is barely noticeable in the sequence charts. All these phenomena are visualised in T2, displayed in Figure 22. The rest of the sequence charts from the other tasks can be found in the Attachment 1.



Figure 22 D4T2 Fixation time to AOIs per participant.

4.2.4 Questionnaire and Interview

The questionnaire and the interview provide information regarding the subjective opinion of the participants. First, they rate how difficult it is to solve the tasks of each dashboard from 1 (very difficult) to 5 (very easy). The results are in Table 2.

Table 2 Difficulty rate.

	D1	%	D2	%	D3	%	D4	%
Very Difficult	3	16%	0	0%	1	5%	0	0%
Difficult	6	32%	3	16%	2	11%	5	26%
Ok	5	26%	3	16%	7	37%	6	32%
Easy	5	26%	7	37%	6	32%	6	32%
Very Easy	0	0%	6	32%	3	16%	2	11%
TOTAL	19	100%	19	100%	19	100%	19	100%

A higher number of participants rates D1 as the dashboard with the most difficult tasks, and for no one is considered very easy. This contrasts with completely the opposite situation for D2. D3 has very equilibrated results, tending to the easy side, and D4's results are closer to the ok rate because even though the correctness rate is high in T1 and T2, T3 is not intuitive.

The participants are also asked to rate aesthetically the dashboards (results in Table 3) from 1 (very ugly) to 5 (very nice). This question is very subjective since, for example, some participants appreciate a dark background with colourful symbology (D1 and D4), but others prefer the simplicity of a clear background (D2 and D3). Nevertheless, D2 shows the highest rates of aesthetic approval, being also the easiest dashboard to solve the tasks. Then D3 is closer to being considered very nice, followed by D1 and finally D4. This means, in general, lighter and simpler dashboards are better aesthetically rated.

Table 3 Aesthetics rate.

	D1	%	D2	%	D3	%	D4	%
Very Ugly	0	0%	0	0%	0	0%	1	5%
Ugly	4	21%	1	5%	1	5%	5	26%
Ok	4	21%	3	16%	7	37%	3	16%
Nice	7	37%	9	47%	7	37%	5	26%
Very Nice	4	21%	6	32%	4	21%	5	26%
TOTAL	19	100%	19	100%	19	100%	19	100%

The interviews also show preferences and suggestions worth considering for the recommendations. The following tables compile the good and the bad assets regarding each dashboard, ordered according to how many (from high to low) people mention them. The most significant ones are taken into account for this analysis. Regarding D1, the highlights are in the negative aspects, since almost half of the participants mention the fact that there is no search option (47%), and the accessibility of the tabs to change regions or time are not easily identifiable (42%). Excess of information and symbology (graduated symbols) are also mentioned as a negative element in a smaller percentage (21% in both cases). Some positive elements are the graphs, numeric metrics and aesthetics, among others, but mentioned by a very small percentage of the participants (lower than 16%). All this information is summarised in Table 4.

Table 4 Positive and negative assets of D1.

D1					
Positive Assets	Participants	%	Negative Assets	Participants	%
Graphs	3	16%	No Search Option	9	47%
Numeric Metrics	3	16%	Tab Accessibility	8	42%
Aesthetics	2	11%	Excess of Information	4	21%
List of Countries	2	11%	Symbology	4	21%
Pop-ups	1	5%	Graphs	3	16%
Symbology	1	5%	Aesthetics	2	11%

The second dashboard's positive elements are more highlighted, as visible in Table 5. Almost half of the participants mention that the symbology (choropleth) is a positive asset (47%), as well as the graph and the aesthetics by 32%. These are commented on the negative side only by 11% of participants, together with other assets in a very low ratio. Only the lack of a search option and the fact that the sidebar occupies a big space in the interface, and contains a lot of text information, are mentioned as negative elements by 26% of the participants.

Table 5 Positive and negative assets of D2.

D2					
Positive Assets	Participants	%	Negative Assets	Participants	%
Symbology	9	47%	No Search Option	5	26%
Graphs	6	32%	Sidebar	5	26%
Aesthetics	6	32%	Aesthetics	2	11%
Legend	1	5%	Graphs	2	11%
			Lack of Pop-Ups	1	5%
			Tab Accessibility	1	5%
			Lack of Information	1	5%

D3's most positive aspect is its aesthetics, pointed out by 32% of the participants, and its sidebar, with a lot of options to choose from, by 21%. Other elements mentioned with a lower frequency are the fact that pop-ups appear when hovering and the countries can be filtered, among others. The negative aspects are headlined by the excess of information according to 32% of the participants, and the not-so-easy tab accessibility to change the parameters by 26%. Again 21% point out the lack of a search option. All these comments are gathered in Table 6.

Table 6 Positive and negative assets of D3.

D3					
Positive Assets	Participants	%	Negative Assets	Participants	%
Aesthetics	6	32%	Excess of Information	6	32%
Sidebar	4	21%	Tab Accessibility	5	26%
Pop-ups when Hovering	2	11%	No Search Option	4	21%
Tab Accessibility	2	11%	Sidebar	2	11%
Filter by Country Option	2	11%	Pop-ups when Hovering	2	11%
Comparison Table	1	5%	Aesthetics	1	5%
Legend	1	5%	Legend	1	5%
			Filter by Country Option	1	5%

Finally, Table 7 shows the results of the fourth dashboard. The search option, unlike the other dashboards, is much appreciated by 68% of the participants. The aesthetics are also positively mentioned in 21% of cases, as well as the time-slider option. This is a tricky element since almost half of the participants mention that it is a negative element (47%), saying it is a good idea, but that it is not properly implemented.

The symbology (dots) is also agreed as a non-positive element by half of the participants (53%), together with the excess of information or aesthetics by a lower rate, 26% and 16% respectively. 16% also point out the list of countries as a positive asset.

Table 7 Positive and negative assets of D4.

		D4			
		Participants	%	Participants	%
Search Option		13	68%	Symbology	10 53%
Aesthetics		4	21%	Time Span Option	9 47%
Time Span Option		4	21%	Excess of Information	5 26%
List of Countries		3	16%	Aesthetics	3 16%
				Legend	1 5%

4.3 Conclusion

The conclusion consists in summarising the results to identify the problematic aspects regarding each dashboard and their elements, and elaborate recommendations for an improved version.

Regarding correctness, the results show that the first dashboard is the hardest and the second dashboard is the easiest. Trial duration relates to the difficulty of the tasks, and the more complex they are, the longer participants need to solve them. There is a common tendency of the first task of each dashboard taking shorter than the last, except in the cases where the previous question enables participants to be familiar with the procedure and answer quicker. When comparing the tasks themselves, between the dashboards, it is visible that for the second task D1 has a statistically significant difference from the rest for showing very high trial duration values, indicating it is the hardest, and the opposite situation happens to D2 in the third task, indicating it is the easiest.

This can be related to the number of elements available to interact to find the information and their accessibility, regarding if they are visible or not. D1 has a lot of elements and makes it hard to find specific information, opposite to D2. D3's situation is similar to D1, but it appeared to be more intuitive to participants, having a higher correctness rate and a lower trial duration. D4 has a similar situation than D2, but the time-slider functionality needed to answer T3 is not user-friendly and decreased the correctness rate and increased the trial duration.

The AOIs show the importance of the map field in all the cases, and how the nature of the question can change which areas are more consulted. The graphs and the list of countries are noticed very often and used to carry out the proposed exercises. The frequency of the numeric metrics varies considering if these are interactive or static, being static a less useful resource. The specific elements of each dashboard, such as the search option and the time-slider in D4 are present in the fixations of the participants, not only because they are required for the task, but because of their usability. Elements like the title, credits and explanatory sidebar are just perceived quickly and, in the last case, it can be a roadblock rather than a helpful element.

An aspect to mention is that D2T3 involves clicking a tab to answer the question regarding if the situation in Norway improved or worsened. The majority do not realise the presence of this tab, which is hidden in the big text block in the sidebar and try to answer the question by looking at the graph. The fact that participants do not use it, does not affect the correctness rate, showing it might not be a necessary functionality, and even if it was, it is not properly noticeable.

The subjective answers regarding the difficulty of the dashboard put D2 as the easiest and D1 as the hardest, which coincides with the correctness rate and the trial duration insights. D3 is considered to be easier, and D4 in the ok rate. The aesthetic rating shows that the lighter dashboards are rated better, even though some participants prefer dark.

The good and bad assets show similar insights. While D1 tends to have more percentage of bad aspects, D2 has the opposite situation. Some negative elements that all dashboards have in common are the following: only D4 has a search option, therefore there are complaints about the rest of the dashboards; tabs accessibility is also a commented issue, especially in D1 and D3 cases, as well as the excess of information; that is why D4's symbology is criticised because it includes a lot of overlapping dots that do not directly visualise the phenomena, together with the time-slider option. This option appears to be a good idea in some cases, but majorly considered not properly implemented. Other good ideas refer to: symbology when the choropleth is used in the map section, light aesthetics, graphs, interactive numeric metrics and sidebars that include lists of countries.

Therefore, according to what both quantitative and qualitative data results show, the following recommendations are formulated. A proper dashboard should include light, simple aesthetics with a choropleth map, a country list sidebar with a search option and numeric metrics that interact with the desired requested information (instead of pop-ups). In case of graphs and a time-slider showing temporal data, they should be user friendly and noticeable; a concise, clear title is necessary, as well as to avoid big blocks of explanatory text. On the other hand, less appreciated elements are darker colours and graduated symbols or dots, static numeric metrics and hard to use/find functionalities. These recommendations are summarised in Table 8.

Table 8 Recommendations for good and bad dashboard design elements.

	✓	✗
Map Field	Choropleth	Dot/Graduated Symbols
Numeric Metrics	Interactive	Static Click and Pop-ups
List of Regions	Yes	No
Search Option	Yes	No
Graph	Yes	No
Date Selector	Choose Option	Time-slider
Tabs	Visible	Non-visible
Title	Big and Clear	Small
Credits	Small	Big
Aesthetics	Light	Dark
Block of Text	No	Yes

These characteristics are used to develop two dashboards. Testing these two dashboards, designed especially following or avoiding these recommendations, leads to final conclusions regarding if these elements contribute or not the user-friendly level of a dashboard.

5 DASHBOARD DEVELOPMENT

5.1 Methods

Following the thesis topic, the self-developed dashboards for experiment II cover COVID-19 cases. COVID-19 global pandemic still needs to be monitored by indicators such as the incidence rate (cases per 100 000 inhabitants) to obtain insights into the situation of the health crisis to take action in its management. In this case, it shows a specific study area, Catalonia (*Catalunya*).

For the development of these dashboards, different steps need to be followed to proceed and build an SDI, with the usage of different standardised tools to communicate this information.

5.1.1 Data Sources

The data source is the Catalogue of Open Data of the Government of Catalonia which provides:

- Data in CSV format for the daily cases for two administrative levels (counties and towns), from the 1st of March 2020 up until the 26th of February 2022.
- The geometry for both administrative levels (shapefile format)
- The population in the year 2020 of the towns, which is used to calculate the COVID-19 incidence rate, in CSV format.

The counties' population data is from another source, the Statistical Institute of Catalonia, in CSV format. Once the information is obtained, the data retrieving, and cleaning process needs to be carried out.

5.1.2 Technologies

To integrate the elements in the different parts of the SDI, several tools need to be employed, these are the following:

The back-end consists of:

- Linux server with Python scripts running.
- PostGIS: the spatial database extension of PostgreSQL that allows the storage of geographic objects.

The middleware is implicit in Tableau, which has a framework allowing the communication between the server and front-end user interface.

The front-end is displayed using:

- Tableau: a visual analytics platform (Tableau, 2022) with a desktop application, which retrieves data straight from the database and allows creation of user-friendly and responsive dashboard interfaces at-a-glance.
- Canva: to design the background of the dashboards.

5.1.3 Implementation

It is necessary to understand how the data sources and the technologies described are integrated in the SDI architecture and their roles through the data flow.

A Linux server is the back-end of the infrastructure which supports all the processes happening: a Python script handles the automatic data retrieval from the data source (the CSV from the official API containing COVID-19 information), the data cleaning and manipulation (calculation of the metrics, in this case, the incidence rate). The shapefiles with the geometry are converted to GeoJSON and uploaded to the server to be used in the Python script. Pandas is used to read the CSV data in a tabular format and perform all necessary manipulations mentioned. The geographic dimension of the data is handled with Geopandas.

The connection to the database is done via SQLAlchemy and GeoAlchemy, which synchronises the resultant data frames after manipulation to PostGIS. Four tables containing all the relevant information to be displayed in the dashboard are pushed and stored to a database that is automatically updated daily: two with the geometry of each administrative level and two with the COVID-19 metrics for each. The first two contain a column with the INE code (*codiine*, the code that identifies towns on a national level), the name of the town/county (*nom_muni/nom_comar*) and the geometry (*geometry*). Figure 23 shows the table containing this information for the towns and the geometry visualisation option of PostGIS. The second two tables, with the COVID-19 metrics, include the INE code again (*municipicodi*), together with the code to identify the counties (*comarcacodi*), the name of the towns (*municipidescripcio*), the date of the metrics (*data*), as well as the number of cases (*numcasos*) and the resultant incidence rate (*incidence*), previously calculated. In Figure 24, it is visible this information for the towns' table.

	codiine text	nom_muni text	municipi text	geometry geometry
1	08175	Puig-reig	081751	0106000020E6100...
2	25044	Bassella	250447	0106000020E6100...
3	17116	Osor	171169	0106000020E6100...
4	08100	Gurb	081000	0106000020E6100...
5	25037	les Avellanes i ...	250371	0106000020E6100...
6	08151	Oristà	081516	0106000020E6100...
7	43149	Tivenys	431495	0106000020E6100...
8	17105	Mieres	171058	0106000020E6100...
9	17163	Sant Gregori	171637	0106000020E6100...
10	17015	Banyoles	170157	0106000020E6100...




Figure 23 Table containing the geometry of the towns in PostGIS.

municipicodi text	comarcacodi text	municipidescripcio text	data text	numcasos bigint	incidence double precision
25053	33	BENAVENT DE SEGRÀ	2021-0...	1	65.88
17015	28	BANYOLES	2021-0...	3	14.96
17184	19	SANTA PAU	2021-0...	1	61.96
17147	31	RIPOLL	2021-0...	1	9.26
08219	21	VILASSAR DE MAR	2021-0...	3	14.26
08086	41	LES FRANQUESES DEL V...	2021-0...	6	29.46
17155	20	SALT	2021-0...	2	6.22
08010	07	ARTÉS	2021-0...	3	52.07
08298	24	VIC	2021-0...	2	4.2
08041	41	CANOVELLES	2021-0...	3	17.83

Figure 24 Table containing the COVID-19 metrics of the towns in PostGIS.

Tableau Desktop, used to design the dashboard's interface, directly retrieves the data from the database, as well as joins the tables containing the geometry with the COVID-19 metrics in both administrative levels, as well as within them. That is why the INE code for the towns and the counties is necessary to perform joins.

The middleware that allows the communication between the server and the user interface does not need external configurations. Tableau Desktop commands allow the construction of dashboard interfaces with different functionalities, but not an automatic refresh of the data. The background of the resultant dashboards is elaborated with Canva. Then it is Tableau Public that allows sharing the link from a browser. All these steps are summarised in Figure 25.

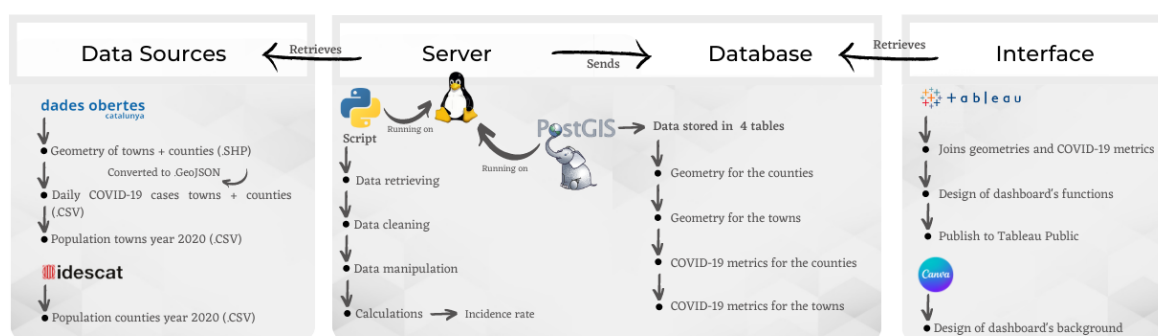


Figure 25 Summary of the steps for the dashboard development.

5.2 Results

The resultant dashboards are designed according to the results of experiment I, which relate to:

- Good assets: light aesthetics with a choropleth map, a country list sidebar with a search option, numeric metrics that interact with the desired requested information, noticeable and user-friendly graphs and tabs, an option to choose a date, and a clear title.
- Assets to avoid: darker colours and graduated symbols or dots for the cartographic symbology, static numeric metrics, clickable elements to display pop-ups, hard to use/find functionalities in the case of the graphs and tabs, as well as big blocks of explanatory text.

The dashboards are named *Light Version (LV)*⁸ and *Dark Version (DV)*⁹ and they are compounded by the detailed elements. In the first place, they share similarities: they both have a title, credits, a map field, a list of countries, numeric metrics, a graph and tabs to change between administrative regions (counties and towns). To communicate the information in a truthful manner, the map fields display a choropleth map showing the incidence rate (cases per 100 000 inhabitants) together with graduated symbols displaying the absolute number of cases, following the cartographic rules.

⁸ Dashboard available in the following link:
https://public.tableau.com/app/profile/anna7816/viz/LV_16463986936330/DashL1

⁹ Dashboard available in the following link:
https://public.tableau.com/app/profile/anna7816/viz/DV_16463985997160/DashD1

They are visible in Figure 26 and Figure 27, and the elements that differ from one interface to the other are the following:

- The LV has light aesthetics and DV's are dark.
- The title is bigger in LV and the credits in DV.
- The DV has a block explanatory text, and the LV's list of counties/towns has a search option.
- The numeric metrics are interactive and change to region/time parameters in LV, whereas they show general static values for last update in DV. To obtain this information, the user must click to the symbols and the values appear in a pop-up shape.
- To display temporal time, connected to the graph and the map (as well as numeric metrics in LV), LV has a drop-down menu to choose a specific date, whereas DV has a time-slider.
- The tabs to change between administrative regions are bigger on the top right of the screen for LV, and smaller on the bottom right (under the list) for the DV.

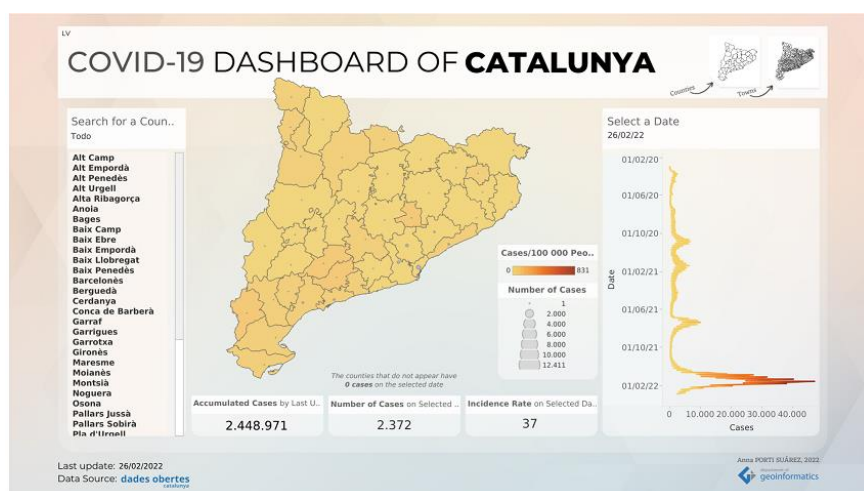


Figure 26 Light version of the COVID-19 Dashboard of Catalunya.

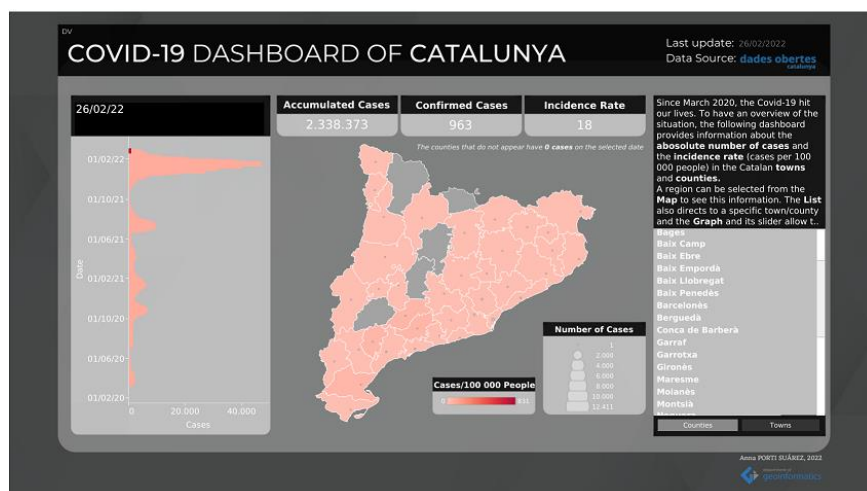


Figure 27 Dark version of the COVID-19 Dashboard of Catalunya.

In general, the elements in the LV are intended to be interactive with each other and show temporal and regional data. DV shows the same, but the behaviour of the elements is static, and the elements need to be clicked to display the information, intending to make it a slighter less user-friendly interface, with other assets considered negative according to experiment I (e.g., dark aesthetics).

5.3 Conclusion

The target of developing two dashboards for experiment II that require displaying COVID-19 cases is achieved. For this, an SDI needs to be built to communicate this information. While developing the dashboard and the required infrastructure there are some challenges encountered.

In the first place, the structure of the data source is complex and needs a lot of manipulation through data cleaning to obtain the desired results to visualise in the dashboards, which is very time-consuming.

The tool selected to develop the user interface is Tableau. As a desktop software, it is an intuitive tool that allows creating interactive dashboard visualisations of different metrics with various widgets and functionalities. On the other hand, it only allows sharing the results in Tableau Public and does not permit the automatic refresh of the data sources, which is a roadblock to displaying up-to-date information that requires constant monitoring.

Despite the challenges, the two dashboards are elaborated. They have different interface appearances according to the insights obtained in the first experiment. They share similarities, since they both need to have the essential elements (e.g., a title, a map field, credits, and other widgets showing temporal data and regions). But the design of the functionalities changes regarding which good and bad assets are identified previously, purposely placed in both of them, in order to be further tested.

In general, the elements in the LV are intended to be interactive with each other and DV's elements' behaviour is static, together with other differences, intending to make it a slighter less user-friendly interface and test the insights from experiment I.

6 EXPERIMENT II

6.1 Methods

The second experiment is a formative study to identify the problematic aspects of the two self-developed dashboards. Details about their development are described in chapter 5. Moreover, this study compares the dashboards functionalities designed on purpose to evaluate if the aspects presented are user friendly or not, following the insights obtained in experiment I.

Experiment Design

The experiment starts with its design meeting the requirements to reach the proposed goals. It is done with SMI Experiment Center 3.7 and it consists of the following steps:

- Calibration: to check that the deviation of the visual angle is under 1° .
- Introduction: including information about the experiment to the participant.
- Free exploration: the participants can explore for one minute each of the two self-developed dashboards that appear in a randomised order.
- Task solving: the participants have to solve ten tasks, five for each dashboard, first for the light version and then for the dark version.
- Interview: the participants express their opinion about the functionalities of each dashboard, as well as a comparison.
- Questionnaire: to obtain information about the aesthetics and difficulty of the dashboard's use, together with the participant's level of expertise in cartography.
- Acknowledgement: thanking the participant and finalising the experiment.

The eye movement, the participant's voice and image are recorded, with a camera and a microphone, to collect data during the interview. The detailed procedure is visible in Figure 28.

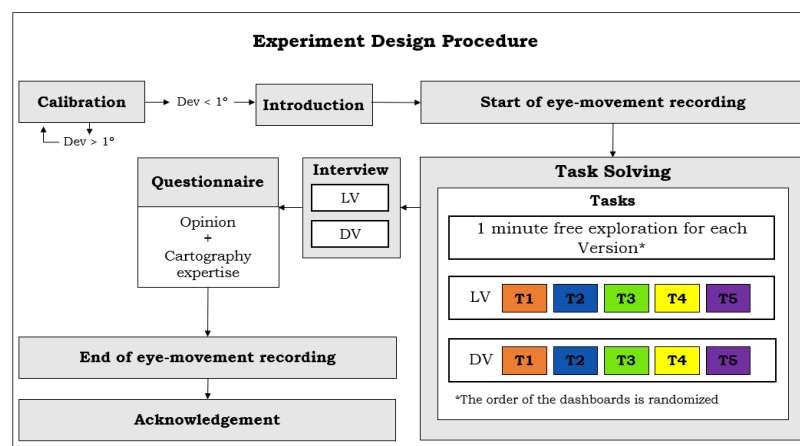


Figure 28 Experiment II design procedure.

Stimuli and Tasks

The two self-developed dashboards detailed in chapter 5 are the stimuli used for this experiment. Five tasks (T1, T2, T3, T4 and T5) are assigned to each of them, so in total there are ten tasks. These involve finding an answer to a specific question by interacting with the dashboard interfaces.

From tasks 1 to 5, the level of difficulty increases: T1 is the simplest and refers to finding a numerical metric displayed on the screen without having to change any parameter, such as the accumulated cases or the number of confirmed cases by the last update; T2 and T3 require to change a parameter, either the region or the date, or both, to find numerical metrics; T4 and T5 require to change the administrative level of the regions by selecting a tab, as well as compare numeric metrics from different dates and regions. The tasks for each of them are further detailed in the following lines.

The tasks for the first dashboard (visible in Figure 26, page 39), from now LV (light version), are the following: accumulated cases in Catalonia by last update (T1); incidence rate in the county of Anoia by last update (T2); the number of cases in the county of Maresme on the 01/01/22 (T3); if the number of cases is higher in the town of Girona, on the 01/10/20 or on the 01/10/21 (T4); and if the incidence rate is higher in the town of Girona or Tarragona, on the 02/02/22 (T5).

The second dashboard's tasks (visible in Figure 27, page 39), from now on DV (dark version), consist in: number of confirmed cases in Catalonia by last update (T1); incidence rate in the county of Bages on the 12/12/21 (T2); incidence rate on the town of Barcelona by last update (T3); if the incidence rate is higher in the town of Barcelona, by the last update, or the same date one month ago (26/01/2022) (T4); and if the number of cases is higher in the town of Lleida or Terrassa, on the 01/01/21 (T5).

Apparatus

To record eye movement, the eye-tracker SMI RED 250 is used, which records at a frequency of 250 Hz, in the eye-tracking laboratory of the Department of Geoinformatics of the Palacký University Olomouc, Czechia. As stated before, a camera and a microphone (Logitech C920) are also employed. SMI Experiment Center 3.7 is the software used to design and record the experiment, combined with iView X. The stimuli are displayed in a Google Chrome browser on the participants' 24-inch screen.

Participants

Between the 9th and 18th of March (2022), experiment II is performed. Twenty university students or recent graduates are the participants. In this case, their level of expertise in the field of cartography is evaluated. As visible in Figure 29, the results are very equilibrated. 25% of the participants do not have any expertise in the cartographic field; 20% a little, medium or good; and 15% are experts. These results allow the analysis of the behaviour of users with better or worse understanding of the presentation of spatial data, without their knowledge having to influence the outcome.

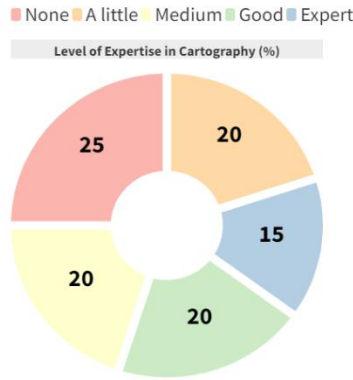


Figure 29 Experiment II participants' level of expertise in cartography.

To identify the ideal number of participants for this study, the Sample Size Calculator for Discovering Problems in a User Interface for Measuring U is employed, as in the first experiment. For the second one, it is also estimated how many respondents would be appropriate to detect at least 85% of the problems encountered. Considering eleven problems are encountered, the result is twelve. Twenty participants are selected, which is higher than the recommended. This information's summary is in Figure 30.

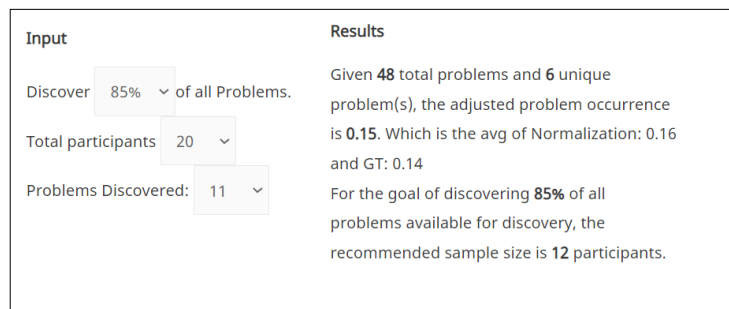


Figure 30 Participants number estimation (Sample Size Calculator for Discovering Problems in a User Interface for Measuring U, 2022).

As displayed, some of these problems are unique (six), and others are more generalised. Almost all participants commit the same two problems. The first one, is that when they are asked to answer a question regarding a specific town, they are not able to distinguish what a county is from what a town is, due to the lack of knowledge regarding the study area. Also, in most of the cases, during the procedure, they finally realise. It is a similar case with the fact that, almost all the participants, when interacting with the DV, do not realise the numeric metrics are static and do not change the variables when selecting a specific region or date. In that case, most of them imply the values shown are correct, when indeed they must click for the pop-up to display the specific numbers. Also, in many cases, they realise this as they progressively reply to the questions.

Other problems are happening on a few occasions, and they are referring to actions like not selecting the proper date or the proper region, misunderstanding values, and mixing up numbers, dates, and names.

Data Pre-processing

After the performance of the experiment, the data obtained needs to be processed. The metrics evaluated are not dependent on the tracking ratio and still provide valuable information to obtain insights for the study, both quantitative and qualitative, so there is no need to exclude any data.

The first step is to identify the fixations and the saccades. For that, the I-DT algorithm is applied, used for data measured at 250 Hz and below (Holmqvist et al., 2011), by applying two thresholds: one for the minimum fixation duration and one for the maximum dispersion. The optimal fixation detection settings for cartographic research in SMI BeGaze 3.7 are 80 milliseconds for minimum fixation duration, and 50 pixels for maximum dispersion, according to Popelka (2014).

The recorded eye movement and screen monitoring need to be processed in a way that allows the data to be separated into different individual tasks according to the dynamics of the experiment. As with the previous experiment, with SMI BeGaze 3.7 the videos can be segmented into custom trials. This is used to analyse eye-tracking metrics in the general context of the tasks, but cannot be used to analyse each participant's behaviour, for which areas of interest are established, manually, using the AOI editor of the same software.

Methods of Analysis

The same methods as in the previous experiment are used to analyse the processed data. As different, this experiment contains one minute of free exploration. To analyse this minute for each dashboard version, the entry time or time to the first fixation to each of the AOIs is observed. This indicates the average duration that it takes a respondent, or all respondents on average, to the first fixation into an AOI (SMI, 2010).

The correctness rate of the task solving evaluates the effectiveness, if the participant can reach the objective of the task or not. The trial duration, or the time needed to solve the tasks, visualised in boxplots created in R studio, evaluates the efficiency, or how fast they solve the tasks (Rubin & Chisnell, 2008). The usage of a correlation test provides more than 0.8 correlation values between the trial duration and the scanpath length of each task, therefore, this metric is avoided for analysis since it leads to obtaining the same insights.

The AOIs are also visualised with sequence charts for each of the ten tasks. These are created in SMI BeGaze 3.7 and show the fixation time of each participant around different elements of the dashboard. Each dashboard has the same elements: a map field, a title, numeric metrics, a list of counties/towns, a date selector (drop-down for LV and slider for DV), a graph, tabs to change the administrative level, and credits including the last update and the data source. These elements are organised in different layouts, visible in Figure 31 and Figure 32.

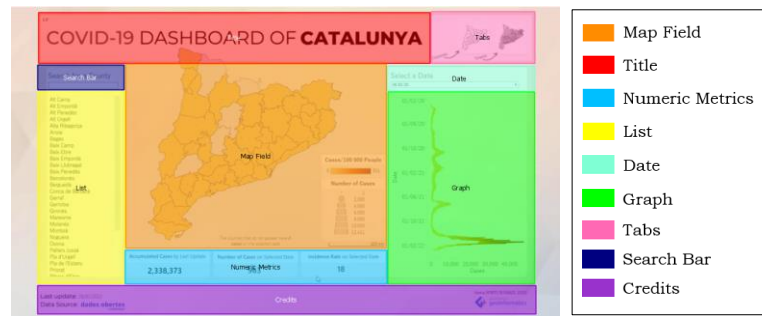


Figure 31 Areas of interest of LV.

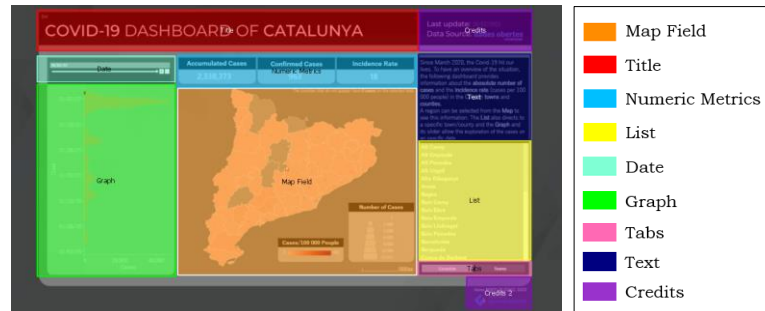


Figure 32 Areas of interest of DV.

The search bar on the top of LV and the DV's block of text are the only two elements unique in each dashboard. Other differences in the layout are the graph's and the numeric metrics' position, opposite in each case, and the size of the elements.

As explained, the experiment also contains an interview and a questionnaire. The interview leads to obtaining a comparison between the same functionalities displayed differently in each dashboard, as well as the opinion on specific elements of them, which enables to measure the users' perception towards the interface. Finally, the questionnaire includes questions to also analyse the satisfaction, regarding the confidence of the user when solving the tasks to test their usability, as well as if they find them aesthetically nice, together with a question to identify their level of expertise in the field of cartography, which is displayed in the participants' overview (Figure 29, page 43).

With these methods, conclusions regarding this experiment can be elaborated, and the insights of the first experiment can be confirmed or rejected, to reach the final conclusion of this project.

6.2 Results

6.2.1 Entry Time

First, the entry time to each AOI during the free exploration time, visible in Table 9, provides insights that relate to the role of the size and position of the elements on the layout.

Table 9 Entry time average to each AOI during the free exploration time.

AOI	Entry Time Average LV (ms)	Entry Time Average DV (ms)
Map Field	2173	1975
Title	2866	10345
Numeric Metrics	15204	3521
List	7933	15723
Date	15174	14675
Graph	4946	6362
Tabs	30579	18746
Search Bar	10110	
Text Block		13130
Credits	16008	14696

The time to the first fixation in the map field (in milliseconds) is similar for both dashboards, since it is the main element, being the first one to be perceived. The title and the list are spotted very quickly in the LV, but it takes very long for the DV, since LV's are bigger, and the list is right under it. The opposite situation happens to the numeric metrics, which can be related to the fact that they are located on the top of the map field for the DV and at the bottom for the LV, being more visible in the first case.

The graph is one of the first elements perceived and shows similar entry times for both versions. So do the credits, being one of the last. The search bar (LV), the text block (DV), and the date selector are not perceived quickly either, also sharing similar values of entry time in both cases. The tabs are a curious case: the ones from LV are bigger and on the top right, and the DV's are smaller on the bottom right under the list, but the DV's are perceived in almost half of the time as the LV ones.

6.2.2 Correctness

In each dashboard the questions go from easy to hard, the LV appears first, and then the DV. Even with the minute to explore the dashboards freely and get familiar with their functionalities, users expect similar outcomes from LV in DV and bring wrong answers.

The results in Table 10 show big differences between the success in task solving in each dashboard. The first task (and easier) of every dashboard is in most cases correctly answered (85% for both versions). Mistakes in the first question can also be related to the unfamiliarity of the participant with the dynamics of the experiment.

Table 10 Correctness of the task solving.

	LV									
	T1	%	T2	%	T3	%	T4	%	T5	%
Correct	17	85%	18	90%	11	55%	15	75%	16	80%
Incorrect	3	15%	2	10%	9	45%	5	25%	4	20%

	DV									
	T1	%	T2	%	T3	%	T4	%	T5	%
Correct	17	85%	2	10%	1	5%	11	55%	11	55%
Incorrect	3	15%	18	90%	19	95%	9	45%	9	45%
TOTAL	20	100%	20	100%	20	100%	20	100%	20	100%

When it comes to the next questions, the difference between both versions is very significant: T2 and T3 have a 90% and 55% success rate, respectively, in the case of LV, being T3 lower because it involves more difficulty, while T2 and T3 success rate for DV is 10% and 5%. This is related to the fact that users are supposed to look for a specific region incidence rate or number of cases, and they expect the numeric metrics to change (like it happens in the LV), but these are static and to find the answer they need to display a pop-up. It takes the participants a couple of questions to realise that, and that is why the success rate increases in the last two tasks.

Finally, T4 and T5 are the hardest questions because they involve comparing values between two regions or two dates. The LV's success rate is 75% and 80% and the DV's 55% in both cases. Again, LV shows a higher success rate, indicating its interface appears to be more intuitive for its users. The success rate of the last two tasks for DV is higher than the previous two even though they are more difficult because the participants realise the functionalities behaviour and familiarise themselves with the interface during the experiment, such as the numeric metrics being static and the need to click for a pop-up display.

6.2.3 Trial Duration

Considering the difficulty increases during the experiment, the trial duration should as well, but this is not always the case. The median of the time required for T1 of LV is higher than for T2. This is related to the fact that it is the first question, and the participants are adapting to the task solving process, because it is not the same in the DV case.

For the LV, the median tends to increase from T2 to T4, but T5 decreases considerably. It can relate to the familiarity of the participants to solve the tasks, as well as that they already have the region regarding the question selected from the previous task, which has to be compared to another region, and saves some time. Whereas for T4, they have to compare two dates and it takes longer to browse through the drop-down date menu. This can be avoided by typing in the date in the search bar, but it is not always noticeable by the participants.

For the DV, the trial duration from T1 to T5 increases, except for T2, being its median higher. This is related to selecting a specific date in the time-slider, which requires a lot of time to achieve with precision. In this case, the last task, and hardest, is the one taking longer and with a higher median, since two regions need to be compared and without the search option, it takes a lot of time to scroll through the list to find them.

In general, tasks in LV take longer to solve than in DV, but the time difference is very small. T1 is the first question of all the experiments and takes 14 seconds longer in LV than in DV. The difference between T3 is about 4 seconds of difference and for T4 is 11 seconds.

On the other hand, T2 and T5 take longer to solve in DV, with a more significant difference, the time is almost doubled: from 26 to 47 seconds in T2, and from 44 seconds to 1 minute and 36 seconds in T5. That it takes shorter to solve DV's tasks can be related to the fact that most of the participants do not realise the numeric metrics are static. This makes them use the same values over again for the first tasks until they realise it is necessary to click the map to display a pop-up, as the correctness rate displays.

The normal distribution of each value is analysed with the Shapiro-Wilk test. All the tasks' values show that they do not have a normal distribution, except for the DVT5. For this reason, a non-parametric test should be performed, in this case `kwAllPairsNemenyiTest` from `PMCMRplus` package of R. In the LV there is a statistically significant difference between T4 and T1, T2 and T3, being the task that took longer to solve, with the values less distributed and statistically different from the rest. For the DV, T1 is statistically different from T2, T4 and T5, as well as T3 differs from T5. T1 differs from the majority because its distribution is not equal, whereas T5's is. The boxplots in Figure 33 show this information.

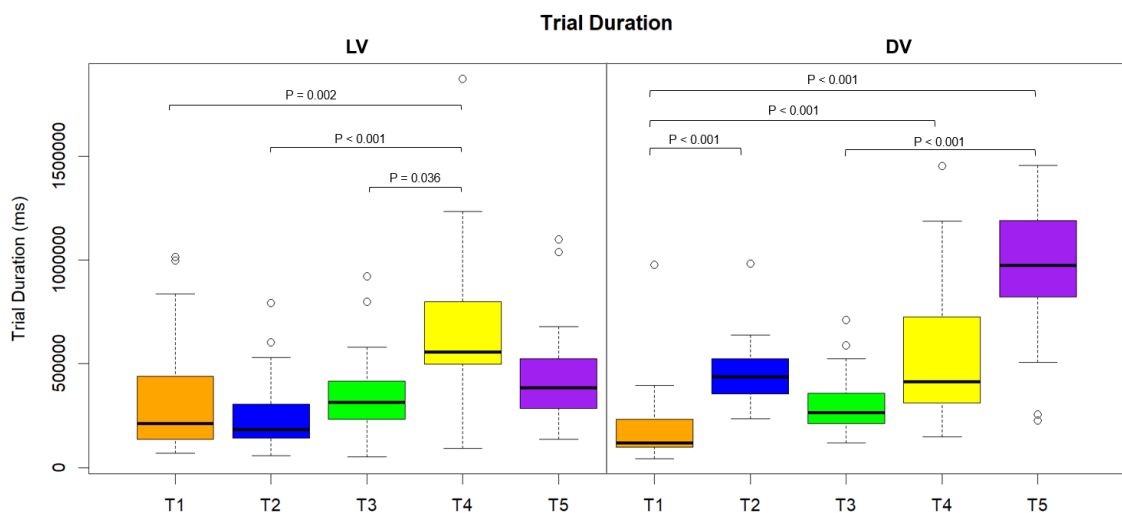


Figure 33 Trial Duration of each task.

To do a comparison between dashboards, the same non-parametric test is used to analyse the statistically significant similarities/differences between the tasks. For T1, it is visible that the LV median and time to solve the task is higher. Again, it is related to it being the first question and the participant adapting to the task solving process, not necessarily to its difficulty. Their statistical significance also differs. It is the same case for the T2, LV and DV differ statistically, but in this case, the median and the trial duration are higher for the DV, due to more difficulty and time required for date selection in the time-slider.

Both for T3 and T4 the median is higher in the case of LV but does not differ in the slightest; and for T5, the two dashboards differ statistically and the difference between the median and the time needed to solve the last task is definitely higher for DV, due to its difficulty and lack of search tools to make the processes shorter and easier. This is visible in Figure 34.

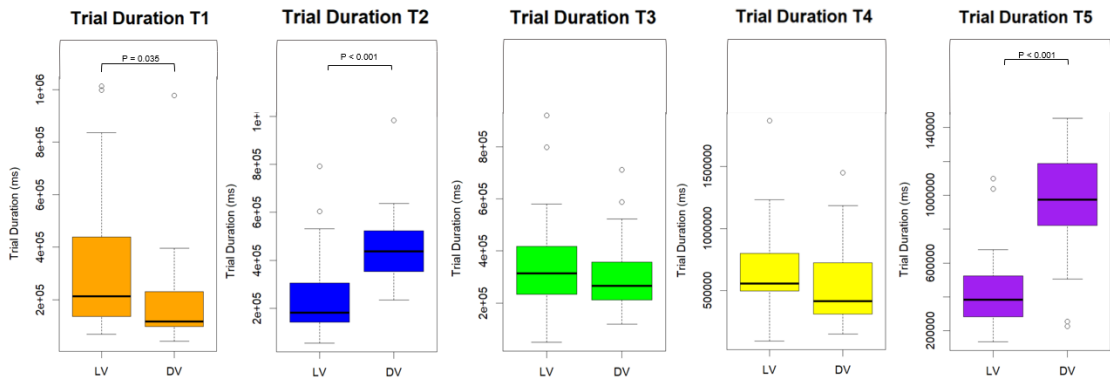


Figure 34 Trial duration of each task per dashboard.

6.2.4 AOIs

The sequence charts show the fixation time of the participants on each of the AOIs and enable to see which elements of the interface they focus more on. Depending on the proposed activity and the elements compounding the dashboard, the participants focus on different elements, even though there are common cases, such as the map field.

For the LV, the title and the credits are barely perceived as in experiment I. Some participants focus especially on them, in the cases where the last update wants to be identified. The map field and the numeric metrics are predominant in all the tasks since these are dynamic regarding the selected region/date. The graph also plays an important role, even though it is not crucial to solve the tasks, people pay a lot of attention to it, also because it is a method of searching the date and includes the drop-down date selector. In most cases, the list takes a lot of fixation time and the search bar option is only perceived by some participants. The tabs to change the administrative level are very important but the participants do not pay attention to them, which leads to a lot of confusion and a low success rate. All this is visible in Figure 35, showing the example of T4, where the participants must compare a numeric metric of a specific region between two periods of time, and the graph and the list have a long fixation time.

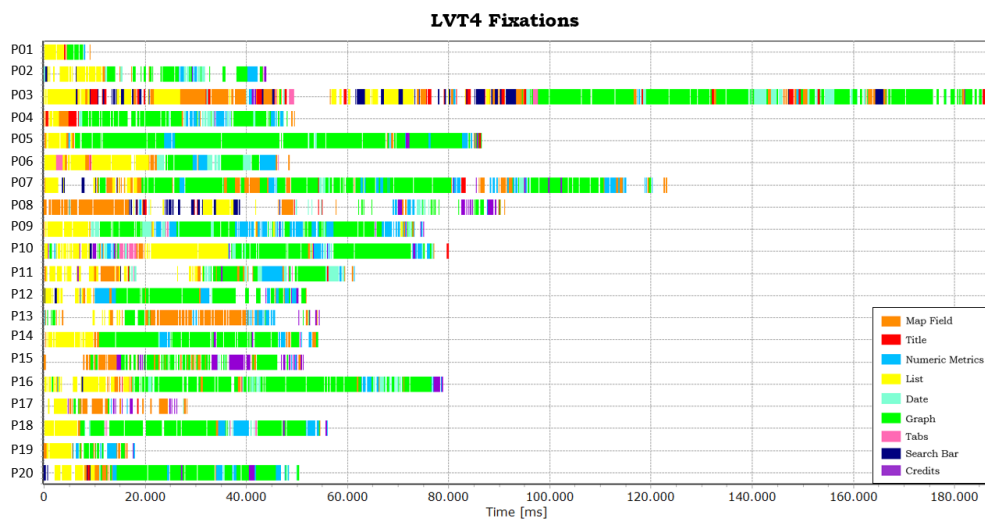


Figure 35 LVT4 Fixation time to AOIs per participant.

For the DV, some things are like the LV. The title and the credits are not an important focus of attention in most of the cases and the map field takes a lot of fixation time since it plays an important role. So does the list, because of its importance, but also because there is no search option and the participants must invest a lot of time to find a specific town/county. In addition, the text block does not have a lot of attention from the participants which indicates they are not interested in the information it might provide. This is visible in Figure 36, which is T3, where participants look for a numeric variable of a specific region by the last update.

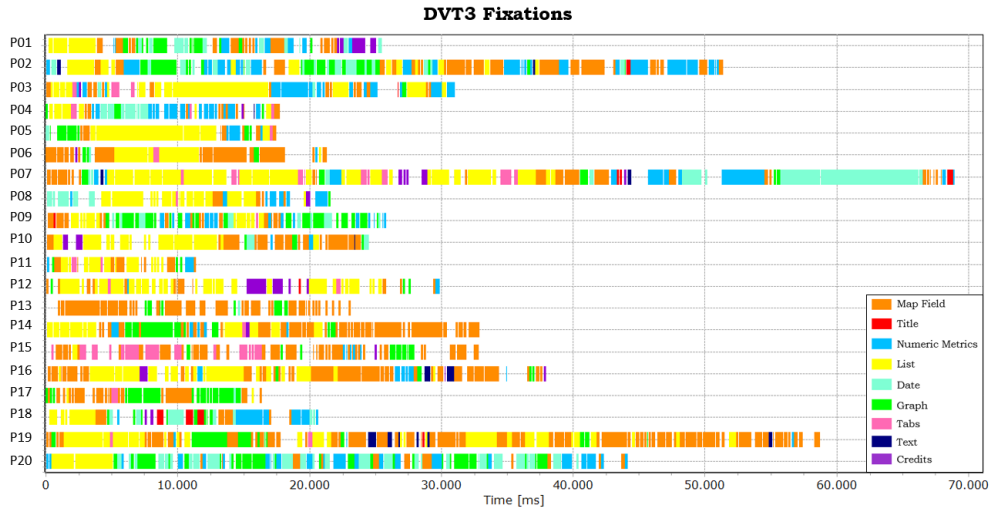


Figure 36 DVT3 Fixation time to AOIs per participant.

On the other hand, the date selector, in this case in slider shape, requires a lot of fixation time, because it is hard to select a specific date and demands a lot of time from the participant. This is visible in Figure 37, which refers to T4, in which the participants must find a numeric variable of a specific region, on a selected date.

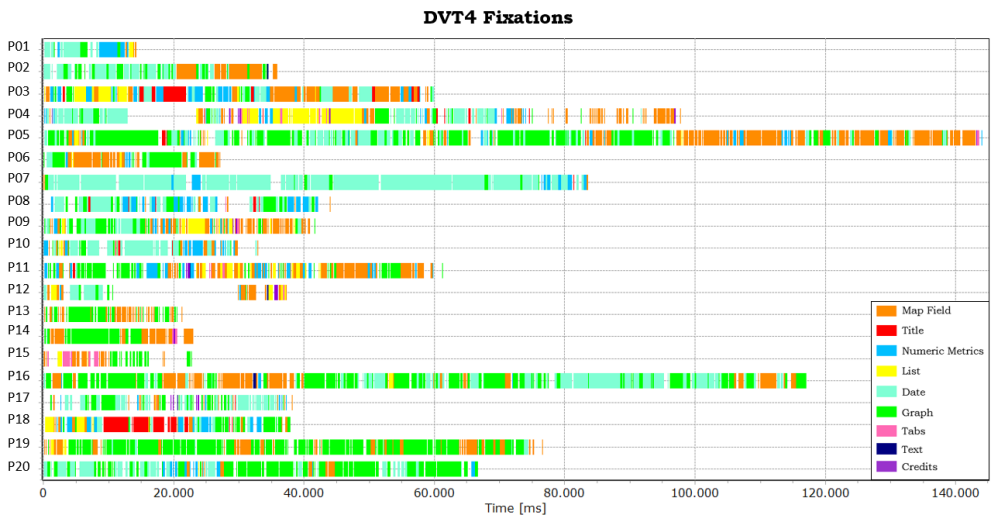


Figure 37 DVT4 Fixation time to AOIs per participant.

Regarding the graphs and the numeric metrics, they get a lot of fixations, but slightly lower than in the LV. Also in the case of the numeric metrics, the attention time decreases during the participants realising they are static and do not change when interacting with other elements of the dashboard, as visible from Figure 36 to Figure 37.

The tabs, which participants find and use to change the administrative level in a higher number of cases compared to LV, are also visible on a few occasions. The rest of the sequence charts from the other tasks can be found in Attachment 1.

6.2.5 Questionnaire and Interview

The questionnaire and the interview provide information regarding the subjective opinion of the participants. First, they are asked to rate aesthetically the dashboards, from 1 (very ugly) to 5 (very nice). The participants also rate how difficult it is to solve the tasks of each dashboard from 1 (very difficult) to 5 (very easy). The results are in Table 11 and Table 12.

Table 11 Aesthetics rate.

	LV	%	DV	%
Very Ugly	0	0%	0	0%
Ugly	0	0%	4	21%
Ok	1	5%	5	25%
Nice	6	30%	9	45%
Very Nice	13	65%	2	10%
TOTAL	20	100%	20	100%

Table 12 Difficulty rate.

	LV	%	DV	%
Very Difficult	0	0%	2	11%
Difficult	0	0%	4	21%
Ok	3	15%	8	40%
Easy	10	50%	3	15%
Very Easy	7	35%	3	15%
TOTAL	20	100%	20	100%

The LV is in 65% of cases considered very nice, 30% nice and 5% Ok. Any participant considers it ugly or very ugly. These rates are higher than the DV's, which only 10% think is very nice. Still, 45% think it is nice, but 25% decided on ok and 21%, which is a fifth of the participants, for ugly. Regarding the difficulty, the results show a similar tendency. The LV is generally considered easier: half of the participants say it is easy, 35% very easy, and 15% Ok. Again, none considers it difficult or very difficult. Whereas DV has a wider range of opinions. The major number is the participants that consider it is Ok, with 40%. The number of participants that consider it difficult or very difficult (21% and 11%, respectively) is higher than the number that considers it easy or very easy (15% in both cases).

The interview also allows obtaining more subjective opinions regarding specific elements of the dashboards. The participants are asked to freely speak their mind regarding the user-friendly/non-user-friendly functionalities of the dashboard and experience during the task solving process. With these opinions, several insights are obtained: a comparison table between the same functionality presented in a different manner in both dashboards (Table 13), specific functionalities of each dashboard mentioned by the participants as good/bad assets (Table 14) and extra interesting recommendations of the creativity of the participants.

Starting with the comparison, some of the participants express their opinion by comparing the same functionality displayed differently in each dashboard. Overall, most functionalities of the LV are widely preferred over the DV ones. The favourite element was the drop-down date option, in opposition to the time-slider: 85% of the participants say the LV's option was a better option. The fact that the numeric metrics change when interacting with the dashboard and the light aesthetics are also the choice of more than half of the participants (60% and 55% respectively). On the other hand, 15% are in favour of the dark aesthetics and 10% support the clicking option to display the values.

When it comes to the list, 37% mention it is a nicer element and more user-friendly in the LV since it is complemented with a search option that the DV does not have. Finally, the DV's tabs accessibility is considered better with the opinion of 25% of the participants.

Table 13 Comparison between LV and DV functionalities.

	LV		DV	
	Participants	%	Participants	%
Date	17	85%	1	5%
Numeric Metrics	12	60%	2	10%
Aesthetics	11	55%	3	15%
List	7	35%	0	0
Tabs Accessibility	3	15%	5	25%

Besides their comparison, specific elements are highlighted as positive or negative. 40% of the participants call the LV user-friendly, while only 5% the DV. The map field and the graphs are commented in a very anecdotic percentage (5-15%), the first mainly as a positive asset, and the second as a negative. Last, the text in the DV, which is not present in the LV, is considered unnecessary in 20% of the cases. Participants express the fact that they would not pay attention to the text.

Table 14 Positive and negative assets of LV and DV.

		LV			
	Participants	%	Participants	%	
User-friendly	8	40%	User-friendly	1	5%
Map Field	1	5%	Graphs	3	15%
		DV			
	Participants	%	Participants	%	
Text	1	5%	Text	4	20%
Map Field	1	5%	Map Field	1	5%
			Graphs	1	5%

Finally, the participant's opinion makes it possible to obtain interesting recommendations. The ones highlighted from this experiment, which did not appear in the last experiment, relate to the necessity of a clear button to clear the selected element, instead of having to unselect it; and the graph to be measured as a logarithmic scale, since there would not be such a big difference between higher and lower values.

6.3 Conclusion

The conclusion summarises the results to identify which are positive and negative elements of the dashboards, comparing which elements of LV and DV are preferred.

The entry time shows that the size and position of the elements can play a role in the user's attention. For example, the map field is the dominant element of the dashboard, so it is the first element to obtain the attention, while the credits are the last. Big elements such as the graphs or the list are perceived before others, like search bars or date selectors, because of their size. On the other hand, some elements, like the numeric metrics are noticed earlier, not only if they are dynamic rather than static, but also depending if they are on the top or on the bottom of the layout. The tabs are more likely to be perceived if they are near the element they refer to (in this case, the list of regions), rather than because of their size or having a predominant position in the layout.

The correctness rate shows significantly better results for the LV, implying it is easier for the participants to answer correctly. Except for the T1, the correctness rate of all the tasks for DV is lower, especially T2 and T3, which are wrong by almost all the participants. This relates to the fact that they do not realise the numeric metrics are static and do not change and answer a wrong value.

The trial duration results are related to the correctness: the time should gradually increase from T1 to T5 in each case, but it is not like that. The first task takes longer in LV than in DV because of the adaptation procedure of the participant to the dynamics of the experiment. This does not affect the correctness rate, though. In both dashboards, when having to select or compare dates, it takes long for the participants to select it, but the DV's time-slider takes even longer than the drop-down-menu (LV). Also, the LV allows typing the date to search for it, but not many participants realise that and save the time to scroll. The fact that the DV does not have a search option either for the list of towns/counties increases the time of solving when having to compare two regions as well.

The AOIs show the importance of the map field and the non-importance of the title and credits. In both cases, the list is also taking a lot of fixation time, even more in the DV, because the lack of a search option involves more scrolling time. Regarding the numeric metrics, the attention decreases through the course of the experiment from early tasks to later tasks in DV, when the participants realise it is a static element that does not interact. As expected, the text in the DV does not take a lot of attention and, unexpectedly, the tabs in DV take more attention than in LV, which leads to people changing the administrative level in more cases, bringing a higher correctness rate. The graphs, even though they are not essential to solve the tasks, take a lot of fixations, especially in the case of LV, because they can be used to select a date and overlaps with the drop-down menu. On the other hand, the date selector (time-slider) of DV takes a lot of attention because it is difficult to select a date with precision and it involves a lot of time.

When asking for the subjective opinion of the participants regarding the aesthetics and the difficulty the dashboards, for both metrics, to a significant major extent, LV is considered nicer and easier, showing it is more user-friendly, intuitive and aesthetically nicer. When the participants freely give their opinion, comparing the two dashboards, the functionalities of the LV are widely preferred over the DV ones. The most mentioned, from high to low, are the date selector element, the dynamic numeric metrics, the light aesthetics, and the search option. Whereas the only element preferred of the DV is the tabs' accessibility, proved by the fact that more participants made use of them than in LV. Besides this comparison, they also express their opinion on specific elements, and the main highlight is that almost half of the participants call the LV user-friendly and the text of DV is considered unnecessary in 20% of the cases.

According to the new outcomes that both quantitative and qualitative results provide, dashboards should contain a map field following the cartographic rules, light aesthetics, a sidebar list with a search option, numeric metrics that interact with the desired requested information (on the top of the layout), tabs near to the element they refer to, graphs with a date selector as a drop-down menu or typing box, and a clear title. Time-slider as a date selector and blocks of explanatory text are roadblocks.

7 RESULTS

The following subchapters summarise the results achieved in the research.

7.1 Experiment I

The first experiment is a formative study testing user evaluation on existing dashboards. According to what both quantitative and qualitative data results show, the following recommendations are formulated. User's preferences indicate dashboards should include light, simple aesthetics with a choropleth map, a country list sidebar with a search option, and numeric metrics that interact with the desired requested information (instead of pop-ups). In case of graphs, date selectors and tabs, they should be user friendly and noticeable, and a concise, clear title is necessary. Less appreciated elements are darker colours and graduated symbols or dots, static numeric metrics, hard to use/find functionalities (e.g., having to hover over a graph to select a date) and big blocks of explanatory text. With these insights, the stimuli for the second experiment are elaborated.

7.2 Dashboard Development

For the development of the dashboards, used as the stimuli for experiment II, different steps need to be followed to proceed and build an SDI employing several tools to communicate this information. The resultant ones, LV and DV, are designed according to the insights of experiment I and relate to user-friendly and non-user-friendly elements' appearance and position in the interface. Both are visible in Figure 26 and Figure 27 (page 39).

The dashboards have a different interface appearance, but they share some similar elements: both map fields are a choropleth map showing the incidence rate (cases per 100 000 inhabitants) with graduated symbols displaying the absolute number of cases following the cartographic rules. They are also compounded by a title, credits, a list of countries, numeric metrics, a graph and tabs to change between administrative regions (counties and towns), which appearance can change. LV has search options and numeric metrics that are interactive, while DV has not, but it has a block with explanatory text and the date selector is a time-slider. Also, the LV's aesthetics, as the name indicates, are light, and the DV's are dark.

In general, the elements in the LV are intended to be interactive with each other and show temporal and regional data. DV shows the same, but the behaviour of the elements is static, and the elements need to be clicked to display the information, being less user-friendly and containing other assets considered negative according to experiment I.

The technologies' familiarisation is part of this research. The usage of Tableau Desktop allows drag-and-drop, user-friendly functionalities to design interactive dashboard interfaces, but on the other hand, it does not allow automatic refresh of the data, which fails to serve the purpose of a dashboard that tries to provide up-to date information on the metrics of a specific topic, in this case COVID-19.

7.3 Experiment II

The user testing of self-developed dashboards confirms some of the insights obtained in the first experiment, while also bringing new outcomes. The objective methods show that the version with lighter aesthetics is more user-friendly, and the subjective option of the participants confirms it. The functionalities added in the LV allow the time to solve the tasks in average to be lower, as well as the success rate is higher.

Regarding the map fields, without considering the insights of experiment I, these are designed following the cartographic rules, which are using relative values for the choropleths and absolute values represented with graduated symbols, to communicate the information in a truthful way. The map field is an important element of the interface that takes most of the attention time. This is the same case for the list of countries and the graph, together with the numeric metrics, which must be interactive with the requested information as stated in experiment I. The new insight obtained refers to the fact that they are paid more attention when located in the top of the interface, instead of the bottom.

The search option and the date selector are two elements that regardless of their size or position in the interface are always appreciated, used and take fixations. Again, the search option is considered essential by the participants and when choosing a date selector, they prefer to drop-down in a menu, or type it, rather than hovering through a graph or sliding through a time-slider. Finally, the tabs are likely to be perceived if they are close to the element they refer to, in this case, the list of regions, rather than having a more noticeable size and position.

Figure 38 is a final version¹⁰ of the dashboard combining the positive elements of the two dashboards employed for the stimuli of the second experiment, compounding an interface with all the recommended elements in the according layout.

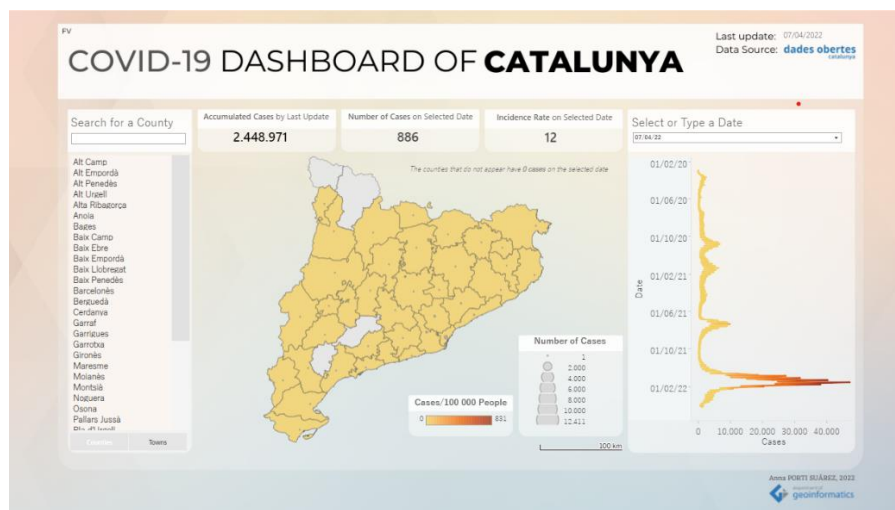


Figure 38 Final version of the COVID-19 Dashboard of Catalunya.

¹⁰ Dashboard available in the following link:
https://public.tableau.com/app/profile/anna7816/viz/FV_16504498862090/DashF1

8 DISCUSSION

During the process of performing user evaluation, together with the development of two dashboards as stimuli for the user testing, several issues encountered need to be pointed out.

When performing experiments with eye-tracking technologies, the tracking ratio of the recorded eye movement needs to be considered, which is defined as “the proportion of time that the eye-tracker recorded point of gaze coordinates over the entire task” (Amso et al., 2014), to further examine the quality of data. For example, Riege et al. (2020) establish a tracking ratio <70% to exclude it. In the case of this thesis, no data are excluded regardless of the tracking ratio, since this quantitative information is combined with qualitative data to formulate recommendations for dashboard design.

A problem encountered during the analysis of the recorded eye-tracking data of the experiment I is the fact that some participants, on few occasions, zoom in and zoom out the web interface, causing a disturbance in the AOIs’ original shape. This is corrected by creating two types of AOIs for the same area, with different sizes, and merging the results. A solution to correct more specific cases would be the usage of dynamic areas of interest, which the software employed (SMI BeGaze 3.7) allows.

Also, during the performance of experiment II, on some occasions, the fact that participants had to interact and solve tasks with an interface showing a study area that they are not familiar with, brought confusion (e.g., distinguishing between a town and a county). This is mitigated by providing little help but could be improved by performing the experiment with a known region or with worldwide data.

Most of the problems are faced during the dashboard creation and its respective SDI. In the first place, the retrieved data from the Catalogue of Open Data of the Government of Catalonia needs a lot of manipulation to structure the data in the required columns to display the information, as well as in an ordered manner to perform the necessary joins between the COVID-19 metrics and the geometries of the administrative levels. This results in a very time-consuming part. A solution to that would also be changing the study area, to mitigate the problem of the unfamiliarity during the performance of experiment II too, since this is the official data source from the government of the study area and there are no alternatives.

The second issue faced is related to the software used to develop the dashboard interface, Tableau. The desktop version of this programme is a very intuitive and easy to use tool, but it has some limitations: the data needs to be refreshed manually, and therefore it is not possible to achieve a dashboard interface that periodically provides up-to-date information regarding the COVID-19 situation in a specific study area. An alternative is Tableau Server, which is a part of the Tableau software that allows automatic refresh of the data extracts by uploading the data source and the workbooks to the Tableau Online platform and scheduling the refresh (Tableau, 2022). To avoid commercial licences, such as Tableau, open-source options have to be further explored. The downside is that they require proficient programming skills.

The automatization of the dashboard is a necessary asset to continuously monitor the COVID-19 numbers and be aware of the development of the situation. The Copernicus Services have different manners to track this situation with EU Space Data which monitors land uses and environmental data, such as NO₂ reduction (CAMS, 2020) from their satellite imagery. Unfortunately, these measures are not related to the absolute or relative values of the COVID-19 variables, such as the number of cases and incidence rate, required for this work, which cannot be obtained to be displayed of any study area.

On the other hand, dashboards displaying other metrics can be developed, such as The Rapid Action on Coronavirus and EO (ESA & European Commission, 2022), a dashboard interface showing metrics regarding earth observation data in European countries, such as air quality. This is combined with the number of cases and vaccinations, retrieved from external sources. The results obtained in this study could be used as insights that would be particularly useful to the Copernicus Services ground segment in understanding the user's way of thinking, to assess further improvement of interfaces of this type.

The qualitative results of this study provide suggestions from the participants to improve dashboard interfaces, and there are two important ones to highlight for better communication of the displayed information: 1) add a button to clear the selected information, instead of having to unselect it manually. 2) use a logarithmic scale in the graphs, to spread the values around the graphic and provide better visualisation in cases like this, where the absolute values are too low and homogeneous, and are not visible.

A clear conclusion in both experiments shows that, when interacting with dashboards, people do not spend time reading the explanatory texts that might introduce them to the functionalities of the interface. The question is if users react like this in web interfaces or just in this case.

The short fixation or attention time employed in the text area might not mean users do not read the explanatory text, it can be that they are scanning for useful information (Manhartsberger, et al., 2005). By tracking a gaze plot, Manhartsberger et al. can identify which text the user reads to fulfil certain tasks. In the same study, they conclude that users face problems in solving certain tasks when the reading text gives instructions of functionality that is not directly placed next to it, bringing usability problems, and recommend displaying the information related next to it, such as the mentioned issue with the tabs in this study, mentioning the Law of Proximity (Gestalt Theory) applied to interactive web interfaces (Graham et. al, 2008).

Together with other results of the user evaluation, these insights provide recommendations for better interface design and suggest improvements. Still, a lot more variables could be further investigated related to dashboard's functionalities, their appearance, size, and placement in a layout, as well as the users' behaviour towards it. This work focuses on dashboards covering COVID-19 metrics, but it could apply to assorted topics as well, to observe if the same patterns are identified within different interfaces.

9 CONCLUSION

The aim of the thesis is to perform user evaluation of dashboards containing geospatial information regarding the COVID-19 topic, establishing three partial goals and using mixed-research methods.

The **first goal** has the aim of obtaining insights about the user interaction with existing dashboards' functionalities. The results identify the problematic elements of these dashboards and formulate recommendations for the elaboration of a more user-friendly dashboard interface that communicates the information appropriately.

The elements recommended according to the insights are the following: the map field's symbology as a choropleth, a sidebar with a list of the regions displayed and a search option, numeric metrics that interact with the desired requested information (instead of pop-ups), noticeable tabs to change parameters, small credits and a big and clear title. Temporal data should be presented in graphs, including a date selector to choose a period of time to visualise. The light aesthetics are preferred to the dark, and big blocks of explanatory text, together with excess of information, are not appealing to the user.

The **second goal** consists of elaborating two self-developed dashboards with the design and placement of the elements according to the insights obtained previously, to be evaluated in the **third goal**. Nevertheless, the spatial information is displayed following the cartographic rules showing a choropleth map with relative values, and graduated symbols for the absolute values, to communicate the information in a truthful manner. The results of the second experiment show most of the insights are repeated, but also provide new outcomes, such as: the numeric metrics must not only be interactive but placed on the top of the interface; the preferred date selector option is through a drop-down menu or typing; and the usage of the tabs does not depend on their size and visibility, but the fact that they are close to the element they refer to, in this case, administrative levels.

A proposed objective, part of the second goal, is the creation of an SDI with the following structure: a back-end server handles the processes with a Python script that retrieves the data, manipulates it, and stores it in a spatial database, from which a visual analytics desktop application (Tableau) retrieves the data to display it in a user interface. The usage of Tableau is an intuitive and easy manner of developing dashboard interfaces to display metrics and spatial information. However, it does not allow the automatic refresh of the retrieved data, which needs to be done manually, and does not serve the purpose of a dashboard interface that provides up-to date information.

To sum up, together with dashboard development tools and functionalities, this process involved the characterisation and familiarisation with analysis methods in cognitive cartography, both quantitative and qualitative, focusing on eye-tracking and all the technologies related. Besides this, the aim is achieved by obtaining insights about user interaction with COVID-19 dashboards, the appearance of their functionalities and their role in communicating and transferring the information properly, in order to formulate recommendations to improve dashboard interface design.

REFERENCES AND INFORMATION SOURCES

- Amso D, Haas S, Markant J (2014). An Eye Tracking Investigation of Developmental Change in Bottom-up Attention Orienting to Faces in Cluttered Natural Scenes. *PLoS ONE* 9(1): e85701. <https://doi.org/10.1371/journal.pone.0085701>
- Bloomberg. (2016, April 4). Understanding Bloomberg and the Terminal. Retrieved October 29, 2021, from <https://www.bloomberg.com/company/stories/understanding-bloomberg-and-the-terminal/>
- City Dashboard (n.d.). London. Retrieved October 29, 2021, from <https://citydashboard.org/london/>
- Copernicus Atmosphere Monitoring Service (2020). Flawed estimates of the effects of lockdown measures on air quality derived from satellite observations. Retrieved April 9, 2022, from <https://atmosphere.copernicus.eu/flawed-estimates-effects-lockdown-measures-air-quality-derived-satellite-observations>
- Dades Obertes Catalunya (n.d.). Comarques de Catalunya i Aran. Retrieved February 9, 2022, from <https://analisi.transparenciacatalunya.cat/Urbanisme-infraestructures/Comarques-de-Catalunya-i-Aran/gsjn-sema>
- Dades Obertes Catalunya (n.d.). Dades diàries de COVID-19 per comarca. Retrieved February 9, 2022, from <https://analisi.transparenciacatalunya.cat/Salut/Dades-di-ries-de-COVID-19-per-comarca/c7sd-zy9j>
- Dades Obertes Catalunya (n.d.). Limits administratius municipals de Catalunya. Retrieved February 9, 2022, from <https://analisi.transparenciacatalunya.cat/Urbanisme-infraestructures/L-mits-administratius-municipals-de-Catalunya/97qg-zvqd>
- Dades Obertes Catalunya (2020). Població de Catalunya per municipi, rang d'edat i sexe. Retrieved February 9, 2022, from <https://analisi.transparenciacatalunya.cat/en/Demografia/Poblaci-de-Catalunya-per-municipi-rang-d-edat-i-se/b4rr-d25b>
- Dades Obertes Catalunya (n.d.). Registre de casos de COVID-19 a Catalunya per municipi i sexe. Retrieved February 9, 2022, from <https://analisi.transparenciacatalunya.cat/Salut/Registre-de-casos-de-COVID-19-a-Catalunya-per-muni/jj6z-iyrp>
- Elliott, D. G. (2003). *Executive Information Systems: Their impact on executive decision making*. UMI.
- ESA & European Commission (n.d.). The Rapid Action on Coronavirus and EO. Retrieved April 9, 2022, from <https://race.esa.int/>
- Few, S. (2006). *Information dashboard design*. O'Reilly.
- Few, S. (2007). *Data Visualization: Past, Present, and Future*. Cognos. City Dashboard (n.d.). Retrieved October 29, 2021, from https://www.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf
- Ghaoui, C., Ball, L. J. & Poole, A. (2006). Eye Tracking in HCI and Usability Research. In *Encyclopedia of human computer interaction*. Idea Group Reference.
- Goldberg, J.H., Stimson, M.J., Lewenstein, M., Scott, N., & Wichansky, A.M. (2002). Eye tracking in web search tasks: Design implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, New Orleans, LA, USA, 25–27 March 2002; pp. 51–58.

- Golebiowska, I., Opach, T., & Rød, J. K. (2020). Breaking the eyes: How do users get started with a coordinated and multiple view geovisualization tool? *The Cartographic Journal*, 57(3), 235–248. <https://doi.org/10.1080/00087041.2019.1660513>
- Google Trends (n.d.). Dashboard. Retrieved November 7, 2021, from <https://trends.google.com/trends/explore?date=all&q=dashboard>
- Graham, L. (2008). Gestalt theory in interactive media design. *Journal of Humanities & Social Sciences*, 2(1).
- Health Map (n.d.). Health Map. Retrieved October 29, 2021, from https://www.healthmap.org/covid-19/?mod=article_inline
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H.n & Van de Weijer, J. (2011). *Eye tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- Humanistic GIS Laboratory of the University of Washington (n.d.). Novel Coronavirus (COVID-19) Infection Map. Retrieved October 29, 2021, from <https://hgis.uw.edu/virus/>
- Institut d'Estadística de Catalunya (n.d.). Population on 1 January. Counties and Aran, areas and provinces. Retrieved February 9, 2022, from <https://www.idescat.cat/pub/?id=aec&n=246&lang=en>
- Jenks, G. (1973). Visual integration in Thematic Mapping: fact or fiction? *International Yearbook of Cartography* 13, 27-35.
- Johns Hopkins Coronavirus Resource Centre (n.d.). COVID-19 Map. Retrieved October 29, 2021, from <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>
- Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, City Benchmarking and real-time dashboards. *Regional Studies, Regional Science*, 2(1), 6–28. <https://doi.org/10.1080/21681376.2014.983149>
- Kitchin, R., Lauriault, T. P., Wilson, M. W. & Matter, S. (2017). *Urban Dashboards*. In *Understanding spatial media*. Sage.
- Krassanakis, V. & Cybulski, P. (2019). A review on eye movement analysis in map reading process: the status of the last decade. *Geodesy and Cartography, Committee on Geodesy of the Polish Academy of Sciences*, 68 (1), 191-209. 10.24425/gac.2019.126088. hal-02146187
- MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design*. The Guilford Press.
- MacEachren, A. M., & Kraak, M.-J. (2001). Research challenges in geovisualization. *Cartography and Geographic Information Science*, 28(1), 3–12. <https://doi.org/10.1559/152304001782173970>
- Manhartsberger, M., & Zellhofer, N. (2005). Eye tracking in usability research: What users really see. In *Usability Symposium* (Vol. 198, No. 2, pp. 141-152).
- Marek, T., Drusch, G., Bastien, C. & Paris, S. (2014). Analysing Eye-tracking Data: From Scanpaths and Heatmaps to the Dynamic Visualization of Areas of Interest. *Advances in Science, Technology, Higher Education and Society in the Conceptual Age*. AHFE Conference.
- Marr, B. (2012). *Key performance indicators (Kpi): The 75 measures every manager needs to know*. Financial Times/ Prentice Hall.

- Measuring U (n.d.). Sample Size Calculator for Discovering Problems in a User Interface. Retrieved December 8, 2021, from https://measuringu.com/calculators/problem_discovery/
- Orlov, P. A., Ermolova, T., Laptev, V., Mitrofanov, A., & Ivanov, V. (2016). The eye-tracking study of the line charts in Dashboards Design. Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. <https://doi.org/10.5220/0005692102030211>
- Popelka, S. (2014). Optimal eye fixation detection settings for cartographic purposes, 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing, Albena, Bulgaria.
- Popelka, S., Herman, L., Řezník, T., Pařilová, M., Jedlička, K., Bouchal, J., Kepka, M., & Charvát, K. (2019). User evaluation of map-based Visual Analytic Tools. *ISPRS International Journal of Geo-Information*, 8(8), 363. <https://doi.org/10.3390/ijgi8080363>
- Riege, A., Gourdon-Kanhukamwe, A. & Vallée-Tourangeau, G. (2021). Covert eye-tracking: an innovative method to investigate compliance with instructions. *Psychological Research* 85, 3084–3093. <https://doi.org/10.1007/s00426-020-01451-9>
- Robinson, A. (2010). *The Look of Maps: An Examination of Cartographic Design*. Esri Press Classics.
- Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing: How To Plan, design, and conduct effective tests*. Wiley.
- Steinke, T. (1987). *Eye Movement Studies in Cartography and Related Fields*. Department of Geography, University of South Columbia, South Carolina.
- Tableau (n.d.). Schedule Refreshes on Tableau Online. Retrieved April 10, 2022, from https://help.tableau.com/current/online/en-us/schedule_add.htm
- Tableau (n.d.). What is Tableau? Retrieved February 26, 2022, from <https://www.tableau.com/why-tableau/what-is-tableau>
- Turner, C. W., Lewis, J. R., and Nielsen, J. (2006). Determining usability test sample size. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (pp. 3084-3088). Boca Raton, FL: CRC Press.
- United Nations Office for the Coordination of Human Affairs(n.d.). COVID-19 Data Explorer: Global Humanitarian Operations. Retrieved October 29, 2021, from https://data.humdata.org/visualization/covid19-humanitarian-operations/?ga1&gclid=EAIaIQobChMI3aCi08Kk8wIVxQwGAB1IfgTgEAAAYAiAAEgJl0fD_BwE
- Zuo, C., Ding, L., & Meng, L. (2020). A feasibility study of map-based dashboard for spatiotemporal knowledge acquisition and analysis. *ISPRS International Journal of Geo-Information*, 9(11), 636. <https://doi.org/10.3390/ijgi9110636>

ATTACHMENTS

LIST OF ATTACHMENTS

Bound attachments:

Attachment 1 Sequence charts for AOIs (experiment I and II)

Free attachments

Attachment 2 Poster

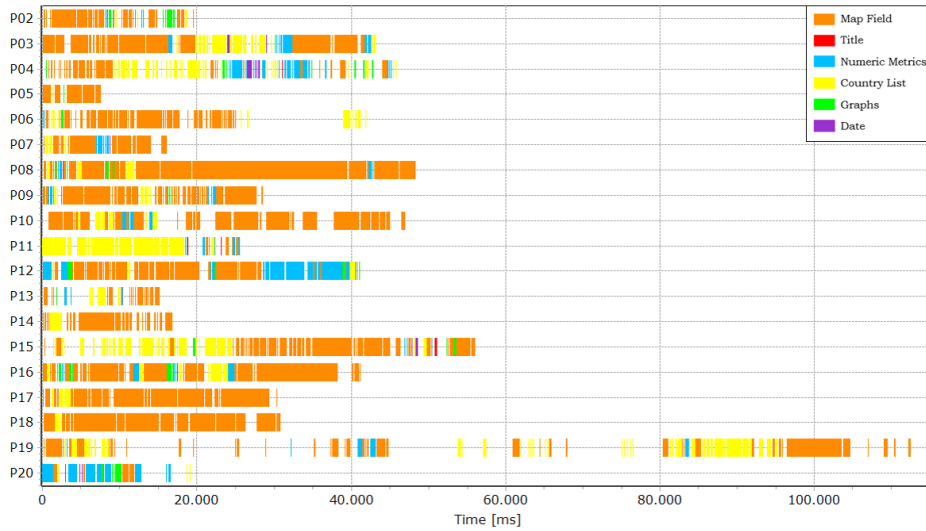
Attachment 3 SD Card

Attachment 3 structure

- Data regarding the experiments' analysis (*Data_analysis* folder)
 - Experiment1 (.R)
 - Experiment1 (.xlsx)
 - Experiment2 (.R)
 - Experiment2 (.xlsx)
- Website (folder)
- Poster (.pdf)
- Project from Tableau (*Project_tableau.twbx*)
- Back-end source code (*Source_code.py*)
- Text (.pdf)
- Eye-tracking data are stored in the Data storage of eye-tracking laboratory of Department of Geoinformatics, Palacký University Olomouc

Attachment 1

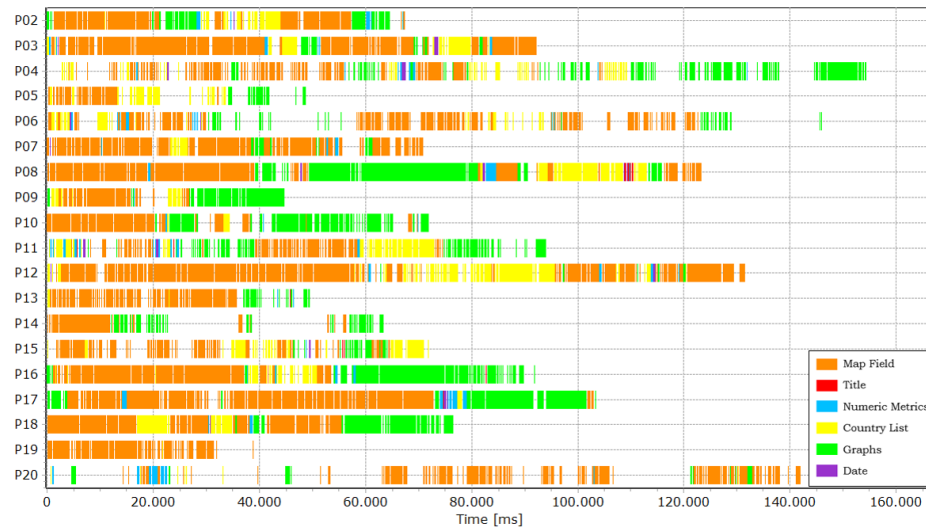
D1T1 Fixations



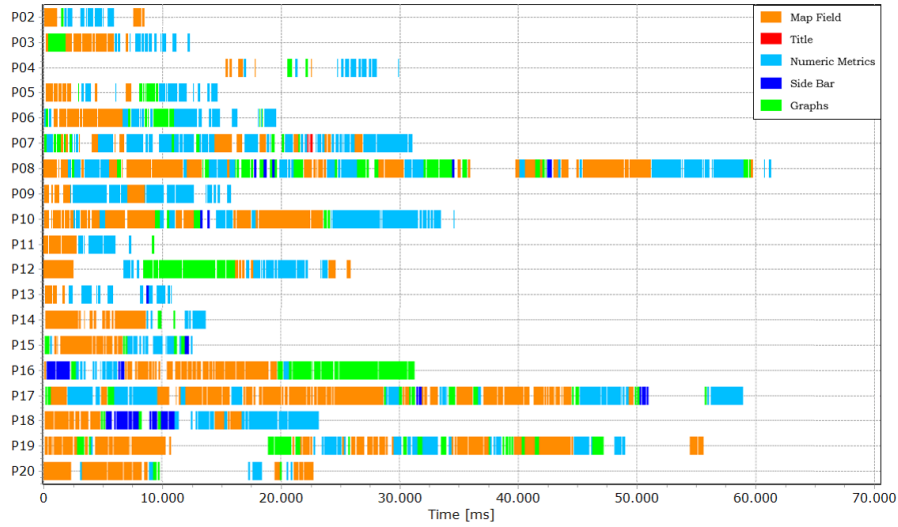
D1T2 Fixations



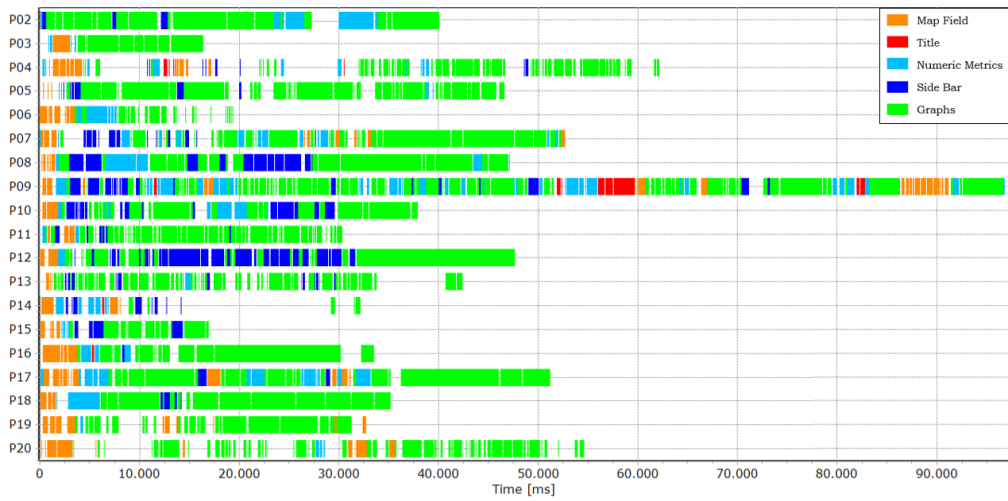
D1T3 Fixations



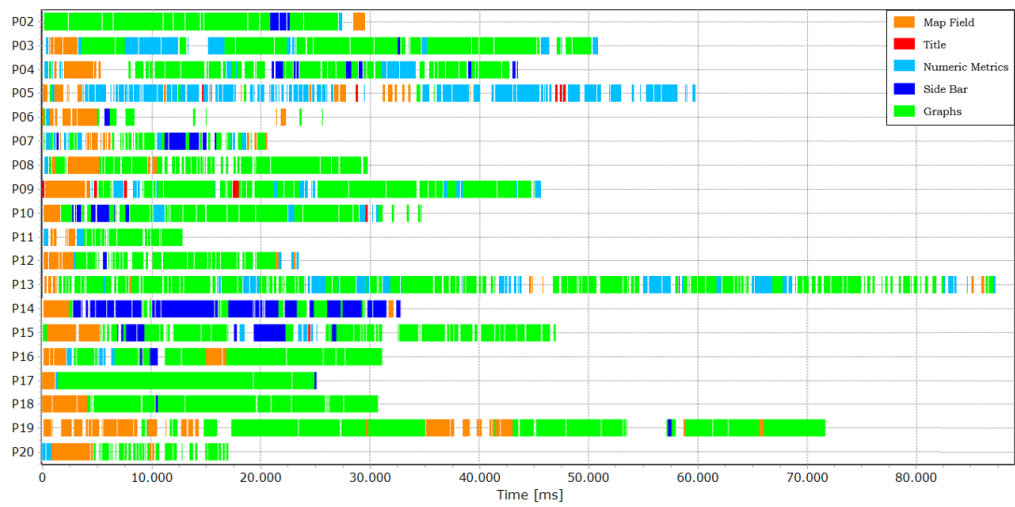
D2T1 Fixations



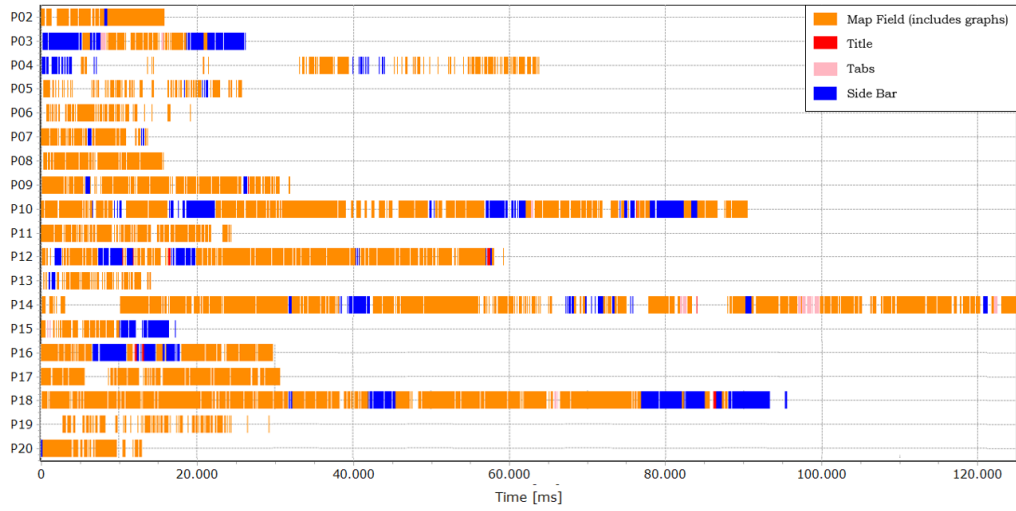
D2T2 Fixations



D2T3 Fixations



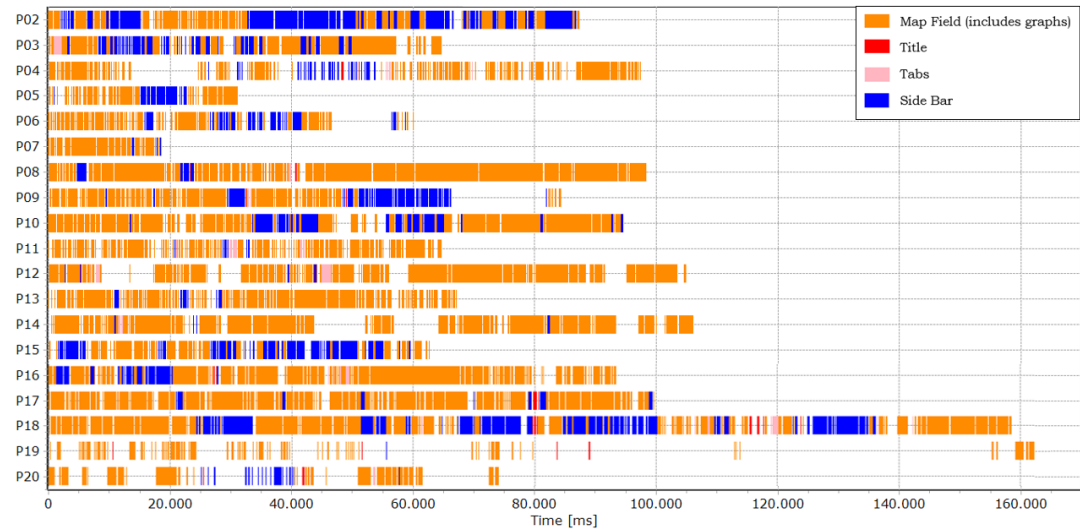
D3T1 Fixations



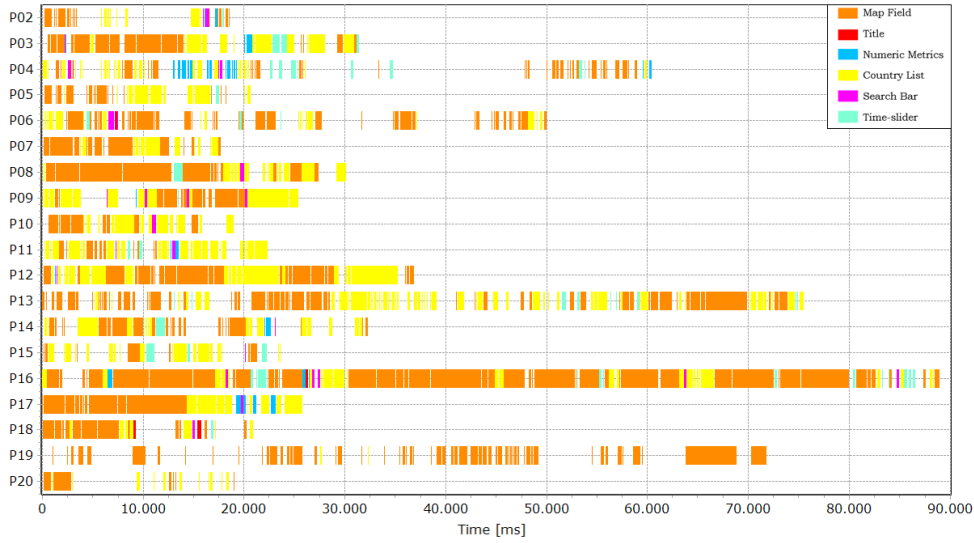
D3T2 Fixations



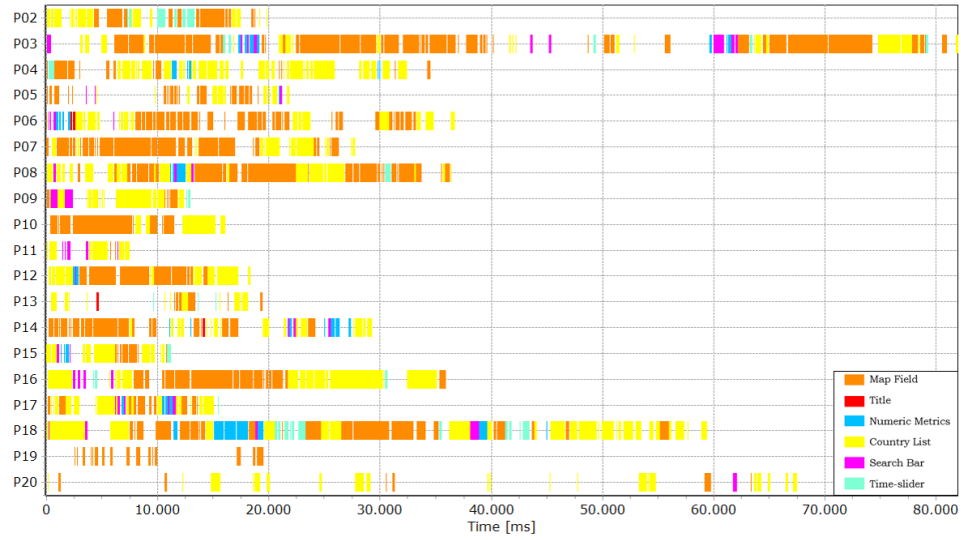
D3T3 Fixations



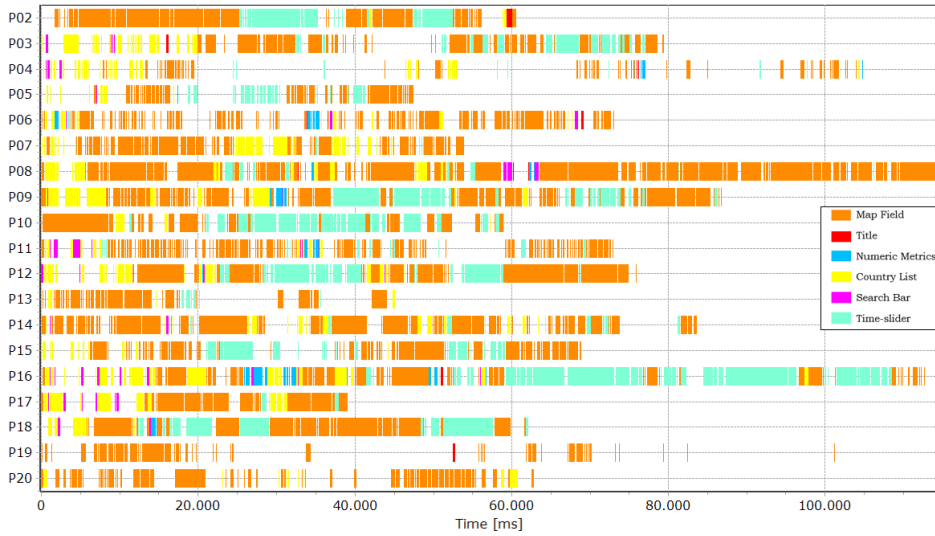
D4T1 Fixations



D4T2 Fixations



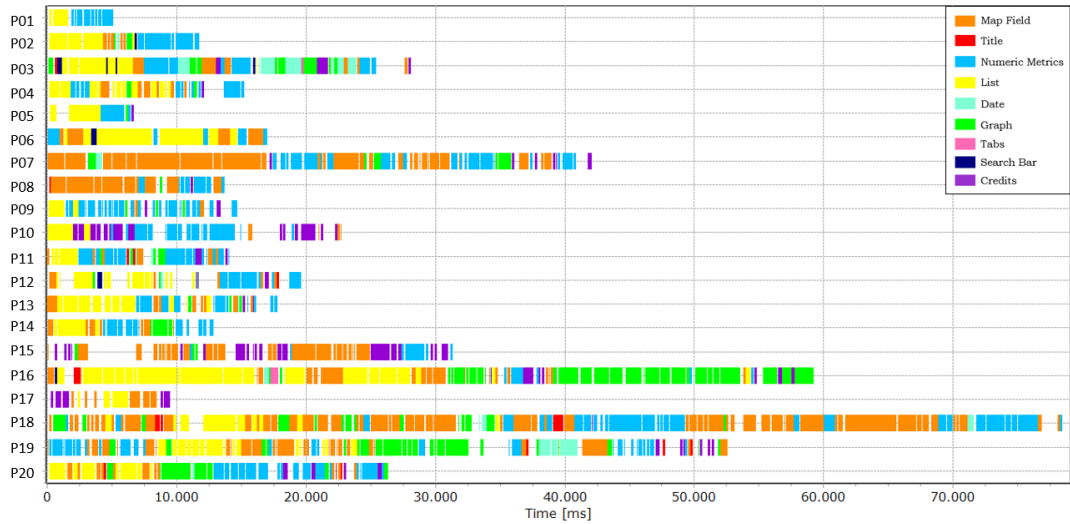
D4T3 Fixations



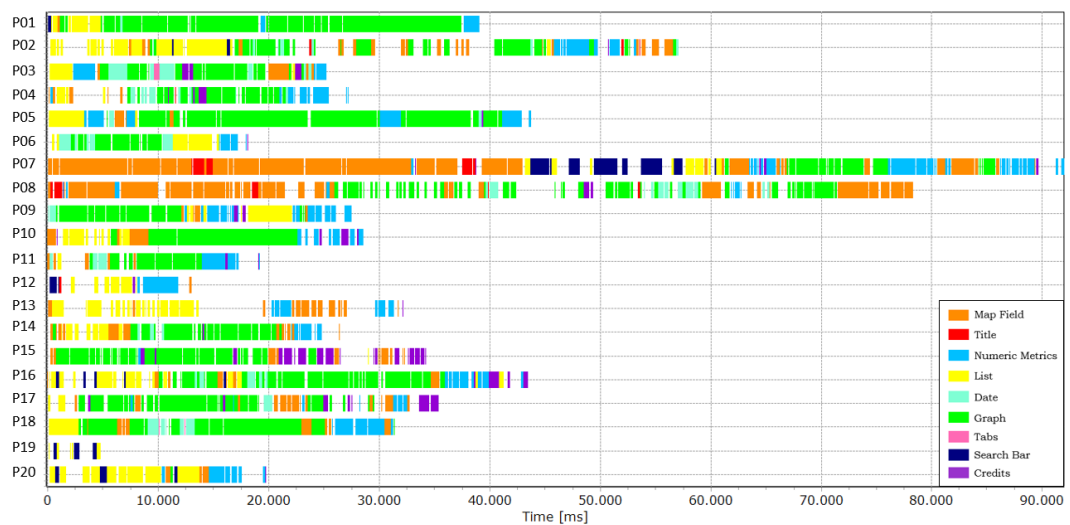
LVT1 Fixations



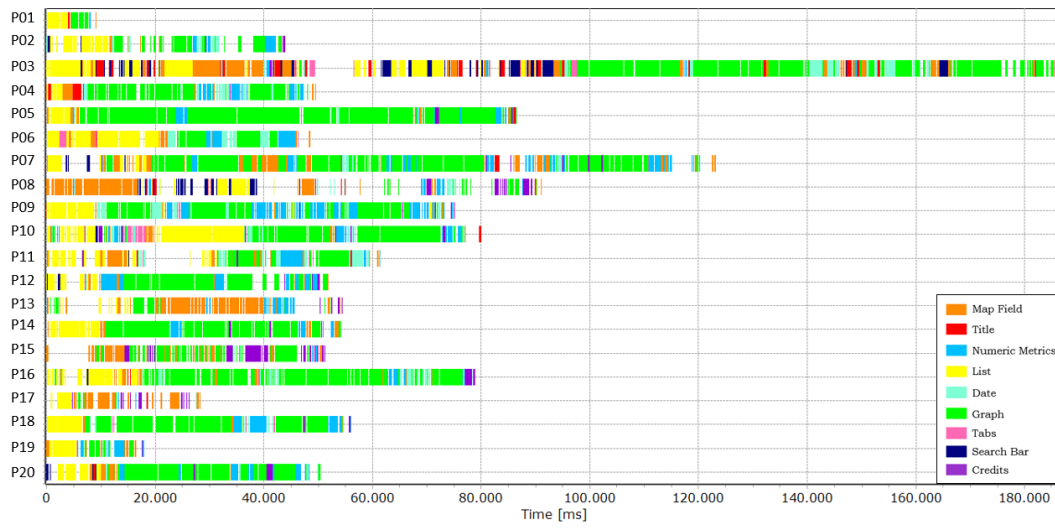
LVT2 Fixations



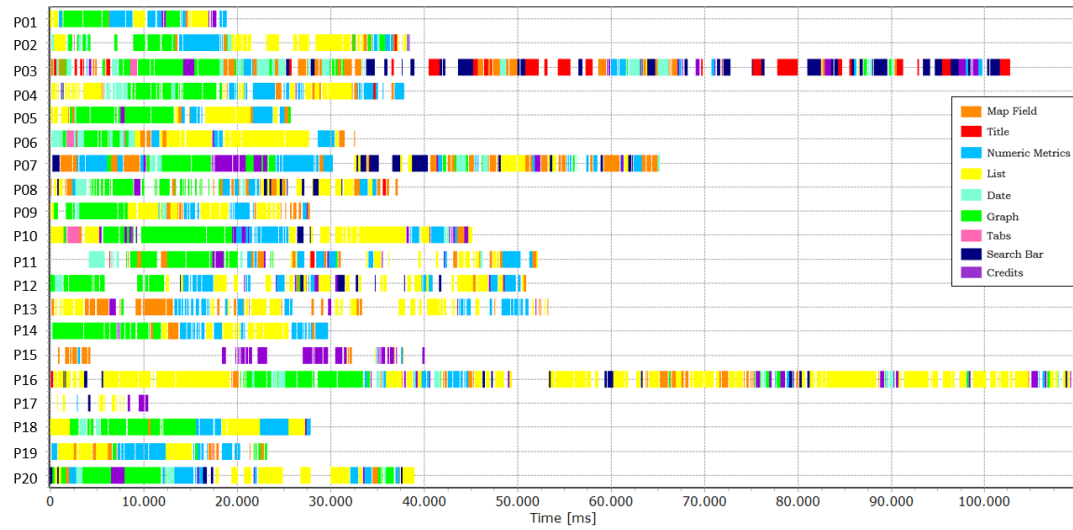
LVT3 Fixations



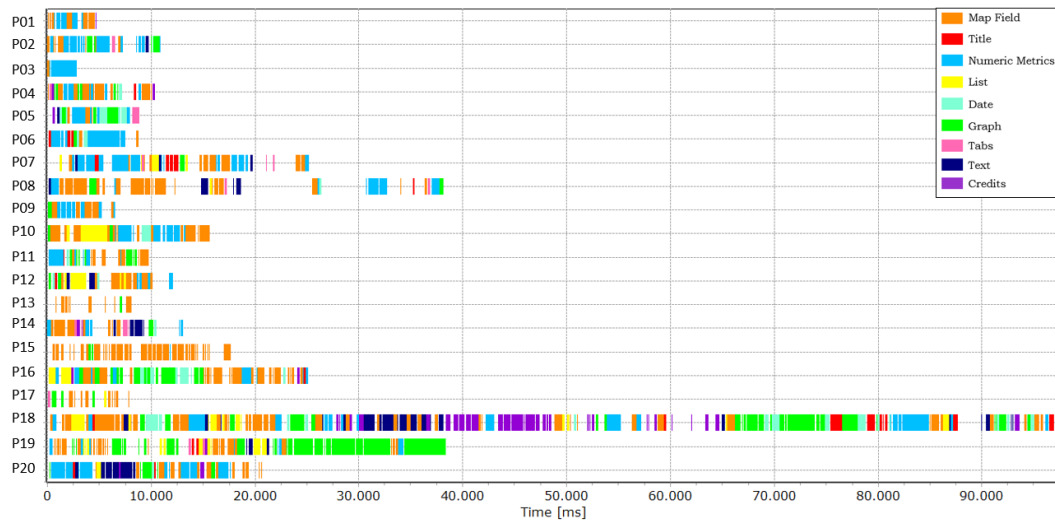
LVT4 Fixations



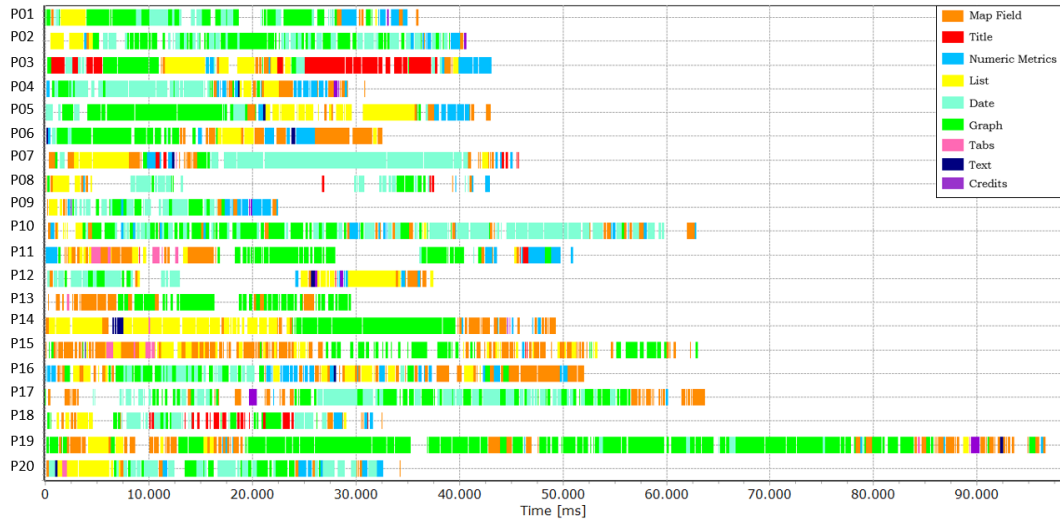
LVT5 Fixations



DVT1 Fixations



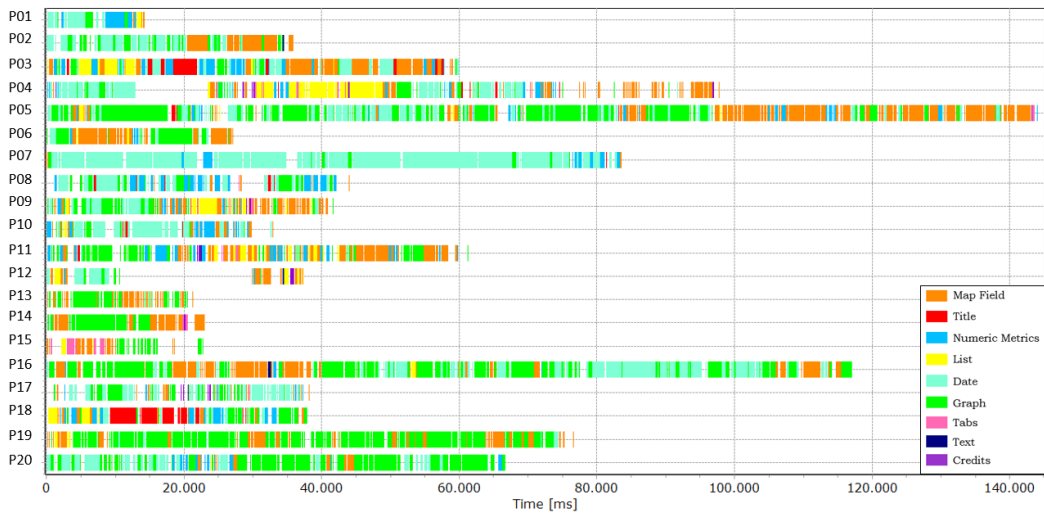
DVT2 Fixations



DVT3 Fixations



DVT4 Fixations



DVT5 Fixations

