Developing data visualization tools to assist government's health policy responses to the COVID-19 pandemic in Portugal

Ekaterina Ignatenko

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Professor Mónica Duarte Correia de Oliveira

Professor Manuel Luís Castro Ribeiro

Examination Committee

Chairperson: Professor Maria Margarida Campos da Silveira

Supervisor: Professor Mónica Duarte Correia de Oliveira

Members of the committee: Professor Maria João Correia Colunas Pereira

October 2021

Declaration

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.

Preface

The work presented in this thesis was performed at Instituto Superior Técnico during the period March-October 2021. The thesis was supervised by Prof. Mónica Duarte Correia de Oliveira and co-supervised by Prof. Manuel Luís Castro Ribeiro.

Acknowledgments

Firstly, I want to express my gratitude to Professor Mónica Oliveira and Professor Manuel Ribeiro for all the motivational support over the last several months, for constantly pushing me to go farther and achieve better. For all the words of kindness, courage, and optimism demonstrated. This dissertation would not have been possible without their advice and unwavering support.

To my parents, Olga and Igor, for making all this possible, for the greatest effort they have made and for always being there no matter what. To my little sister, for always being by my side. To Daniel, for the encouragement and for all the invaluable support. To my grandmother Maria, for always believing in me. I will forever be grateful.

Last but certainly not least, to my friends for their unconditional support and the joy they bring into my life during all these years.

Abstract

Due to the large amount of data generated by the rapid emergence of new technologies, data visualization tools, such as dashboards, have emerged, allowing the representation of information that stimulates visual perception capabilities. Despite the growing popularity of such data visualization tools, little is known about the methods used to support their design. Despite the recognized importance of users' opinion in dashboard design, the number of studies that combine participatory methods with visualizations options is very limited.

This study aims to explore methods to inform the design of data visualization tools. Specifically, so as to understand which visualisation formats should be incorporated within dashboards for the COVID-19 pandemic, research is developed to understand the preferences and views of general population about distinct data visualization formats.

A novel approach to improve the process of selecting visualization formats was developed, considering not only the theory-based evidence, but also the perception of final users. The approach incorporates a Delphi process to understand preferences and acceptability of alternative visualizations. The developed approach was applied to select appropriate data visualization formats for public web-based COVID-19 dashboards – with 47 individuals participating in the Web-Delphi process, and afterwards, preference gathered in the Delphi process were applied to the design of the DGS COVID-19 dashboard, so as that it accounts users' preferences. A platform to run the Delphi process was programmed.

The aftermath of this study is a contribute to data visualization selection literature in the context of COVID-19 pandemic and of dashboard design.

Keywords: Dashboard; Delphi process; Data visualization tools; COVID-19; Portugal.

Resumo

A grande quantidade de dados gerados pela rápida evolução das novas tecnologias, fez surgir métodos de visualização de dados, como dashboards, que permitem a representação de informação de modo a estimular as capacidades de perceção visual. Apesar da crescente popularidade desses métodos, pouco se sabe sobre os processos utilizados para auxiliar o seu design. Apesar da reconhecida importância da opinião dos utilizadores sobre o design de dashboards, o número de estudos que combinam métodos participativos com opções de visualização é muito limitado.

Este estudo tem como objetivo explorar processos que auxiliem o design de métodos de visualização de dados. Especificamente, de modo a perceber quais visualizações podem ser incorporadas em dashboards sobre a pandemia de COVID-19, é desenvolvida uma pesquisa sobre as preferências da população em relação a diferentes formatos de visualização.

Foi desenvolvida uma nova estratégia para melhorar o processo de seleção de formatos de visualização, considerando não só evidências baseadas na teoria, mas também a perceção dos utilizadores. Esta abordagem incorpora um processo Delphi para perceber as preferências e aceitabilidade de visualizações alternativas. A estratégia desenvolvida foi aplicada para selecionar métodos de visualização apropriados para dashboards públicos de COVID-19. Posteriormente, as preferências obtidas no processo Delphi foram aplicadas ao design do dashboard de COVID-19 da DGS, de modo a integrar as preferências dos utilizadores. Uma plataforma para executar o processo Delphi foi programada.

O resultado deste estudo é uma contribuição para a literatura de seleção de visualização de dados no contexto da pandemia de COVID-19 e do design de dashboards.

Palavras-chave: Dashboard; Processo Delphi; Métodos de visualização de dados; COVID-19; Portugal.

Contents

Acknowledgments *vii*

Abstract *ix*

Resumo *xi*

Contents *i*

List of Figures *iv*

List of Tables *vi*

Abbreviations *viii*

1 Introduction *10*

1.1 Thesis objectives 10

1.2 Thesis outline 11

2 Context *12*

2.1.1 Data visualization tools 12

2.1.2 Data visualization tools in healthcare 13

2.1.3 Data visualization tools and COVID-19 14

2.1.4 SCOPE - Spatial Data Science Services for COVID-19 Pandemic 16

3 Literature review *17*

3.1 Data visualization formats 17

3.1.1 Comparing categories 19

3.1.2 Assessing hierarchies and part-to-whole relationships 20

3.1.3 Showing changes over time 21

3.1.4 Plotting connections and relationships 22

3.1.5 Mapping geo-spatial data 23

3.1.6 Dashboard as a combination of multiple data visualizations 24

3.2 Dashboards 25

3.2.1 Definition 25

3.2.2 Purpose and domains 25

3.2.3 Key indicators 26

3.2.4 Development Stages 26

3.2.5 Data selection and visualization design 28

3.2.6 Target audience 29

3.2.7 Challenges 29

3.2.8 Dashboards in Healthcare 30

3.3 Participatory methods in dashboard design 30

3.3.1 Overview of participatory methods 31

3.3.2 Participatory methods for data visualization selection 33

4 Proposed approach 35

4.1 Designing a novel approach 35

4.1.1 Objectives 35

4.1.2 Designed selection process 35

4.1.3 Selection of a pre-set of data visualization formats 36

4.1.4 Participatory process 38

5 Case study design 41

5.1 Indicators' selection 41

5.2 Implementation 43

5.2.1 Selection of pre-set of data visualization formats 43

5.2.2 Delphi process 45

5.2.3 Implications of the Delphi results for dashboard design 51

6 Case study results 52

6.1 Delphi participation 52

6.2 Delphi results 52

6.3 Implications of the Delphi results for DGS dashboard design 63

7 Discussion 67

7.1 Novel approach 67

7.1.1 Implementation 67

7.2 Results 68

7.2.1 Implications of the Delphi results for dashboard design 69

7.3 Advantages of the proposed approach 70

7.4 Limitations 70

8 Conclusions 71

References 72

Annexes 85

Annex A. 85

Annex B. 86

Annex C. 88

Annex D. 90

Annex E. 90

List of Figures

Figure 1. WHO COVID-19 Dashboard (top left) [40], COVID-19 Dashboard by CSSE at Johns Hopkins University (top right) [41], COVID-19 Dashboard by CSSE (bottom left) [42], and DGS COVID-19 Dashboard (bottom right) [43]. 14
Figure 2. A process model for effective data visualization [20]. 18
Figure 3. Bar Chart [9]. 20
Figure 4. Grouped Bar Chart [54]. 20
Figure 5. Pie Chart (left) [9] and Donut Chart (right) [53]. 21
Figure 6. Stacked Column Chart (left) [53] and 100% Stacked Bar Chart (right) [9]. 21
Figure 7. Line Chart [9]. 22
Figure 8. Area Chart [9]. 22
Figure 9. Scatter Plot [9]. 23
Figure 10. Bubble Chart [9]. 23
Figure 11. Point Map (top left), Bubble Map (top right), Dasymetric map (bottom left), Choropleth Map (bottom right) [58]. 24
Figure 12. Flow diagram. 34
Figure 13. Block diagram representing the steps made for the implementation of the proposed approach. 36
Figure 14. Part 1 of first-round questionnaire – Question 1 (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering three-month time interval (March-June)?") 47
Figure 15. Part 2 of first-round questionnaire – Question 7 (trans. "What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by age group?"). 47
Figure 16. Part 3 of first-round questionnaire – Question 11 (trans. "What map do you consider to be the most appropriate to represent a number of new confirmed COVID-19 cases per municipality (incidence) since the beginning of pandemic?). 48
Figure 17. Part 1 of second-round questionnaire – Question 1 (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering three-month time interval (March-June)? Your response in the first round was: "Column Chart"; you can maintain or change your answer. First-round comments: 1. Answer: Table. I prefer to read numbers."). 49
Figure 18. Question 1 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering three-month time interval (March-June)?"). Frequency and percentage of votes for first and second rounds. 53
Figure 19. Question 2 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering one-week interval?"). Frequency and percentage of votes for first and second rounds. 53

Figure 20. Question 3 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a cumulative number of confirmed COVID-19 cases, considering three-month time interval (March-June)?”). Frequency and percentage of votes for first and second rounds. 54
Figure 21. Question 4 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a cumulative number of confirmed COVID-19 cases, considering one-week interval?”). Frequency and percentage of votes for first and second rounds. 55
Figure 22. Question 5 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases for three regions – Norte, Centro e Alentejo, considering three-month time interval (March-June)?”) Frequency and percentage of votes for first and second rounds. 55
Figure 23. Question 6 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases for three regions – Norte, Centro e Alentejo, considering one-week interval)?”). Frequency and percentage of votes for first and second rounds for Question 6. 56
Figure 24. Question 7 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by age group?”). Frequency and percentage of votes for first and second rounds. 57
Figure 25. Question 8 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by sex?”). Frequency and percentage of votes for first and second rounds. 58
Figure 26. Question 9 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by region?”). Frequency and percentage of votes for first and second rounds. 59
Figure 27. Question 10 of Delphi questionnaire (trans. “What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by sex and age group?”). Frequency and percentage of votes for first and second rounds. 60
Figure 28. Question 11 of Delphi questionnaire (trans. “What map do you consider to be the most appropriate to represent a number of new confirmed COVID-19 cases per municipality (incidence) since the beginning of pandemic?”). Frequency and percentage of votes for first and second rounds. 61
Figure 29. Question 12 of Delphi questionnaire (trans. “What map do you consider to be the most appropriate to represent a number of new confirmed COVID-19 cases per 100000 inhabitants (cumulative incidence), per municipality since the beginning of pandemic?”). Frequency and percentage of votes for first and second rounds. 61
Figure 30. DGS COVID-19 Dashboard [43]. 63
Figure 31. Adjusted DGS COVID-19 Dashboard. 66

List of Tables

Table 1. Categorization according to communication purpose (source: [9]). 19
Table 2. Decision table of data visualization format’s selection. 37
Table 3. Overview of analyzed participatory methods. 39
Table 4. Indicators Themes and their frequencies (source: [36]). 42
Table 5. Example of indicator "Cases by day". 43
Table 6. Summary of second round results of Delphi process. 62
Table 7. The percentages of how many people changed their opinion throughout the Delphi process. 62

Abbreviations

ARS	Administração Regional de Saúde
COVID-19	Coronavirus Disease 2019
DGS	Direção-Geral de Saúde
ECDC	European Centre for Disease Prevention and Control
GIS	Geographic Information System
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SCOPE	Spatial Data Science Services for COVID-19 Pandemic
SDS	Spatial Data Sciences
WHO	World Health Organization

1 Introduction

Nowadays, due to the massive data generation in all areas, new challenges are required to be able to represent relevant information and derive the value from it. In healthcare, evolving data-related concerns present analogous challenges of information integration, putting public health practitioners under more stress [1]. As a consequence, data visualization tools have emerged, allowing the representation of data that stimulates visual perception capabilities in order to magnify cognition [1]. Since visual systems are powerful mechanisms to detect patterns, they can be crucial to help healthcare decision-makers achieve new discoveries, and consequently, conduct them to the problem resolution. In this way, well-designed data visualization tools increase drastically that ability to analyse huge amounts of data and optimize its processing, even when the questions of decision-maker are still unspecified [2].

The primary goal of data visualization tools is to provide a clear vision of selected information, however, deciding on the right visualization methods is not easy at all. The effective visualization tool must be perceptible and render the main characteristics of the metric in question, otherwise, it can be misleading and guide to wrong conclusions. The evolution of visualization methods culminated in creation of dashboards, which incorporate different visualization formats on a single screen in order to facilitate understanding of information.

Despite the growing popularity of data visualizations tools, little is known about the methods to support their design for dashboard Not only there is a few number of studies regarding the data visualizations tools, but also existing ones do not provide recommendations for theorizing the selection of data visualization formats for dashboard [3]. Moreover, the design of data visualization tools is confirmed to be inextricably linked to end-users' perception and attention [4], [5]. Despite the significance of users' input for dashboard design, up to our knowledge, the number of research that utilize participatory techniques to investigate dashboard visualizations is very low.

In this thesis, attention is given to methods for selecting data visualization formats in the context of dashboard design and applied to information about the COVID-19 pandemic. The research is performed to address the limitations of current dashboards, namely the lack of a well-founded rationale for selecting appropriate visualizations, either considering theory-based evidence and the perspective of dashboard end users.

1.1 Thesis objectives

The main goal of this thesis is to explore methods to inform the design of data visualization tools and select the data visualization formats, that can be incorporated in dashboards and used to transmit relevant information about COVID-19 pandemic. This way, the research will be developed regarding the preferences and views of general population for distinct data visualization formats. Specifically, the visualization literacy of dashboard design will be investigated to identify most appropriate data visualizations for selected COVID-19 data, in order to increase the learning outcome and improve dashboard effectiveness. Furthermore, the research will give insights on differences in stakeholders' views that may go unnoticed when focusing on individual opinions, highlighting the need for involving different stakeholders not only in the data collection for decision-making dashboard, but also in visual formats' selection.

1.2 Thesis outline

The present dissertation is structured into nine chapters. After this introduction, Chapter 2 sets the context by giving scientific background about data visualization tools and its applications in healthcare and specifically, in COVID-19 pandemic.

The objective of Chapter 3 is to provide an understanding of existing literature about data visualization tools in order to identify main limitations and areas for future improvement. Chapter 4 develops a novel approach to assist in the selection of the most appropriate data visualization formats for dashboard indicators. The application of the developed approach to improve the design of the public web-based COVID-19 dashboards is described in Chapter 5, and the results of this implementation are presented in Chapter 6. It also describes real-life application of case study results to DGS COVID-19 Dashboard with the objective to analyze and improve its design. The results, advantages, and limitations of the established approach are discussed in Chapter 7, and finally, Chapter 8 presents final conclusions about the work developed in this thesis and reflects upon future work and possible methodology improvements.

2 Context

In this chapter key concepts about data visualization tools and its applications in healthcare are introduced. Specifically, a brief description of relevance of data visualization tools is elaborated. Also, the context of COVID-19 and its contribution for the development of data visualization tools are introduced.

2.1.1 Data visualization tools

Advances in technology and computer graphics have introduced a multitude of techniques for visually representing data, offering a wide choice of visualization methods to enable users to obtain information from a data set. Here, the dashboard emerges as a combination of multiple data visualization formats to improve efficiency of information transmission. Dashboard designers face the problem of presentation format as there are alternative ways of displaying metrics and trends on a dashboard. Therefore, it is necessary to obtain guidance to select data visualization formats for certain indicators [4], [6].

Research on dashboards and visualization principles is still in its early stages, with just a few publications starting to appear recently, like Yigitbasioglu & Velcu [7]. Some design and presentation recommendations can be found in information systems literature, including Cairo, 2012 [8] and Kirk, 2019 [9]. Few [10] provides a detailed survey of contemporary software-based data visualizations but it only goes to a descriptive level and makes no recommendations for theorizing or reflective formalizations concerning data representations in dashboards. On the same level of abstraction, there are a number of contributions in the literature that focus on data or information visualization, such as [9], [11], [12]. These studies are primarily concerned with visual design and do not integrate their findings into a broader focus of reasoning about developing software-based visualization techniques for improved information systems [13].

Therefore, the majority of studies regarding dashboard design does not focus specifically on selection of data visualizations without using standardized guidelines, relying only on the opinion and perception of the developer [4], [14], [15], [16], [17]. Some articles select visual formats for certain type of dashboard based on the previous studies, so the existing methodologies are highly specific [2], [18], [19]. Additionally, some dashboard designs focus on goal orientation approach and select the visualization format according to its intended purpose and the typology of outcome expected from the final user [4], [5]. However, this approach is very limited and does not provide specific guidelines regarding the choice of presentation formats to be used.

Despite its high relevance, the theoretical reflections on visualizations are rarely performed. This leaves an open research question on how to describe and depict visualizations independently from their concrete graphical appearance, but that are specific to the information that they convey [13]. Previous research studies on theoretical frameworks to guide visualization choices for diverse dataset domains suggest a variety of classifications to consider when creating visual representations in general [4]. In some approaches, the intrinsic characteristics of the dataset are employed as a main directive for the visualization choice, considering that the structural correspondence criterion is a necessary condition without which the visualization cannot be guaranteed to be effective [9], [20], [21].

Therefore, many researchers in the field of information visualization use a technical approach to identify appropriate data visualizations by categorizing them according to data properties. Earlier studies by Bertin [22], Tufte [23], and others enumerated different types of visual representations and connected them with intrinsic characteristics of dataset. Moreover, some suggested that the selection of appropriate display formats for users is based on certain rules, i.e., on structure of data and whether the data attributes are categorical or quantitative [9], [21], [24]. However, in majority of cases this approach is not applied to dashboard design in a systematic way, without providing a step-by-step guideline to determining the most appropriate visualization type for a particular situation.

As a new subject, it is expected that some studies focus on the user's perception of the dashboards [25]. This way, the choice of visualization format to use depends not only on the types of data to be displayed, but also on the types of insight people are expected to be able to gain. Since the design of data visualization tools is highly related with learning theories of cognition, the choice of interactive components such as charts and maps, depends on end-users' perception and attention [25]. Therefore, some studies integrate participatory methods, such as surveys, workshops, Delphi, etc., where stakeholders are directly invited into the design process [26], [27], [28]. However, despite the highlighted importance of stakeholder's opinion for dashboard design, the number of studies that integrate participatory methods to explore visualization of dashboard is very limited and the adopted participatory methods are not transparent.

According to the collected information, the challenges regarding the design of data visualization tools were identified. The main concerns are specified below.

- The lack of a use of a transparent step-by-step approach as a mean to support and ease the dashboard design. In this specific case, it could be very useful a process that could help reinvent and structure the process of data visualization format's selection in a more competent and effective way, providing an evidence-based guidance.
- Very limited research regarding the use of participatory methods for data visualization format's selection.

2.1.2 Data visualization tools in healthcare

In healthcare, constantly growing data-related issues present similar challenges of information integration and visualization, leading to increased pressure on public health practitioners. Since health data is complex, the visualization of data with associated geospatial information can be very efficient in discovering the patterns and trends of population health. Therefore, data visualization tools in healthcare can give overwhelming number of opportunities for policy development and academic research, and different studies demonstrate the effectiveness of incorporating visuals in health and medical communication [29], [30].

However, the development of healthcare visualization tools is a complex task due to the heterogeneity and complexity of data. A selection of geospatial, temporal and text information to integrate within a visual context involves collaborative processes with multidisciplinary research groups and wide range of stakeholders [17]. These processes are particularly important as decisions are made collaboratively across various situational and personnel contexts. However, they are very time-consuming and difficult to perform, and little research exists regarding the contribution of these collaborations for tool development and the preferences and views of final users for distinct data and visual formats [31].

Consequently, according to literature, there are few data visualization tools for healthcare that have been created as a result of collaboration with stakeholders [32]. It is therefore essential to identify effective practices for visualization selection and use and investigate the opinion of stakeholders regarding the implementation of certain visualization elements.

2.1.3 Data visualization tools and COVID-19

In December 2019, a local outbreak of initially unknown respiratory illness was detected in Wuhan (Hubei, China) and was rapidly identified to be caused by coronavirus SARS-CoV-2. On 11 March 2020, coronavirus disease 2019 (COVID-19) was announced a pandemic by the World Health Organization (WHO) [33]. Today (25/10/21), 245 million confirmed cases and almost 5 million deaths are identified, and consequently, there are severe concerns about the social, economic and health impact of this virus [34], [35].

Therefore, different policies for detection and containment of clusters of infection were established to control the propagation of pandemic throw community transmission. For that reason, many data visualization tools for decision support, such as web-based dashboards have been developed to facilitate the transmission of relevant information to the general population and comprehension of the COVID-19 data by the public. Therefore, the public web-based COVID-19 dashboards ultimately have shared a common objective: to serve as both a communication tool and call for individual and collective action to respond to the COVID-19 pandemic [36]. This type of solution is essential to guarantee the monitoring of the disease evolution, not only regarding the number of new cases and deaths but also

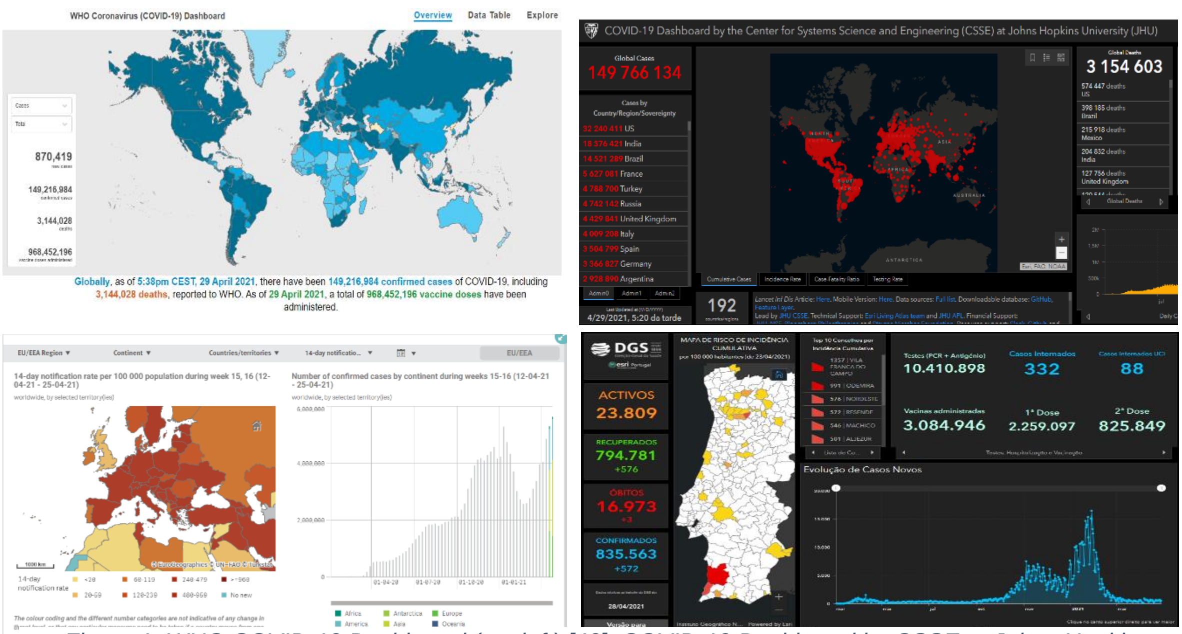


Figure 1. WHO COVID-19 Dashboard (top left) [40], COVID-19 Dashboard by CSSE at Johns Hopkins University (top right) [41], COVID-19 Dashboard by CSSE (bottom left) [42], and DGS COVID-19 Dashboard (bottom right) [43].

other indicators and at different geographic scales, this way enabling the public to understand current situation and develop certain behavior [37]–[39].

The first web-based public COVID-19 dashboards were published by WHO and Johns Hopkins University, to represent the pandemic situation by illustrating number of confirmed cases and deaths for each country and worldwide [40], [41]. Another globally notable dashboard was developed by European Centre for Disease Control and Prevention (ECDC) to monitor disease progression specifically within the European Union [42]. In Portugal, Direção Geral da Saúde (DGS) developed a dashboard which integrated risk map of cumulative incidence for each region, and linked graphs to visualize evolution of cases in country [43].

Most existing web-based public COVID-19 dashboards have focused on epidemiological data, e.g., number of confirmed cases and mortality. Less frequent indicator themes have been related to hospital care (admissions to infection control units), testing (total test, testing rates), and spread and death (recovered and active cases [36]. However, only the epidemiological data can be insufficient to discover more specific relationships regarding the influence of other factors to coronavirus propagation. Therefore, the demographic and socioeconomic indicators, such as age, social deprivation, population density and ethnicity, started to be adopted by COVID-19 dashboards, although not frequently reported [36], [38].

COVID-19 dashboards predominately report indicators on different geographic levels, representing spatial variations of pre-determined COVID-19 indicators among different regions. On the other hand, these indicators are also represented over time, predominantly by day, identifying temporal differences among certain periods of time to show the effects of implemented policies, such as stay-at-home strategy [39], [37]. In addition to geographic and temporal breakdowns, dashboards are used to analyze data by other breakdowns, the most common of which are age and sex [36].

Visualization techniques have been front-and-center in the efforts to communicate the science around COVID-19 to the general population. By definition, public web-based COVID-19 dashboards utilize the data visualization formats to transmit indicators in the most perceptual way, usually in the form of graphs or charts, maps and tables [36]. However, an overall underuse of known and proven delivery visual techniques is a common issue for COVID-19 dashboards, that can mislead both unintentionally and intentionally if the data visualizations are not accurately represented [44].

The use of spatial analysis in tracking and understanding the propagation of infectious illnesses is well-known, reaching the peak of its' popularity with development of Geographic Information Systems (GIS), which significantly increased the opportunity to visualize and detect disease patterns [34]. The spatial epidemiological data related to COVID-19 is usually based in aggregated data by region or area (e.g. counts or rates) and represented by bubble or choropleth maps [37]. The first one uses bubbles located at the center of each region depicting coronavirus indicators values, e.g., total number of confirmed cases, by the radius of the correspondent bubble. Alternatively, the choropleth map utilizes variations in coloring to depict different values of COVID-19 indicators associated with each region or area. Moreover, the maps can be complemented with other data visualization formats, e.g. column or bar charts, that are frequently used to compare certain information between different regions [37].

On other hand, temporal information (e.g., daily or cumulative numbers of new cases, confirmed cases and deaths) is usually represented by line graphs, column charts (or stacked column charts), and area charts. The representation of breakdowns such as age and sex is highly diverse, utilizing the wide range of visualization formats not only to compare categories, but to assess the part-to-whole relationship [45].

This type of dashboards represents several challenges regarding their design. The review of 158 public web-based COVID-19 dashboards demonstrates that clear reporting of a dashboard's purpose (its "why") and audience (for "whom") was infrequent [36]. On the other hand, an overall underuse of known and proven data delivery techniques, e.g., visualization techniques, was found. Therefore, this study highlights that there is no single approach to develop a dashboard and suggest that introducing certain features, such as "use storytelling and visual cues" to improve interpretation may enhance the dashboard efficacy. A study also underlines that the perspective of the target audience is needed to be investigated to obtain insights from firsthand use, by highlighting "know the audience and their information needs" feature [36]. Furthermore, one can assume that the web-based public COVID-19 dashboards represents the same challenges as dashboards in general.

Therefore, the challenges hereby described can be surpassed by applying the novel approach developed in this thesis, to improve the efficiency of web-based public COVID-19 dashboards. In other words, the final data visualization tool will be validated not only by evidence reported in literature, but also by incorporating the values and preferences of the population, crucial for visualization's acceptance and legitimacy. This way, the approach guarantees that the developed data visualization tool is adequate, functional, and meets the needs of end-users.

2.1.4 SCOPE - Spatial Data Science Services for COVID-19 Pandemic

Spatial Data Sciences (SDS) can be used to understand and predict spatial patterns of infectious disease and its transmission dynamics, for subsequent development of optimal resources allocation strategies and control interventions. The SCOPE project (Spatial Data Science Services for COVID-19 Pandemic, from the FCT call "AI 4 COVID-19: Data Science and Artificial Intelligence in the Public Administration to strengthen the fight against COVID-19 and future pandemics - 2020", 2021-2023) aims to utilize SDS to develop a functional software that can be used to manage spatial risk during epidemic events and inform policy analyses. The SCOPE project already provides daily updates on maps with local averages of infection risk, as well as the uncertainty associated with the spatial predictions. These updates help inform the strategies and actions needed to manage the pandemic.

The COVID-19 maps produced by the SCOPE project are included into a group of maps entitled "Disease mapping", which gives approximate information on the infection risk occurring on certain location. Normally, the COVID-19 maps that are being created to inform the public are defined by the administrative unit, with constant infection risk value within each area. However, the maps produced by the SCOPE depict spatially varying infection risks distributed over a spatial continuum. This way, a map appears as the gradient of risk values, not exhibiting sharp discontinuities at the limits of each area.

Although these maps are used by health authorities, the SCOPE project has the objective of developing new tools to spatially analyze the evolution of pandemic events, and to enable an analysis of policy, being important to understand if the maps transmit information to the general population in friendly formats. This thesis aims to develop methods to inform the SCOPE project.

3 Literature review

This chapter has the goal of presenting key concepts and literature in order to frame the available options and address issues identified in the previous section. The first section will cover a vast spectrum of data visualization formats, emphasizing the underlying benefits of their implementation in decision-making tools. Furthermore, the possibility of combining data visualization formats in dashboards will be explored. Afterwards, a review of the dashboard as a visualization tool will be provided in second section, specifically, its definition, purposes and development stages. Moreover, the second section will cover the applicability of dashboards in healthcare and the main challenges related with the development process. Lastly, an overview is made regarding the use of participatory methods in dashboard design, enhancing the importance of stakeholders' participation in the development process. Therefore, the concept of participatory approach is briefly described, and the most common participatory methods are enumerated. In the last part of section, the application of participatory methods to data visualization formats' selection for dashboards is investigated by gathering the scientific articles that have explored this issue.

3.1 Data visualization formats

Human visual capabilities are a great instrument to identify trend and patterns [46]. Therefore, the proper visualization significantly increases the perception of data elements, stimulating human´s perceptual and cognitive abilities of problem solving [7].

The sudden increase in the quantity of data available in the last ten years has created new analytic challenges regarding data visualization formats, or data visualizations, defined as a "representations of data in some systematic form, including attributes and variables for the unit of information" [47]. Data visualization format allows to use visual elements, with the purpose of representing large amounts of information in the way that facilitates data interpretation. Through data visualization, the data is depicted in ways that allow viewers to experience it in a new light, exploring the unseen patterns and relationships within the data. Selection of data visualization format is not an easy task, since it requires a deep knowledge in various subjects, including cognitive science, statistics, graphic design, cartography, and computer science [9].

The selection of appropriate data visualization format to satisfy a specific objective is essential for effective decision-making. [30], [48]–[50] According to Sedrakyan [4], good data visualization needs to be clear, appropriate and memorable for the chosen audience, without containing a redundant information. The modern technology and graphics offer a large choice of techniques for visual representation, however the selection of the most suitable visualization for a given dataset is highly subjective and complex.

Despite of growing popularity, there is essentially no design rationale for visualization's selection in existing theory, lacking proper grounding of the design of visual languages. The current state shows that there are significant limitations to the methods and approaches used to select data visualizations. Therefore, development of methodological standards for data visualization selection can provide foundation to specify a rationale for selection of choices [51].

The traditional view holds that visualizations can be selected by mapping conceptual elements (e.g., datasets) and visual representations (e.g., charts and graphs). In this approach, the essential properties of a visual representation are abstracted from its individual graphical expressions, such as shape, color or position, and focus on the underlying data structure [52].

Therefore, various studies state that data visualization format's choice depends on the nature of the data set [20], [22], [24], [53]. The data visualization process model for effective data visualization proposed by Dastani [20] highlights two main steps (Figure 2):

1) Determine the data structure.
2) Determine visual elements that represent data elements in such a way that the perceptual structure of the decided visual elements represents the structure of the data.

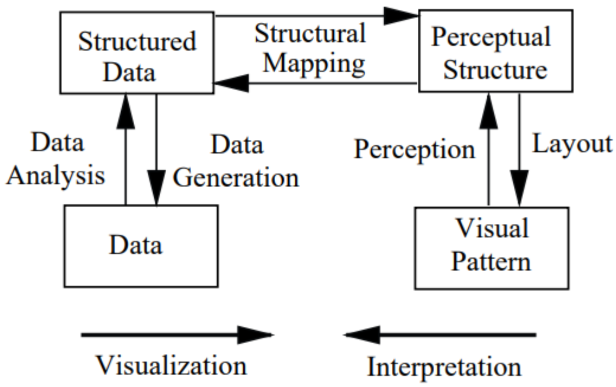


Figure 2. A process model for effective data visualization [20].

Identification of the data structure is not a trivial decision, since there are different structural properties that can be defined for certain data set. The data familiarization and preparation need to be done, and afterwards the fundamental structure must be understood. In order to facilitate the data set structuring and identification of variables, a data set can be structured in the form of table that are made up of rows and columns, where each column correspond to the certain data set variable, and each row represent an item [9], [21].

Several studies ([13], [22], [24]) suggest the identification of variables' type in order to proceed with the refinement of data visualization choice. According to Jacques Bertin's, "Semiology of Graphics" [22], structural data variables comprises data values that can be classified as categorical or quantitative. A categorical variable contains data rows with a limited number of distinct values, or a limited number of aggregated ranges of quantitative values. In contrast, a quantitative variable corresponds to a dataset's column whose row values can take on a full range of numeric values.

The second step of Dastani model integrates the selection of correct visualization method, taking into account the physical properties of the data and mapping them to perceptual structure of visualization, considering the nature of variables. However, Kirk [9] outlines that the choice of visualization method must examine not only the data structure, but also the final outcome intended with the use of data visualization format. Therefore, the visual analysis work needs to be performed to explore the data set and identify the possible comparisons, trends, patterns, and relationships that the data visualization format is intended to demonstrate.

In this respect, the taxonomy suggested by Kirk [9] categorizes the data visualization formats by the primary communication methods, focusing on variety of possible outcomes (Table 1):

Table 1. Categorization according to communication purpose (source: [9]*).*

Method classification	**Communication purpose**
Comparing categories	To facilitate comparisons between the relative and absolute sizes of categorical values.
Assessing hierarchies and part-to-whole relationships	To provide a breakdown of categorical values in their relationship to a population of values or as constituent elements of hierarchical structures.
Showing changes over time	To exploit temporal data and show the changing trends and patterns of values over a continuous time frame.
Plotting connections and relationships	To assess the associations, distributions, and patterns that exists between multivariate datasets.
Mapping geo-spatial data	To plot and present datasets with geo-spatial properties via the many different mapping frameworks.

This categorization is flexible, since there are examples of a charts that spans across two method classifications. Kirk [9] suggests that the choice of visualization format must be firstly driven by its desired outcome, and afterwards, the accommodation of physical properties must be performed. This way, the nature of the variables in question reduces the range of suitable chart types within each method family.

In the next sections, an organized collection of some of the most common chart types and graphical methods being used today is presented, taking in account taxonomy presented by Kirk. It is to consider that many of the chart types represent several presentational characteristics and could be included into more than just one classification method. Additionally, each chart type refers the type and quantity of typical data variables normally used with it, with the objective of posterior refining by dataset nature.

3.1.1 Comparing categories

Bar and Column Charts

Data variables: 1 x categorical, 1 x quantitative.

Bar chart is one of the most commonly used data visualization formats, being utilized for comparison of values of quantitative column across values of categorical column. The categorical values are grouped by their x-axis label and are represented by bars, where their length is proportional to the quantitative values that they represent. Column chart is analogous to bar chart, except the x-axis and y-axis are swapped so that the bars are represented vertically instead of horizontally. This chart can be also used to show evolution of data over discrete time intervals [47], [50], [53].

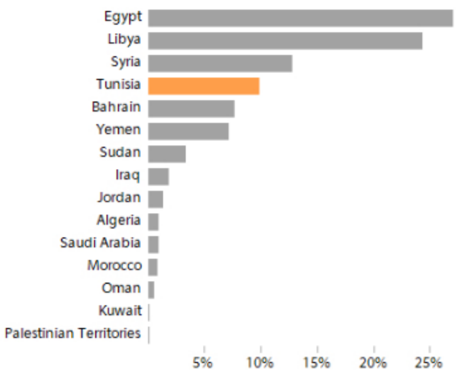


Figure 3. Bar Chart [9].

Grouped Bar and Column Charts

Data variables: 2 x categorical, 1 x quantitative.

A grouped bar chart extends the bar chart, plotting quantitative values for levels of two categorical variables instead of one. Bars are grouped by position for levels of one categorical variable, with color indicating the secondary category level within each group. When constructing a grouped bar chart, one of the most essential considerations to make is to decide which of the two categorical variables will be the primary variable (dictating the axis positions for each bar cluster) and which will be the secondary variable (dictating the number of bars to plot in each cluster). The grouped column chart can also be used to compare distinct items at distinct time intervals, showing the trends over time [53].

Another variation, bi-directional or two-sided bar chart, compares two quantitative values of one secondary categorical variable by "growing" two columns from the axis in the center in two opposite directions for each primary category [21].

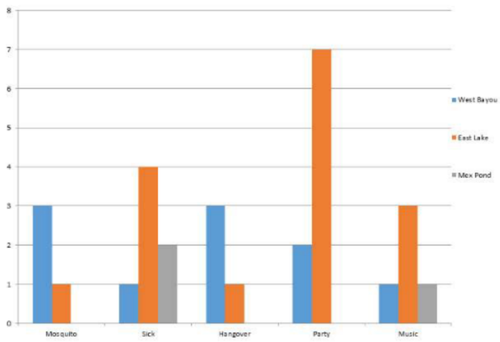


Figure 4. Grouped Bar Chart [54].

3.1.2 Assessing hierarchies and part-to-whole relationships

Pie and Donut Charts

Data variables: 1 x categorical, 1 x quantitative.

Pie chart is represented by circle divided into different sectors, where each sector shows the proportion to the whole quantity. Therefore, the labels for the sectors correspond to categorical column values, while the quantitative column values are summarized into contribution per the categorical column value.

One of the variants of the pie chart is a donut chart, with a hole in its center, displaying categories as arcs rather than slices [47], [50], [53].

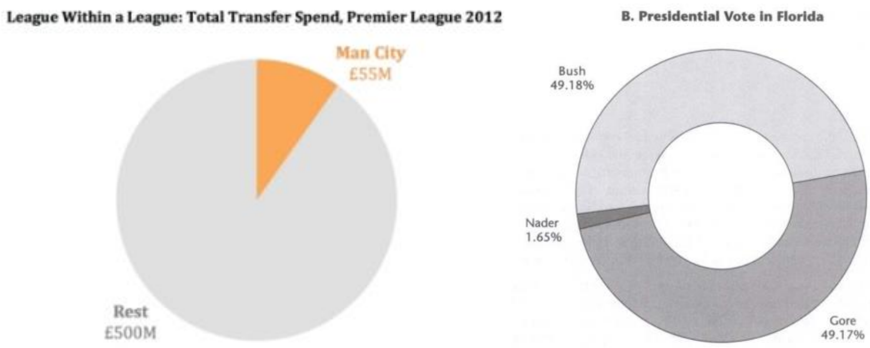


Figure 5. Pie Chart (left) [9] *and Donut Chart (right)* [53].

Stacked Bar and Column Charts

Data variables: 2 x categorical, 1 x quantitative.

The stacked bar chart extends the regular bar chart from observing quantitative values across one categorical variable to two. Each bar in a standard bar chart is divided into a number of sub-bars stacked end to end, each one corresponding to a level of the second categorical variable. As in the case of grouped bar chart, one must choose which of the two categorical variables will be the primary variable (dictating major axis positions and overall bar lengths) and which will be the secondary (dictating how each primary bar will be subdivided) [21]. Stacked column charts can be also used to show the part-to-whole relationship of distinct categories at distinct time intervals, demonstrating evolution over time [53].

Additionally, 100% Stacked Bars shows the percentage contribution of each value of quantitative column to the corresponding sum of values for each category [53].

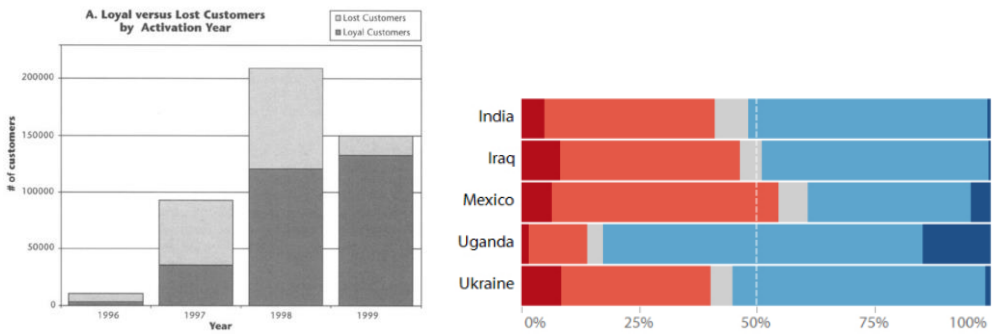


Figure 6. Stacked Column Chart (left) [53] *and 100% Stacked Bar Chart (right)* [9].

3.1.3 Showing changes over time

Line Chart

Data variables: 2 x quantitative, 1 x categorical.

Line graph represents a set of data points plotted in two-dimensional coordinate system, connected by line segments. Line graphs enable to visualize a trend in data over time interval, and to compare a continuous quantitative variable on the x axis and the size of values on the y axis [47], [50], [53].

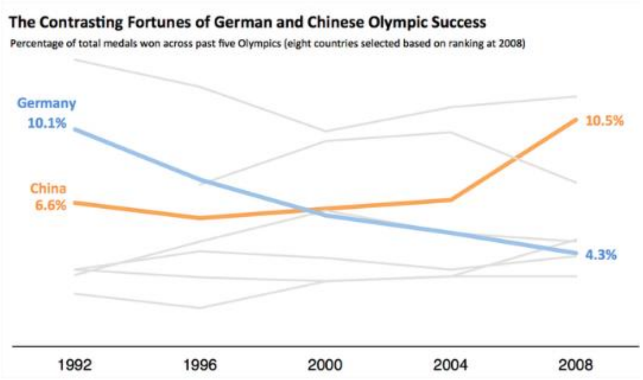


Figure 7. Line Chart [9].

Area Chart

Data variables: 2 x quantitative, 1 x categorical.

Another variation of line chart is an area graph, which is usually used to discover the relative contributions of the values of each quantitative column by emphasizing the total of all values combined [53].

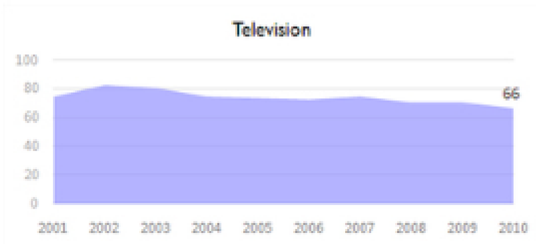


Figure 8. Area Chart [9].

3.1.4 Plotting connections and relationships

Scatter Plot

Data variables: 2 x quantitative.

Scatter plot is described as two-dimensional plot which displays data in Cartesian coordinate, showing the relationship between values of two or more quantitative columns, where one is represented as a vertical distance (independent variable) and the other as horizontal distance (dependent variable). It enables to examine trends in the data and identify outliers in a simple way [47], [50], [53].

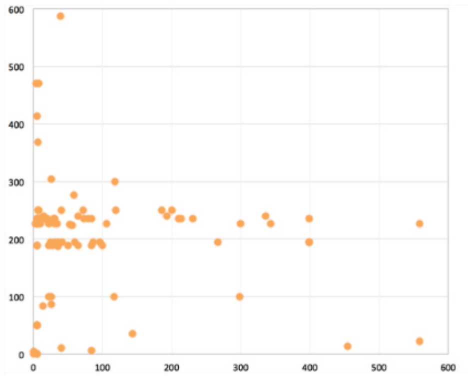


Figure 9. Scatter Plot [9].

Bubble Chart

Data variables: 3 x quantitative.

Bubble chart is another variation of a scatter plot where dot is substituted with a bubble. This situation is possible when the set of data points contains three quantitative values for each data item. This way, the relationship between three variables can be established, since two of them are represented by the plot axes, and the third one by the bubble size [47], [50], [53].

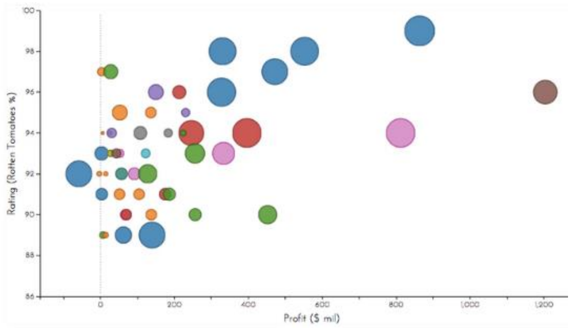


Figure 10. Bubble Chart [9].

3.1.5 Mapping geo-spatial data

Map visualizations allow to investigate spatial, or geographical, relationships, when dataset contains a geographic data dimension. Therefore, each column value is mapped on a visual map based on a corresponding spatial key [53].

Evolution of map technologies and a growing number of spatial decision problems triggered the development of Geographic Information Systems (GIS). GIS combines geospatial data from a broad variety of sources, allowing various types of spatial analysis. Map visualizations are playing a big role in GIS, since they not only interactively display the geospatial indicators provided by GIS, but also present the final results of the spatial analysis, facilitating the posterior process of decision-making [55]

The most common visual representations of geospatial indicators can be categorized into point, bubble, choropleth and dasymetric maps. Point maps use point symbols to represent locations of individual-level events (e.g., residential addresses in case-control studies), that can be colored accordingly to attribute value. In bubble map, the focus relies in representing aggregated data by region or area, using a filled circle located at the center of each region, where the symbol size varies with the attribute value or class of values. Finally, choropleth maps represent statistical data for predetermined geographical

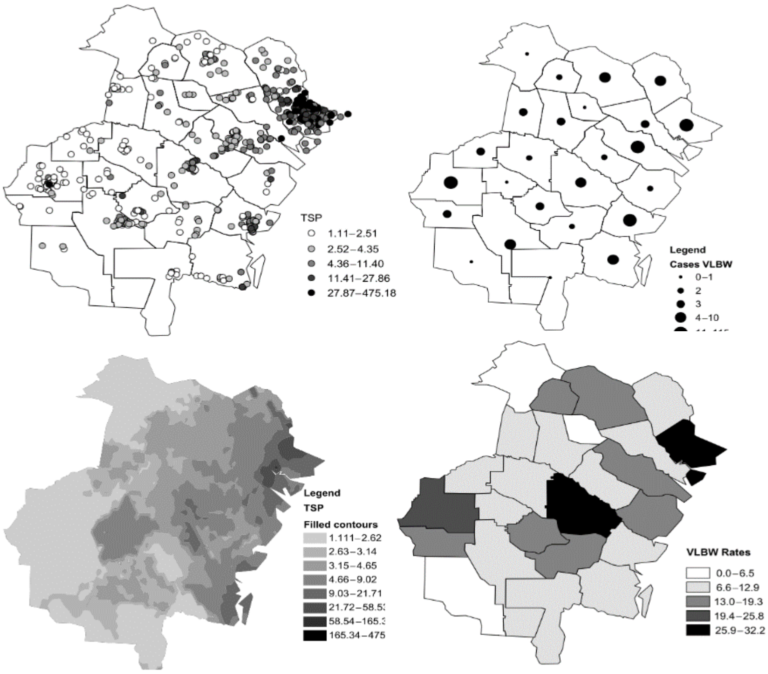


Figure 11. Point Map (top left), Bubble Map (top right), Dasymetric map (bottom left), Choropleth Map (bottom right) [58].

units through different coloring, shading or hatching, whereas dasymetric maps depict quantitative data areal data using boundaries that divide the mapped area into zones of relative homogeneity [56], [57], [58].

Usually, these types of maps are represented by a variable indicating the magnitude of a quantitative value attached to longitude and latitude data coordinates to position marks on the map [56], [57].

With the emergence of new types of users like Data Scientists, Geographic Information Systems is starting to evolve more outside of its traditional realm. This is allowing more sophisticated spatial analysis to take place in the context of new Data Science & big data solutions. SDS, earlier mentioned, is a field of Data Science that investigates how and why the events happen in space, emerging as a discipline that draws on the unique characteristics of spatial data [59].

3.1.6 Dashboard as a combination of multiple data visualizations

Each data visualization has a specific objective to fulfil, and therefore it is essential to select the datasets to be explored, and afterwards, determine the most suitable graphic representation [49]. However, for multi-objective decision-making, there is no single data visualization mentioned before, that comprises all requirements for data analysis and design decision support. Consequently, the analysis of output of multiple data visualization requires navigating between different platforms, which is a time-consuming and inefficient process. Therefore, a single tool, that provides the capability to compare and analyse different visual formats at the same time, could be essential to establish relations between objectives and variables, and consequently, to improve efficiency of decision-making [60].

In order to satisfy this necessity, a new data visualization tool – dashboard, was created. A dashboard provides a rich user interface that exhibits the data in a graphical form using a range of visualization formats already mentioned, e.g., charts, tables and maps [61].

3.2 Dashboards

3.2.1 Definition

Dashboards are one of the most common-use cases for data visualization tools, delivering information in a particular format to the decision-maker. The interest in dashboards has increased recently, due to their ability to enhance decision-making by taking advantage on human perceptual capabilities, and solving the issues of presentation format and data load [7].

The definition of a dashboard suffered various changes through the history, depending on the specific role and its objective. In his article for Visual Business Intelligence, Few [62] states: “A dashboard is a predominantly visual information display that people use to rapidly monitor current conditions that require a timely response to fulfil a specific role”. The big Book of Dashboards [63] expands this definition, describing a dashboards as a display which can incorporate infographic elements in order to facilitate understanding.

In the light of latest developments, a more precise definition of a regular dashboard might be: “visual and interactive tool that displays on a single screen the most important information to achieve one or several objectives, allowing the user to identify, explore, and communicate problem areas that need corrective action”[7].

However, big data and new visualization technologies have expanded the dashboard concept to new domains. This way, the dashboard idea has transformed from single-view report screen to interactive interfaces with numerous views, adding learning, communication, and motivation to traditional purposes such as monitoring and decision support [64]. Using new sophisticated software, it is possible to apply abstraction or zoom into details, allowing user to focus on specific elements of interest.

At the same time, modern dashboards allow to incorporate predictive analytics, implement a pattern recognition and other machine learning models, to improve the perception of data and elevate it to the next level [65].

3.2.2 Purpose and domains

As has been already mentioned, dashboards were developed as a response to the increasing complexity and diversity of data. A weak organization of potentially decision-relevant data that leads to inefficient resource allocation, increase of biases in information processing and lack of performance reporting practices are some of the factors that drive the utility of dashboards [66].

Consequently, there can be several purposes of a dashboard, in order to reenforce decision-making in various aspects. Firstly, a dashboard allows consistency across departments and units, synchronizing different metrics. Secondly, a dashboard helps to evaluate and correct data constantly in real time, i.e., allows data monitoring. At the same time, the simulation of future scenarios can be incorporated in dashboard, allowing the planning of possible strategies. Lastly, a dashboard may be used to communicate values of an organization to important stakeholders [7].

Dashboard are used across different domains due to their beneficial contribution for organizations. The business companies present the major proportion of dashboard’s users, especially within business intelligence (BI) area. BI dashboard's objectives are diverse and can incorporate multiple analytical tasks, including the enhancement of operational efficiency, the cost reduction and the strategy planning [7], [67], [68].

Healthcare organizations have been adopting dashboards at both hospital and patient-care levels, improving quality of care by enhancing decision-making across various units [32], [69]. (Section 3.2.8) Moreover, urban informatics and smart cities movement have been implementing dashboards in order to monitor and control the cities' overall health and performance through real-time streamed data [70], [71], [72]. The implementation of dashboards as personal behavior tracking tool is gaining popularity, with purpose to support reflective learning for daily activities [73].

3.2.3 Key indicators

Once the purpose of dashboard is defined, the next step is to determine the information that should be included on the dashboard. Key indicators are considered one of the vital components of dashboard, since they provide the capability not only to assess data needed to focus attention on, but also to acquire information beyond the direct relationship of the parameter and its value. Indicator, by definition, is "a measurement or a set of measures to evaluate a complex social, economic, or physical context" [74]. It reflects parameter or value that gives information about certain phenomenon. At the same time, indicators can evaluate performance of policies in terms of effectiveness by measuring changes in indicator values over time. When the decision-making requires the study of certain geographic area, e.g., in order to allocate resources or develop policies for certain regions, the values of key indicator can be associated with the spatial key, allowing the posterior mapping of this values on the dashboard [72].

Well-defined key indicators reflect the most essential metrics, that are crucial for efficient and consistent decision-making. Usually, their number is limited in order to eliminate unnecessary data and concentrate only on dashboard purpose [74], [75].

3.2.4 Development Stages

The development of a dashboard is a complex process, which involves a range of decisions regarding technical and design aspects. The main steps vary depending on the particular type and purpose of dashboard, however, there are 5 core considerations:

A. **Study setting**

Study setting describes the physical, social or experimental context within which dashboard development will be conducted, i.e., defines interactions, data requirements, privacy and confidentiality issues [76]. This step also includes the identification of organizational aims, conditions and readiness of the organization to proceed with dashboard development [77].

B. **Dashboard purpose and concept**

This step includes the definition of actual purpose of dashboard and identification of the stakeholders, target audience and their expectations regarding the use of a dashboard. [76] Specifically, the needs and daily activities of stakeholders are identified, in order to have a more profound comprehension of how dashboard will support stakeholders' activities [77].

C. **User Interaction and flow stage**

At this stage, the dashboard developer needs to identify the actions expected from the user interaction with the dashboard interface. The visual design guidelines i.e., overall sequence of entry points of user's exploration must be identified. Using Visual Information Seeking Mantra, developed by Shneiderman [78], the basic order of user interaction steps is resumed by "Overview first, zoom and filter, then details-on-demand". The concept is remarkably simple: the overview component displays the entire dashboard with all features, zoom is responsible for zoom-in on elements of interest, filter removes unrelatable items and details-on demand enables to browse the details about the group or individual items. Therefore, the objective of this step is to map the particular user action to corresponding component of the Mantra, taking in account the dashboard goals.

D. **Data selection and visualization design**

As the title says, this step involves the design and construction of the dashboard, including the identification of key indicators and corresponding data visualization formats, the posterior extraction of selected key indicators from database and their implementation within dashboard design [76].

The process of selecting the key indicators is usually an iterative one carried out with consultation of interested stakeholders. Distinct selection methods are used as a tool to select and evaluate the key indicators from experts' perspective, and ensure their relevancy, scientific rigor and reliability in terms of the objectives in question [79].

A dashboard design should be cautiously considered while developing a dashboard. If user interface is not able to display the data in an efficient way, then the dashboard is pointless as a decision-making support tool. Consequently, it is important to analyze and select the most relevant visualizations to optimize the delivery of necessary information, taking into account not only the designer opinion, but also the stakeholders' considerations [2], [80].

This development stage is explored in greater detail in Section 3.2.5.

E. **Framework architecture**

This stage includes technological features, such as architecture identification and appropriate platform selection [76].

The architecture most adopted by dashboards is a three-tier one [72]. It is a conventional software architecture that separates dashboard into three logical and physical computing tiers: data, logic and presentation. The data tier is accountable for data storage and management. It also includes data models, responsible for mapping the data to indicator measurements. The second, logic tier, connects the presentation and data tiers to make communication faster and more efficient, and can also incorporate different analysis models for the indicators. Finally, the presentation tier represents the graphic user interface, where end-user interacts with the dashboard [72], [81], [82].

Dashboard platform is a computerized tool that integrates the mentioned architecture and serves specifically to develop and deploy dashboards. This platform must offer interfaces and common functionalities and be highly customizable and expandable to handle complex requirements. A presentation tier of such platform must present all data visualizations required by stakeholders, such as charts, graphs and maps. The platform must also facilitate the connection of dashboard data tier to third-party data sources in order to load the native databases with a needed data [83].

F. Maintenance

Lastly, since dashboard is a dynamic tool that requires constant update of data, it is essential to ensure its maintenance by monitoring metrics, analyzing feedback, and performing necessary improvements to increase its efficiency [77].

3.2.5 Data selection and visualization design

As briefly explained before, data selection and visualization design integrate the process of key indicators' selection for a dashboard and their reporting through the use of data visualizations. This process is generally an iterative one conducted with selected stakeholders and incorporates distinct stages [76].

According to Brown [79], the guidelines regarding the development of key indicators and their integration within dashboard are organized in five stages:

A. Establishing the purpose of the indicators

This step is highly interconnected with the purpose of dashboard and includes the definition of key stakeholders and target audience, and certainly, the identification of the purpose of key indicators [79].

B. Designing the methodology

The methodology is an important tool that incorporates a set of different methods for key indicators' development to guarantee the clear guidance for their selection [79], [84].

C. Selecting and designing the indicators

At this stage, the implementation of previously developed framework is realized, and each selected indicator is evaluated by applying the set of criteria to ensure their relevancy and scientific rigor.

The following criteria, proposed by Schomaker [85] are the most commonly identified, representing the indicator scientific quality. He suggests that indicators should be SMART, i.e.:

- Specific, i.e., be clearly and unambiguously specified.
- Measurable qualitatively or quantitatively.
- Achievable in terms of the available resources.
- Relevant for the issue in question.
- Time-bound, i.e., be responsive to changes within policy timeframes.

D. Interpreting and reporting the indicators

The main purpose of this stage is to create connection between measurement and comprehension by using appropriate data visualization formats [79]. To achieve this goal, the methods concerning selection of these formats need to be previously added to initial methodology process, or the separated methodology must be developed [86]. The presentation of the indicators must be carefully chosen, in order to be transparent and objective [4].

E. Maintaining and reviewing the indicators

The selected indicators need to be open to discussion and modification, and the addition of new indicators must be considered. This process should be continuous involving the stakeholders, expert groups and target audience [79].

3.2.6 Target audience

The interpersonal circulation of a dashboard depends on target audience, previously defined on development process. This audience can be divided into four groups according to the specificity of the context: public, organizational, social and individual.

Public dashboards are focused on government decision-making and accountability to monitor certain metrics, to measure outcomes of implemented policies or to transmit the relevant information to the general population. They have different goals in comparison with the profit-oriented private ones, and their adoption by public sector is a complex task. A concept of public value (e.g., equal access, rights, etc.) is difficult to measure, so there is a large disaggregation of public organizations in terms of governmental regulations and laws. At the same time, the aspects such as the data quality, consistency, and anonymization require high-level control by government [2].

On the contrary, dashboards for organizations are intended to integrate internal processes of organization, in order to improve their management and successfully achieve strategic goals. In majority of cases, data sources are private and protected, enabling the confidentiality of information. Moreover, it allows more efficient control of the data accuracy and veracity [87].

Lastly, social dashboards have a limited access to certain group of individuals, meanwhile the individual dashboard captures personal data of a certain individual [64].

3.2.7 Challenges

The development of dashboard and its posterior utilization present various challenges, at both functional and design levels. First of all, a purpose of dashboard is difficult to identify due to the diversity of consumers. Automatic adaptation of dashboard for different users is hard to achieve, and still is an open research challenge [64], [88].

Another difficult challenge of dashboard development is related with data collection. Since dashboards are generally developed for multiple stakeholders with different needs and activities, the analysis is needed to identify and select the key indicators required by each stakeholder. The complexity of key indicators' selection and identification of viable data sources prejudices the quality of dashboard, and consequently, the decision-making. Usually, selected key indicators are highly subjective, and so their posterior filtering and ponderation, therefore it is a complex task to eliminate ambiguity and optimize the selection methods [66], [89]. At the same time, data availability, quality and security are crucial for effective functioning of dashboard but are challenging to achieve [7].

Design is another challenge of dashboard developer since the representation of selected data and key indicators plays a great role in its analysis. The main purpose of efficient design is to organize a high volume of data into a single screen, in the way it is meaningful and useful for the user. Therefore, the maximization of visual perception without losing functionality and simplicity is a primary goal of dashboard, which is, however, difficult to achieve [2], [7]. As in the case of data collection, the design process also requires different stakeholders, and consequently, the research is needed to identify the data visualizations expected by each stakeholder. Failure in this complex task can be devastating for dashboard design since the outcome can be inconsistent with stakeholder expectations [77], [90].

3.2.8 Dashboards in Healthcare

The adoption and use of decision tools in healthcare domain allow healthcare professionals to examine and determine trends in data and assist them in better identification of anomalies and their interpretation. Moreover, the cognitive studies regarding the improvement of information perception through visualization, triggered the growing interest in the use of interactive visualization tools in healthcare, particularly to support decision making [6], [91], [92].

A clinical dashboard is designed to “provide clinicians with the relevant and timely information they need to inform daily decisions that improve the quality of patient care. It enables easy access to multiple sources of data being captured locally, in a visual, concise and usable format” [93]. Therefore, dashboards in healthcare provide the access and monitoring of real-time health data, the identification of data patterns for better resource allocation, the improvement of communication, guidelines’ adherence and transparency in healthcare organizations, and consequently, the enhancement of performance and efficiency of healthcare services [92]. This visual tool has been also used by patients, in order to improve patients’ safety, satisfaction and communication with health professionals [94].

In healthcare fields, there are numerous examples of application of dashboards to real-life problems. This data visualization tool has been already used for allocation of resources and cost management in operation rooms[95], in emergency units to reduce patient’s hospitalization time [96], in surgical intensive care unit to reduce pneumonia by improving ventilator bundle efficiency [97], and in radiology departments to improve X-rays dosage [98]. Additionally, a dashboard provides an opportunity to analyze key relationships between the health characteristics of population and both socio-economic and environmental features, contributing to spatial epidemiology [99]. Last is defined as being “concerned with describing, quantifying, and explaining geographical variations in disease, especially with respect to variations in environmental exposures” [100]. This way, synchronous spatial visualization of health data and both socio-economic and environmental indicators facilitates the comprehension of distribution of various diseases [101].

Nowadays, the healthcare organizations are forced to communicate their efforts, initiatives and activities not only to stakeholders, but also to the public though social media. Dashboards also provide the solution as an efficient tool to transmit the needed information to a generic population in a simple and transparent way [102].

3.3 Participatory methods in dashboard design

As has been already mentioned, the participation of various stakeholders in dashboard design is a key requirement so that it assists the needs of various users with distinct objective in an effective way. To optimize the integration of key users into development process, participatory methods started being developed to improve the usability of interactive systems by collecting and analyzing a direct input from users [103], [104]. The integration of users during the process ensures the acceptability of the developed solution, since participatory methods combine the information from a diversity of sources more efficiently than quantitative or qualitative methods alone [105].

3.3.1 Overview of participatory methods

A selection of participatory methods is a complex process, where one or more methods can be selected and combined. The next part of this section explores the most common participatory methods that can be used in dashboard development, in which surveys and interviews do not require communication between different stakeholders, meanwhile brainstorming, workshops and Delphi involve certain interaction among participants.

A. Survey

Survey is a participatory method of obtaining information from a sample of individuals of interest. It is a particular type of research that “involves the collection of data from a sample of elements drawn from a well-defined population through the use of a questionnaire” [106]. This process can integrate a range of distinct designs, where data collection can be performed at a single point in time from a specified population (cross-sectional surveys) or at two or more points in time from the same people (panel surveys) [106]. The questionnaire includes standardized questions with an objective to present a homogeneous stimulus to individuals of interest so that their answers are comparable [107].

The surveys can be performed in different ways, including telephone, mail or in person. All the survey’s results should be provided in entirely anonymous summaries, such as statistical tables and charts [108].

B. Interviews

The interview, by definition, is “face-to-face verbal exchanges in which one person, the interviewer, attempts to acquire information from and gain an understanding of another person, the interviewee” [109].

Interviews are often classified according to their structure. The structured interviews present certain similarities to questionnaires, except that the interviewee asks the question instead of leaving interviewee to complete and return the questionnaire. The unstructured interviews are based on a restricted number of topics, with an objective to stimulate the interviewee to talk around a subject. The most common type, semi-structured interview, focuses on 6-12 appropriate and well-formulated questions to be provided in a certain order, but accepting some flexibility [109], [110].

The main advantage of using interviews is that they need less prior knowledge than for a well-build survey, and data collection is much easier [109].

C. Brainstorming

Brainstorming is a broadly used participatory technique, based on exploration of new meanings, proposals and idea within small group of experts and relevant stakeholders [111]. This method was firstly introduced by Alex Osborn in 1938, and experienced different changes through history. However, the classic process follows a number of basic rules, named Osborn’s rules, which instruct the group members to be sustained by previous ideas, to not criticize any ideas, to produce a significant quantity of ideas and do not hesitate to contribute original ideas [112].

The panel format includes leader, which maintains a rapid flow of ideas, recorder - to list the ideas as they are presented, and variable number of panel members. It is a simple process, which doesn't require special expertise or knowledge required from the leader, however, due to non-anonymity of members, the credit for another person's ideas may inhibit participation [113].

D. Workshops

Workshop is a participatory method that emphasizes the group learning by obtaining new knowledge, executing creative problem-solving, or innovating regarding specific topic. The main purpose of workshop is to generate reliable and valid data about the issue in question [114].

Usually, a workshop consists of four phases:

1. Preparation phase: the workshop purpose, its rules and schedule are introduced.
2. Critique phase: actual beginning of the workshop, where the issue in question is investigated by stakeholders.
3. Fantasy phase: the creation of new original ideas is performed.
4. Implementation phase: the proposed ideas are analyzed and validated, and the action strategy is developed in case of solution's discovery [115].

This method resembles brainstorming; however, it is more organized and has a certain structure.

E. Delphi

The Delphi is a structured communication process which uses a number of questionnaires or rounds to collect and deliver information with the objective to achieve a group consensus [116].

There are several variations of the Delphi technique, however, all follow a predetermined set of both behavioral and statistical procedures. Generally, three rounds of questionnaires are presented to pre-selected participants, however, the number of rounds can vary depending on other factors, e.g., availability of participants, costs and duration restrictions [117].

The classical Delphi process usually begins with unstructured open-ended questionnaire, which stimulates an idea generation to identify the issues to be considered in future rounds. Afterwards, the qualitative and quantitative analysis of the answers is performed, and feedback is provided. In subsequent rounds, the participants are asked to review and (if they want) change their responses in the light of the results and feedback from previous round. This process is repeated until consensus is achieved [116], [117]. The consensus is highly subjective, so its definition needs to be thoroughly specified, e.g. by defining percentage level of agreement or by measuring the degree of dissent and divergence amongst experts' opinions [117].

However, Delphi method has undergone various modifications in use to improve particular concerns, some of them enumerated in Annex A.

The growing popularity of Delphi method can be explained by few basic characteristics that advantageously differentiate it from traditional participatory methods, which are: anonymity, interactivity, reflexibility and flexibility [118], [119]. Firstly, the anonymity of participants is preserved throughout the process, encouraging experts to communicate independently from personal conflicts or status, and this way, decreasing the possibility of negatively influence outcomes from group interaction. On the other hand, the Delphi is an interactive, flexible and reflexive process since its structure can be adapted to the research context. Moreover, the expert panel can reflect through the problem between the rounds, review, change or leave feedback on their responses, thereby enhancing the data validity and credibility, and enabling the collection of a rich and varied data set. Lastly, the flexibility of process is improved by its independence from participants' proximity, expanding geographical barriers and reducing travel costs [118], [119]. This flexibility can be particularly boosted by using web-based platforms, such as online survey tools, that bring numerous advantages for the Delphi processes, such as a wider reach of expert participants contributing to the process, a shorter timeframe to study completion, cheaper study costs and possibility of embedding of multimedia in the questionnaire [120].

However, same characteristics can lead to disadvantageous outcome. For example, the access to other participants' responses can increase the experts' biases and tendency to choose the most voted option without solid argumentation. At the same time, anonymity can enhance the lack of participants' responsibility regarding their answers. About interactivity, it can provoke exhaustion and demotivation [119].

3.3.2 Participatory methods for data visualization selection

The major objective of data visualization tools is to offer a clear view of selected indicators; nevertheless, selecting the appropriate visualization methods is not easy at all. The successful visualization, as previously said, must be perceptible and convey the key properties of the indicator. Otherwise, the chosen visual formats may be deceiving and lead to incorrect conclusions. Therefore, it is necessary to obtain guidance to select data visualizations for certain indicators [4], [6].

Despite the growing popularity of data visualization tools, little is known about the methodologies to support their design. However, the methods described in Section 3.3.1 can be used to identify visual formats for indicator's representation.

The theory-based approach such as literature review, is usually used as a starting point of research, in order to learn existing types of visual representations [4]. For example, the literature review has been performed to develop a set of features of visual design and related elements (tables and charts), associated with key performance indicators for medical imaging department [121]. However, the design of data visualization tool is considered to be highly related with learning theories of cognition, where the choice of interactive components such as charts and maps, depends totally on end-users' perception and attention [122].

The next step was to recognize which participatory techniques have been used to select data visualizations and particular situations where they are applicable and useful. During May 2021, databases as PubMed, ScienceDirect, Web of Science and Google Scholar were used to find the best articles and understand the applicability of participatory methods for data visualization selection. Keywords like "visualization", "dashboard"," data visualization", "data visualization tools", in combination with each of the participatory method supposed to evaluate, such as "survey", "interview"," brainstorming", "workshop", "Delphi", were used to find the articles.

There were no time limitations set, but it was clear that this was a current topic of research. As indicated in Figure 12, publications were evaluated one by one, titles and abstracts were reviewed to identify relevant research, and the others were eliminated. The citations in these publications were also evaluated to see if they were valuable or not.

The remaining articles were then thoroughly examined, and the ones that summarized information already obtained in previous ones as well as the ones in which data visualization was not the main focus of the study, were discarded. After that, a set of 18 articles relative to the application of the participatory methods for data visualization selection were selected for their posterior analysis (Annex B).

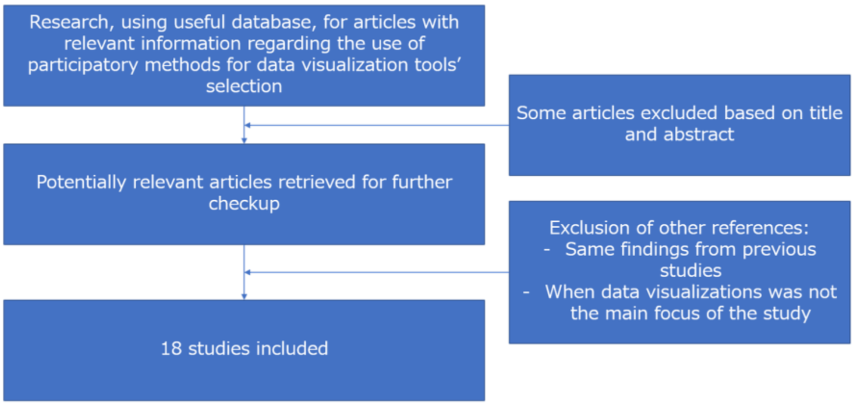


Figure 12. Flow diagram.

Despite the highlighted importance of stakeholder's opinion for dashboard design, the number of studies that integrate participatory methods to explore visualization of dashboard is very limited. Most methods were focused on determining the needs, requirements and preferences of stakeholders for dashboard design and functionalities rather than determining the most adequate visualization for certain indicator. The participatory methods mostly explored the feedback of stakeholders through direct questions about the certain design details, such as element's position in dashboard framework and their interactivity. In other cases, the participatory methods have been used to evaluate and validate the pre-developed dashboard, reducing the stakeholders´ level of involvement in dashboard design.

4 Proposed approach

This chapter is structured around a development of a novel approach that aims to assist the selection of the most adequate data visualization formats for dashboard's indicators. It provides the rationale behind the proposed approach to guide the selection process; it includes an identification of the most appropriate data visualization formats, based on literature review, combined with participatory methods to integrate stakeholders in dashboard design. It is noteworthy to mention that the proposed approach integrates most common data visualization formats that are easy to implement in any dashboard platform. Therefore, it is highly applicable to inform the construction of public health dashboards in order to transmit the important information to the general population in the simplest and most perceptual way.

4.1 Designing a novel approach

4.1.1 Objectives

The proposed approach intends to tackle the previously identified challenges in the ongoing process to improve efficiency, optimize the selection process and validate the outcome with tangible evidence. Therefore, it is relevant to develop an approach which assists the selection process by integrating the knowledge acquired from literature review and combining it with participatory methods in order to access perceptual capabilities. Since the developed approach integrates the most common data visualization formats, it would be most appropriate for public dashboards, e.g., for public health dashboards, due to the simplicity of information and their transmission. More precisely, the main objectives will be:

a) Implementing a step-by-step process to aid the selection process by establishing the criteria for indicator's classification and identifying the pre-set of most adequate data visualization formats for indicator in question.
b) Implementing participatory methods to enable exchange of perspectives and values of participants through selection of the preferred data visualization format from available options (it can be applied with distinct groups of participants).

4.1.2 Designed selection process

Choosing the right data visualization format is crucial to identify the relevant information or patterns from a given dataset, however, this task is far from trivial. A given data set can be interpreted by alternative data visualization formats, therefore the most adequate visualization should be selected to allow easy comprehension of graphical illustration, and to avoid either the loss of important information or the inclusion of unnecessary data [4], [123].

To develop this approach, an assessment of the current situation was already made and, subsequently, the objectives for this approach were established in a previous section 4.1.1. Afterwards, the design of the selection process was performed, considering the challenges of previous ones. A flowchart representing the outline of the proposed approach is presented below (Figure 13). It is important to note that the indicators' selection stage must precede the proceedings regarding the identification of visual tools.

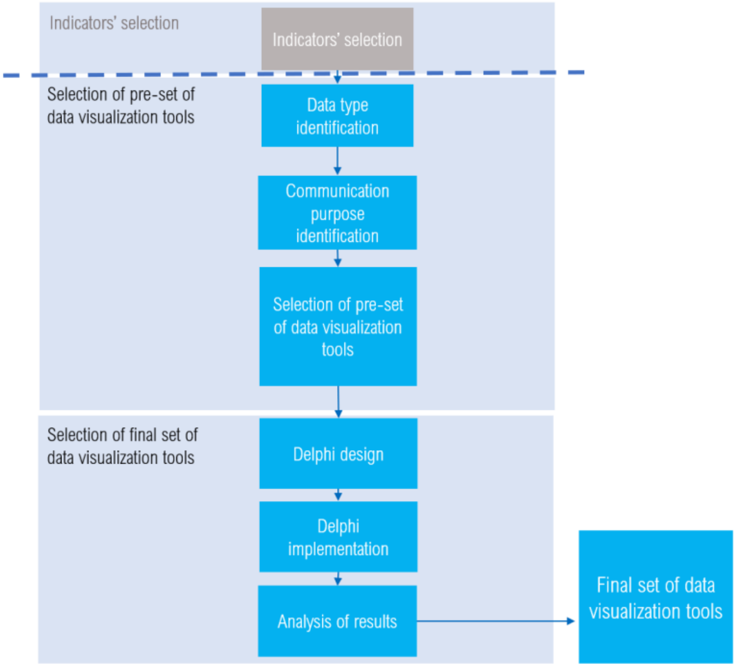


Figure 13. Block diagram representing the steps made for the implementation of the proposed approach.

4.1.3 Selection of a pre-set of data visualization formats

Despite its broad diffusion in actual software applications, theoretic reflections on visualizations are rarely done in Information Systems science. As a result, there is ongoing research concerning how to define visualizations apart from their physical graphical appearance, yet in a way that is detailed enough to capture the characteristics that distinguish each one in visually communicating information [2]. Explaining these core features would allow the creation of an approach, where different types of visualizations can be arranged with the objective to make them comparable [51].

It is clear that there is no "best" design solution, however, there are numerous theories and practical evidence that help to understand which visualization techniques work well for certain situations and less so for others. Therefore, the objective of this step is not to choose the best solution, but to select the data visualization formats that are most appropriate, considering the pretended outcome and indicator's structure.

This stage is based on previously developed literature review and integrates data visualization process from Dastani [20] with Kirk's approach [9], where the first step covers the identification of the data structure. This step in the process of data visualization transforms the input data to what is called structured data. For that, each indicator is structured into $m \times n$ data table consisting of m columns and n rows. A column A_i ($i = 1,\ldots,m$) consists of values of one data variable and a row B_j ($i = 1,\ldots,n$) consists of m values each belong to one of variables $A_1,\ldots,A_m$.

Afterwards, an attribute-based classification of identified variables is performed. The structural data attributes can comprise data values that can be classified as quantitative (Q) or categorical (C). Alternatively stated, the Q variables contain numerical values, whereas C variables have individual values (e.g., geographic regions, names, or products). For further simplification, one can use the designation used by Helfman [24], where the variable's types can be represented by a string, such as

“CQQ” or “CC”, where the length corresponds to the number of variables, and the letter – to the variable’s type.

However, Kirk [9] highlights the necessity to “telling stories” with data visualizations, and determines different methods to appropriately categorize them according to their primary communication purpose, focusing on variety of possible outcomes. Therefore, the identification of the purpose that the data visualization format pretends to transmit was considered to be important, in order to reduce the range of suitable chart types within each method family by the nature of the variables in question. The identification of the purpose requires the in-depth visual analysis, that integrates an understanding of the key communication dimensions of visualization problem.

The second step in data visualization process is to determine visual elements that represent data in such a way that the perceptual structure of the decided visual elements represents the structure of the data. Therefore, the decision table (Table 2) was developed, using as the base the previous literature research, to enumerate data visualization formats and the correspondent variable’s types they can compare according to their communication purpose. This table is elaborated based on Kirk’s categorization of data visualization formats according to communication purpose, combined with the knowledge about quantity and type of typical data variables normally used with these visualizations [124]. This decision table allows to predefine a set of most appropriate data visualization format for certain indicator, however, it doesn´t provide a unique answer and requires the validation from stakeholders. Since suggested process offers a set of alternative data visualizations for a certain dataset rather than identify the most adequate visualization, it is crucial to restrain the options by applying participatory methods. This way, the stakeholder’s value judgments are considered to weigh in on each alternative to modulate the results of this step.

Table 2. Decision table of data visualization format’s selection.

Variables’ type / **Communication purpose**	**CQ**	**QQ**	**QQQ**	**QQC**	**CCQ**
Comparing categorical values	Bar (Column) Chart	-	-	-	Two-sided Bar Chart, Grouped Bar (Column) Chart
Assessing hierarchies and part-to-whole relationships	Pie Chart, Donut Chart, (100%) Stacked Bar Chart	-	-	-	(100%) Stacked Bar Chart
Showing changes over time	Column Chart	Line Chart, Area Chart	-	Line Chart, Area Chart	Grouped Column Chart, (100%) Stacked Column Chart
Plotting connections and relationships	-	Scatter Plot	Bubble Plot	-	-
Mapping geo-spatial data	-	-	Point map, Choropleth map, Bubble plot map, Dasymetric map	-	-

4.1.4 Participatory process

As has been already mentioned, the participation of stakeholders in dashboard design is a key requirement to assist the needs of various users with distinct objective in an effective way. To optimize the integration of key users into development process, this step implements the participatory process. Because participatory methods incorporate information from a variety of sources in more efficient way, the inclusion of stakeholders assures the usefulness and acceptance of the created solution.

a) Selection of appropriate participatory method

As has been already mentioned, the heterogeneity of stakeholders is a pertinent issue in various domains, including healthcare. Therefore, the selected participatory method for dashboard design not only needs to take in account stakeholders' divergent interests, but also reach an agreement concerning different decisions, in a legitimate and dynamic way.

To evaluate the appropriateness of each participatory method for the issue in question, the set of criteria was developed by Wittmer [125]. Firstly, the information management of the participatory method must be clearly defined since the knowledge comes from different scientific disciplines and needs to be integrated. Therefore, the participatory method should elucidate different types of information in a transparent and organized way [125]. On another hand, the legitimacy of decision must be guaranteed, by including all the relevant interests and affected stakeholders and ensuring their equitable participation in the process [125].

Despite the existence of participatory methods that does not require the communication between participants, the group participatory methods present advantage not only when the outcomes impact a group of people, but also when the decision itself can be improved by including more people in the process [126]. Therefore, social dynamics of participatory method can contribute to idea generation and collective learning. At the same time, understanding the position of others as well as developing and/or changing one's own position improve the final outcome of the process [125].

Finally, the costs of participation should be considered to address cost-effectivity of the participatory method, and a panel size needs to be analyzed in order to understand the number of experts that can be involved in the process [125].

After defining the set of criteria, the evaluation of the participatory methods (Table 3) was performed, considering selected criteria and previous literature research, performed in Section 3.3.1. The evaluation ranges from "0", i.e., "the method does not address explicitly this issue", to "+++", i.e., "the method address this issue in systematical and explicit form".

Table 3. Overview of analyzed participatory methods.

	Interview	Survey	Brainstorming	Workshop	Delphi
Information management					
Integration of different types of information from stakeholders	+	+	++	++	++
Legitimacy					
Inclusion/representation	+	+	++	++	+++
Equitable participation in the process	+	+++	+	+	+++
Social dynamics					
Changing behavior, changing perspectives/learning	0	0	++	++	+++
Facilitating convergence	0	0	+	+	+++
Panel size					
Size of expert's panel[1]	++	+++	+	+	+++
Costs					
Costs of the approach[2]	++	+	++	++	+

The results of an evaluation of different participatory methods, presented in Table 3, demonstrate that Delphi is considered highly appropriate for the dashboard design, where the heterogeneity of stakeholders implies high number of experts, and high integration and inclusion of different types of information. Last one is vastly facilitated by anonymity, which enables the free expression of opinions and equitable participation in the process. Moreover, Delphi facilitates convergence between participants by constant learning during the process and flexibility to change certain behavior by analyzing the other stakeholders' opinions. It is also noteworthy to mention that Delphi becomes the method of choice where there is little previous research, which is the case of this study [127].

b) Delphi process and data visualization formats' selection

Lately, the growing popularity of dashboards stimulated the implementation of Delphi process during their development, in order to capture heterogeneous stakeholders' opinions. It have been used not only to select relevant indicators, but also started to be implemented in visualization formats' selection for dashboards, since the cognitive and perceptual capabilities of stakeholders has been confirmed to be a main key to efficient visualization of data [122].

During the literature review regarding the use of Delphi for data visualizations (Annex B), three academic articles were identified and analyzed. Al-Hajj et al. [28] focused on development of Analytical Injury Dashboard to inform policies and to strengthen injury surveillance, prevention and future research. Particularly, this article utilized the previous study, that used Delphi approach to develop injury indicators, and focused on visualization analytics, in order to facilitate data exploration and knowledge construction. Firstly, the types of visualizations were selected to efficiently illustrate trends and patterns

[1] "0" means low number of experts, and "+++": very high number of experts.
[2] "0" means low costs, and "+++" very high costs.

in injury data by using literature review. Afterwards, a dashboard was developed by Visual Analytics Expert (VAE) and was tested by applying Delphi techniques incorporated in Group Analytics approach, where injury stakeholders interacted with VAE and provided the feedback about integrated data visualization visualizations by filling-out the questionnaire. Finally, the dashboard visualizations were manipulated by VAE to customize the views according to the stakeholders' needs and preferences. However, this study incorporated a number of limitations, since the injury stakeholders were not making actual decisions, but validating and customizing already selected visualization formats.

The second study explored data visualization tools for Cybersecurity Data Visualization Dashboards by using tailored Delphi technique [128]. Firstly, the literature review was done to identify visualization challenges used in cybersecurity. Thereafter, the first round of Delphi was performed using interviews with open-ended questions, where participants have been asked regarding visualization challenges that they consider important, and their desirable features in the security visualization tools. Second round was more restricted to gather opinions and feedback on specific items, discussed in the first round. However, the study paid more attention to challenges, e.g., data quality and integration, rather than analyze the visual features.

In the last article, "Developing an emergency department crowding dashboard: A design science approach" [129], Niels et al. described a development of a dashboard, using modified two-round Delphi approach to specify not only the indicators, but also the requirements regarding their visualization. In the first round, a questionnaire was distributed to the experts, and afterwards, the responses were summarized and presented to participants in the second round. Therefore, this study didn´t focus specifically on the design issues but combined the indicators' selection with visualization requirements in a single step. Moreover, the data visualization formats were not clearly identified, since their selection was made by developer, considering the requirements defined by stakeholders in Delphi process.

In summary, these studies presented various limitations, and consequently, the possibility for future improvement regarding the use of Delphi in data visualization selection for dashboard. Namely, the data visualizations were not clearly identified, focusing more on generic preferences and needs of stakeholders rather than to validate the set of proposed data visualization formats for certain indicator. Therefore, the approach proposed in this thesis should take into consideration these studies' limitations and suggest the efficient Delphi for visualizations' selection to support stakeholder's problem-solving and decision-making processes.

To surpass the presented difficulties, the Delphi questionnaire should include well-defined questions, investigating the stakeholder's perception of visualizations and their preferences. For that, the set of candidate visualization formats for each indicator, provided by previous step, must be designed and indicated in questionnaire as possible options to answer. Particularly, each question should ask to choose the most preferred visualization between the options presented for the selected indicator. Finally, the Delphi process should be constructed considering various issues in question such as the objectives of dashboard in question, the number of selected stakeholders and their availability. The consequent analysis of results should investigate the stakeholder´s feedback not only to select the most appropriate data visualization for dashboard, but also to explore different perceptions that can be used for future research.

5 Case study design

The development of the present thesis has as basis a real-world case study and aims at optimizing the selection of appropriate data visualization formats for presenting information commonly displayed in public web-based dashboard of evolution of COVID-19 pandemic. Thereafter, the present chapter has the objective of describing the implementation of novel approach, based on a Delphi process, in order to understand the preferences and views of the general population regarding distinct data visualization formats. First, the indicators mostly used in COVID-19 are established, and finally, the implementation of the developed approach is described. Specifically, the selection of pre-set of data visualization formats is performed, taking in account the theory-based evidence regarding the data to be considered. Finally, the Delphi process is described and implemented.

5.1 Indicators' selection

As has been already mentioned in Chapter 4, this step antecedes the proceedings regarding the data visualization format's selection. The appropriate and well-designed indicators are vital for dashboard effectiveness and performance, so their selection is highly complex process which requires specific attention.

Usually, the indicators are selected for and not with its users, and consequently, are not consistent with user's requirements. The usability and relevance of these indicators are questionable, since the disaggregation between developers, experts and users influences the quality of chosen metrics. Therefore, it is essential to involve stakeholders throughout the entire indicator's design process [130].

The common practice to extract candidate indicators is a literature review, where the search engines are used to locate and retrieve relevant indicators from web-located documents with simple keyword-based searches [131], [132]. However, the literature review does not incorporate the values and interests of stakeholders, crucial for indicators' acceptance and legitimacy. Therefore, the outcomes of literature reviews require the validation by participatory methods in order to evaluate their relevancy for the subject, since group opinion is more reliable than an individual belief. Through these methods, one guarantees that developed indicators are adequate, functional, and meet the needs of end-users, and a final dashboard is more usable as a result of direct input from stakeholders [133].

Since the focus of this study is not related with indicator's selection, the simpler approach was applied to define the set of indicators commonly used for web-based public COVID-19 dashboards. For this, the study entitled "Features Constituting Actionable COVID-19 Dashboards: Descriptive Assessment and Expert Appraisal of 158 Public Web-Based COVID-19 Dashboards" was analyzed, which explores characteristics of 158 public web-based COVID-19 dashboards from 53 countries by assessing their features, namely the key performance indicators and their frequency [36]. For this study, the presented indicator's themes will be filtered by the frequency, considering only high and medium frequencies (>34%) (Table 4).

Table 4. Indicators Themes and their frequencies (source: [36]*).*

Information type and cluster	Indicator Themes	Frequency
Public health and epidemiological		
Spread and death		
	Cases	High
	Deaths	High
	Recovered (healed, cured)	Medium
	Active cases	Medium
Testing		
	Testing (total number tested, PCR tests)	Medium
	Testing rates (positivity, negative tests)	Medium
Health system management		
	Hospitalized (admissions, discharges)	Medium
	Admitted to ICU (critical condition)	Medium

The study has also summarized the types of analysis and presentation of data, i.e., explored the frequency of considerations such as time trends, geographic levels and disaggregation options (sex, age, etc.). The majority of dashboards (87,3%) reported indicators over time, showing new cases or events appearing daily, and/or cumulative number of cases. In addition to these breakdowns, the categorization by age (82,3%) and sex (74,0%) was performed, considering a subset of 96 dashboards from 158. Besides of the fact that 70,3% of dashboards integrated maps with associated indicators, they have also represented these indicators by region using data visualization formats, facilitating the comparation of values depending on location.

The analysis of dashboards also showed a tendency to combine two or more presentation types. One of the most notable examples is the combination of age and sex to visualize indicators such as confirmed cases, deaths and vaccinations, with the purpose of more perceptible comparation and space saving. Additionally, the temporal evolution of indicators' values for different locations was displayed to understand the trends for each location. Moreover, various dashboards represented daily indicators as a cumulative function, where each day represents the total number of events related with indicator since the beginning of the pandemic. This function was usually used for confirmed and active cases, deaths, testing and vaccinations, for better comprehension of the COVID-19 evolution.

Therefore, to simplify the posterior Delphi process, one will consider only the indicator theme from the first row of Table 6 ("Cases"), at the same time comprising most common presentation types included in dashboards. The selection of indicator theme "Cases" can be justified by its highest frequency in public COVID-19 dashboards, while the inclusion of most frequent presentation types enables to examine different perceptions by the public. This way, the selected indicators for the posterior application of approach are the following:

- Daily number of new confirmed cases.

- Cumulative number of confirmed cases.
- Daily number of new confirmed cases by region.
- Total number of confirmed cases by age group.
- Total number of confirmed cases by sex.
- Total number of confirmed cases by region.
- Total number of confirmed cases by sex and age group.

5.2 Implementation

Afterwards, the previously detailed approach was implemented, consisting in two main steps:

1. The data structure of each COVID-19 indicator was identified, and the pre-set of data visualization formats was selected according to correspondent communication purpose and variable's type.
2. The Delphi process was designed and implemented in order to understand the preferences of the general population and to select the most appropriate data visualization formats for public COVID-19 dashboard.

5.2.1 Selection of pre-set of data visualization formats

The objective of this section is to provide a set of data visualization formats that are considered most appropriate for each of the selected indicators, taking into consideration the theoretical evidence. For that purpose, the selection method developed in Chapter 4, was implemented in three main steps:

1. The data structure of each indicator was identified, i.e., number of variables and their attribute type was determined.
2. The communication purpose of data visualization formats for each indicator was discovered, considering the categorization provided by Kirk [9].
3. The set of data visualization formats was defined by mapping previously determined data structure and communication purpose using a decision table developed in Chapter 4.

Firstly, the indicators required the analysis regarding their data structure, which was achieved by organizing data into attributes ensuring relevant relationships representation. Therefore, each indicator was modeled as $m \times n$ data table. Taking as example the indicator "Cases by day", it can be expressed as follows:

Table 5. Example of indicator "Cases by day".

Number of cases	Day
A_{11}	A_{12}
A_{21}	B_{22}
…	…
A_{n1}	B_{n2}

Afterwards, the identification of variables and their categorization was performed. Considering the previous example, the first variable "Cases" attribute comprises the quantitative data values, along with quantitative values of temporal variable "Day", representing the data set as string "QQ" (2 quantitative variables). It is noteworthy to mention that "Day" can alternatively be represented as categorical variable, considering a time-based scale as the limited number of aggregated ranges of quantitative values.

Thereafter, the communication purpose's method pretended to be acquired with visual representation of this data set was identified. In the presented example, the objective of data visualization is to show the changes over time, i.e., to investigate how quantitative values of variable "Cases" change over units of time ("Days"). Therefore, other temporal indicators were integrated by the same method classification, since their main purpose is to exploit temporal data and show the changing trends regarding COVID-19 information. The primary communication purpose of indicators containing categorical variables (Age Group, Sex and Region) was to compare categorical values, nevertheless, it was decided to research the part-to-whole connection as well, in order to investigate more about the public's view regarding this issue.

It is worthy to mention that the selection of data visualization for indicators with presentation type "Regions" is determined by their geospatial context: if the indicator's data set contains coordinates for the exact location with an associated value in the form of QQQ (3 quantitative variables), the communication purpose shifts from "Comparing categorical values" to "Mapping geo-spatial data", where the suggested set of data visualizations contains different maps.

Since the Dastani process [20] focuses on the structural correspondence condition for effective data visualization, one can assume that all the indicators that has equal data structure and communication purpose, would be perceived in the similar manner, using the same data visualization formats for their representation. This statement gives the certain flexibility regarding the selection of indicators, and consequently, of the corresponding set of data visualizations, included in Delphi process. This flexibility confers the right of choosing the restricted set of indicators for which the data visualization formats will be selected, and afterwards, apply them to the indicators with similar dataset structure and communication purpose. This way, the selection process is simplified, at the same time widening opportunities for future investigation.

The same logic was applied to other indicators, and the detailed description of data structure's and communication purpose's identification for indicators' theme "Cases" were derived in Annex C. This way, the selected indicators and corresponding data visualization formats, which will be validated by the Delphi process, are:

- Daily number of new confirmed cases: Line Chart, Area Chart, Column Chart.
- Cumulative number of confirmed cases: Line Chart, Area Chart, Column Chart.
- Daily number of new confirmed cases by region: Line Chart, Area Chart, Grouped Column Chart, Stacked Column Chart.
- Total number of confirmed cases by age group: Bar Chart, Column Chart, Pie Chart, Donut Chart, 100% Stacked Bar Chart.
- Total number of confirmed cases by sex: Bar Chart, Column Chart, Pie Chart, Donut Chart, 100% Stacked Bar Chart.
- Total number of confirmed cases by region:
    - Bar Chart, Column Chart, Pie Chart, Donut Chart, 100% Stacked Bar Chart.

    - Bubble Map, Choropleth Map, Dasymetric Map, Point Map.
- Total number of confirmed cases by sex and age group: Two-sided Bar Chart, Grouped Column Chart, Stacked Column Chart, 100% Stacked Column Chart.

5.2.2 Delphi process

Drawing on the outcome of previous step, described above, one can distinguish a gross list of potential indicators for further analysis regarding their visual representation. Furthermore, the modified two-round Web Delphi process was designed to refine the set of data visualizations corresponding to each indicator. Namely, the main objective of this process is to identify the most preferred visualization formats among the public, in order to improve the quality of the transmission of information regarding the evolution of the COVID-19 pandemic in Portugal through public web-based dashboard.

Since its inception, the Delphi has undergone a series of modifications with no universally agreed guidelines surrounding its appropriate design, "expert" definition or appropriate expert panel size. In the next sections, some of the preliminary considerations regarding Delphi design are elaborated.

a) Panel selection

Delphi uses non-probability sampling to recruit participants or create an 'expert panel'. For the purpose of this research, participants were defined as a sample of possible dashboard users. Users were recruited according to the 'snowballing' method using social media, which consists in identifying persons matching the selected profiles in the immediate network of contacts and then, if necessary, asking these persons to recruit others to represent the missing profiles. This method for recruiting respondents was not intended to generate a representative sample of the study population but to constitute a sample of individuals with very divergent opinions, thereby representing a wide spectrum of points of view. The diversity of these views can then be indirectly confronted with each other by the Delphi process. It is noteworthy to mention that the heterogeneity of the selected panel suggests the better performance as it allows for a wider range of alternatives and perspectives to ensure generalizability.

b) Delphi type

The Delphi process used in this study can be considered as modified, since the credibility of the questions elaborated for the first round is ensured by scientific background provided from the initial stage.

According to research, panel attrition between Delphi iterations owing to panel fatigue is one of the most significant contributors to methodological bias, particularly in studies with large participant numbers. It's likely that the advantages of using web-based Delphi survey methodologies reduce panel fatigue and attrition between rounds. At the same time, Delphi research utilizing online surveys has been proven to lower costs and facilitate data processing. Overall, using online surveys has shown to be a successful method for conducting Delphi research [120]. Therefore, the web-based Delphi was the preferred option for this study, considering the heterogeneity and high number of participants.

c) Iteration

Attrition can be a problem in Delphi studies and various techniques have been suggested to minimize it. As attrition is likely to increase with each round, to ensure against participant fatigue, but to guarantee results are meaningful two rounds seemed optimal, therefore an *a priori* criterion of two rounds in this Delphi study was established. Additionally, the quick turnaround of data was established as a mean of maintaining participant interest. This way, the deadline of two weeks for each round was determined, and new round was sent out one week after closure of the previous one.

d) First Round

To determine the preferences of general population regarding alternative data visualization formats, a list of initial indicators and corresponding visualization options was previously generated. First round of Delphi took the form of a structured questionnaire, including statements generated from the previous list, with the objective of acquiring collective knowledge about which data visualization format is preferred for each indicator from the pre-selected set.

The elaborated questionnaire consisted of four main parts. In the first part, the generic information about Delphi process was provided to clearly show what is required and to remove ambiguity. The participants were asked about some socio-demographic characteristics, namely, their age group, sex, highest level of education completed. Additionally, they were asked if they have an experience working within health sector, and if they are familiar with COVID-19 dashboards.

Second part explored the data visualization formats of temporal COVID-19 indicators, namely the daily number of new confirmed cases (Figure 14). Additionally, the cumulative function derived from this indicator was interesting to explore due to its abundancy in COVID-19 dashboards, which represents the cumulative number of confirmed cases since the beginning of the COVID-19 pandemic. Moreover, the analysis of perceptions captured from observing different time intervals (one week versus three months) in certain data visualizations was considered to be useful to understand the way cognition evaluates the variance in the COVID-19 cases on different screen amplifications, e.g., while using zoom function, commonly integrated in dashboards. Finally, the comparison of evolution of pandemic by some criterion, e.g., by ARS, was considered to be valuable for future investigation. Data visualizations suggested for the temporal indicators by previous analysis were line, bar and area charts. Furthermore, the stacked column chart was considered to be effective for comparison of cases by different criterion.

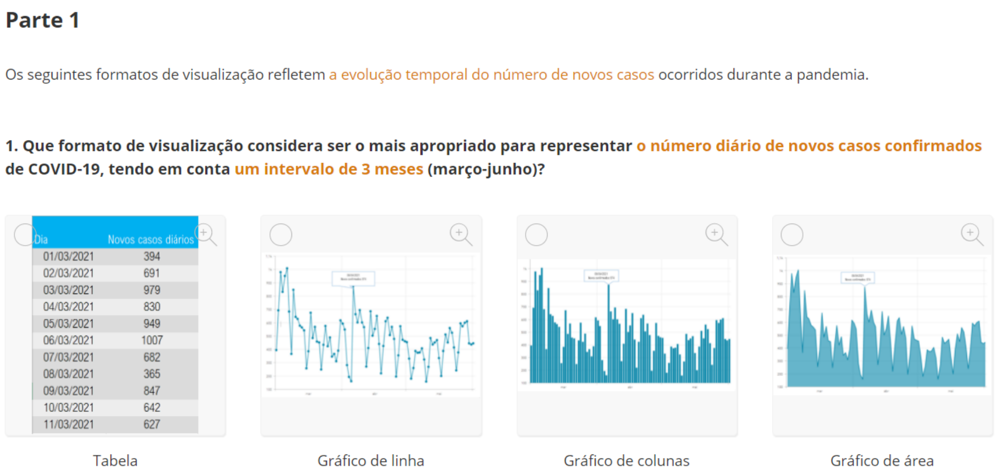


Figure 14. Part 1 of first-round questionnaire – Question 1 (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering three-month time interval (March-June)?")

The third part explored the data visualizations corresponding to total number of confirmed cases since the beginning of the COVID-19 pandemic, aggregated by age group, sex or ARS (Figure 15). The column, bar, pie and donut charts were suggested as possible alternatives for correspondent indicators. Additionally, the 100% stacked bar chart was considered important in order to visually compare the percentage of each category. Furthermore, it was considered relevant to analyze the perception of public regarding the representation of two categories at the same time, using different positioning of columns and bars – grouped column chart, two-sided bar chart and (100%) stacked column chart. It is trustworthy to mention that the options for each question were completed by integrating the corresponding data set table as a reference point.

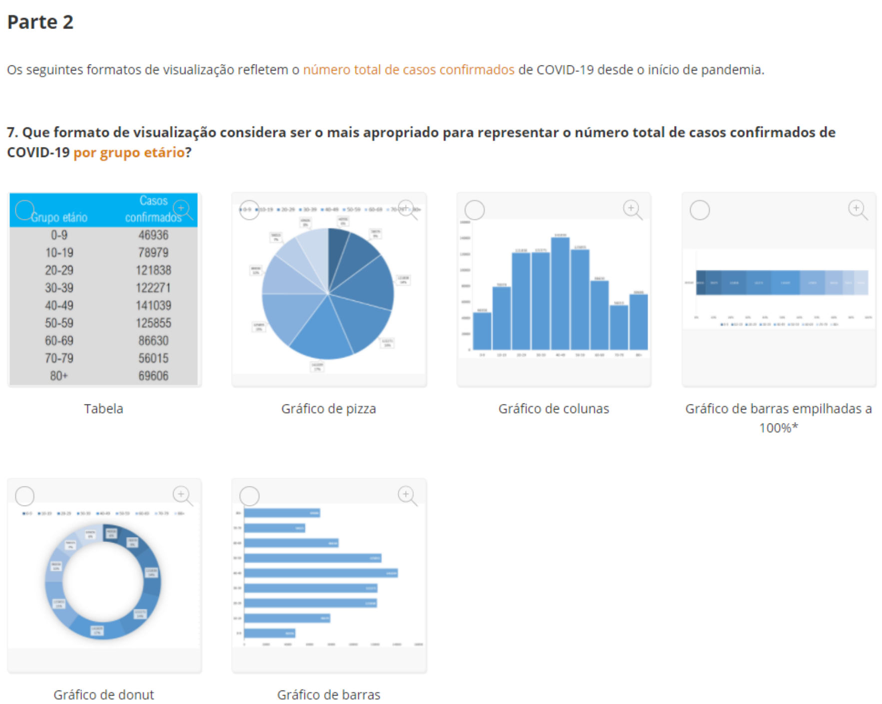


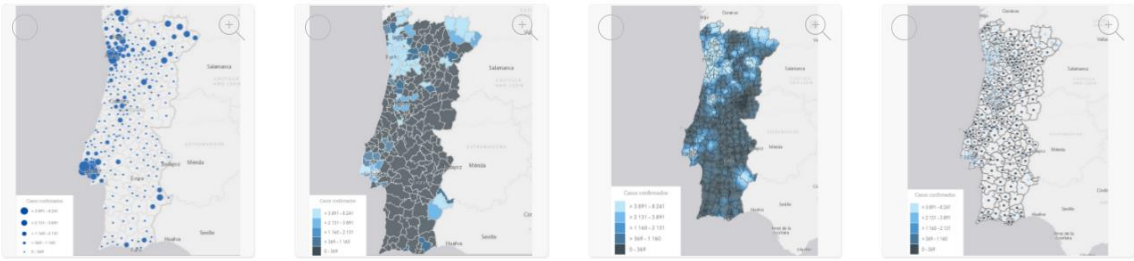
Figure 15. Part 2 of first-round questionnaire – Question 7 (trans. "What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by age group?").

Finally, the last part of questionnaire corresponded to the map visualizations, integrating the project SCOPE (Figure 16). The first question offered the set of map visualizations regarding the indicator "Total number of confirmed cases by municipality since the beginning of the COVID-19 pandemic", including choropleth, bubble, dasymetric and point maps. The last question approached SCOPE project, with an aim to understand the preferred map format for spatial distribution of COVID-19 infection risk. Therefore, two map's formats were suggested: the first one corresponded to the map with constant infection risk within the administrative unit, and another represented the SCOPE map with infection risk, which is continuous in the space. The developed questionnaire for first round of Delphi process can be found in the Annex D.

Parte 3

11. Que mapa considera ser o mais apropriado para representar o número de novos casos confirmados de COVID-19* por concelho desde o início da pandemia ?

*incidência



Mapa de bolas | Mapa coroplético | Mapa dasimétrico | Mapa de pontos

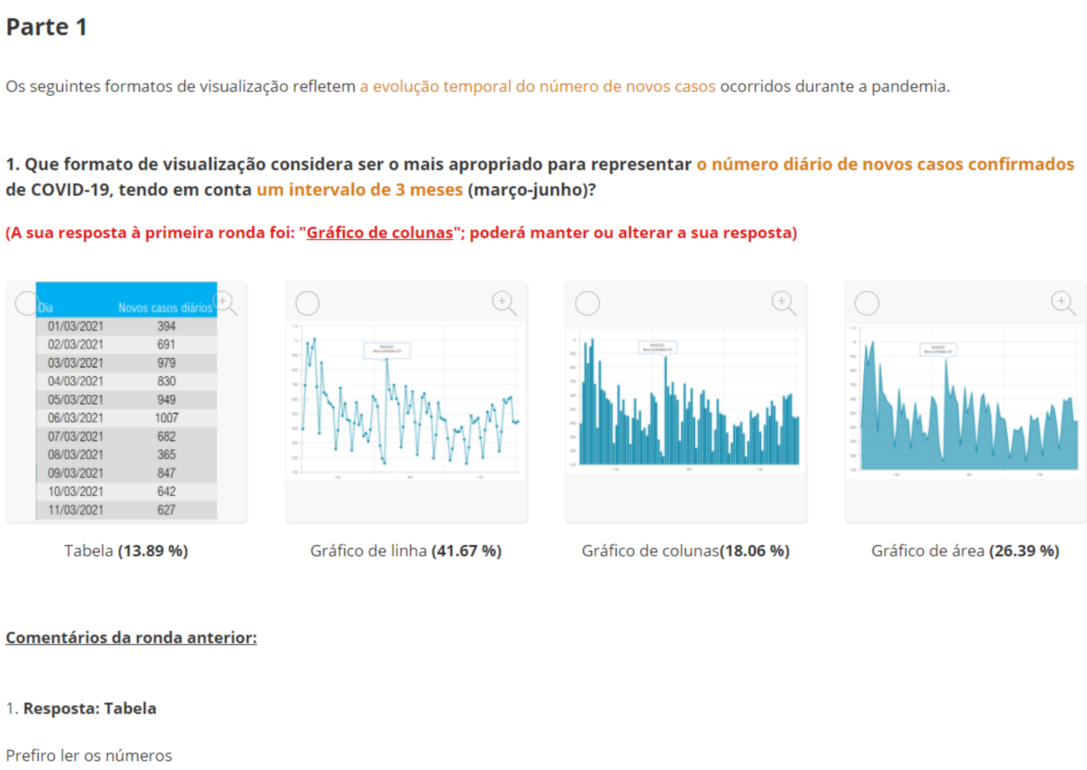
Figure 16. Part 3 of first-round questionnaire – Question 11 (trans. "What map do you consider to be the most appropriate to represent a number of new confirmed COVID-19 cases per municipality (incidence) since the beginning of pandemic?).

The questionnaire gave participants the opportunity to provide further comment regarding each question, if desired, as the rounds progress. Additionally, they were allowed to write free-text comment at the end of questionnaire including suggestions for rephrasing, combining, or reformatting. This process reduced attrition by making participants feel like they are partners in the study, ensuring participants take ownership. Also, contact details were provided for any points requiring clarification.

e) Subsequent round

After concluding the first Delphi round, individual participants' answers were synthesized, and the statistical summary was elaborated in a form of percentage scores for each item.

In round 2 of the Web-Delphi process participants were reminded in the web-platform of their own responses and could additionally visualize a synthesis of the percentage of respondents for distinct data visualizations (Figure 17). The objective of this round was to give participants the opportunity to confirm or change their answers, considering the group information provided, within a collective learning task. The participants could also visualize the comments of the other respondents from round 1. The example of developed questionnaire for second round of Delphi process can be found in the Annex E.



Parte 1

Os seguintes formatos de visualização refletem a evolução temporal do número de novos casos ocorridos durante a pandemia.

1. Que formato de visualização considera ser o mais apropriado para representar o número diário de novos casos confirmados de COVID-19, tendo em conta um intervalo de 3 meses (março-junho)?

(A sua resposta à primeira ronda foi: "Gráfico de colunas"; poderá manter ou alterar a sua resposta)

Dia	Novos casos diários
01/03/2021	394
02/03/2021	691
03/03/2021	979
04/03/2021	830
05/03/2021	949
06/03/2021	1007
07/03/2021	682
08/03/2021	365
09/03/2021	847
10/03/2021	642
11/03/2021	627

Tabela **(13.89 %)**

Gráfico de linha **(41.67 %)**

Gráfico de colunas**(18.06 %)**

Gráfico de área **(26.39 %)**

Comentários da ronda anterior:

1. **Resposta: Tabela**

Prefiro ler os números

Figure 17. Part 1 of second-round questionnaire – Question 1 (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering three-month time interval (March-June)? Your response in the first round was: "Column Chart"; you can maintain or change your answer. First-round comments: 1. Answer: Table. I prefer to read numbers.").

f) Implementation

The most important feature of survey tool considered for this study was the possibility to integrate image as a multiple-choice option. Additionally, images should have a zoom-in option for better analysis of the proposed data visualizations. Besides that, the survey tool should have user-friendly interface in order to be easy and attractive for the participants.

Therefore, the survey tool selected was the one designated as SurveyHero due to its compatibility with requirements presented above. It is known for its user-friendly interface, free or low-cost packages and for being one of the most used survey tools. Additionally, SurveyHero presents the "Image Choice" question type, which allows to create a multiple-choice question using images as the possible answers. The user can also add a caption below each image as well as let participants zoom-in on the image to see all of the details.

g) Analysis of Delphi results

There are several design characteristics and techniques for analyzing Delphi procedures that have been documented in the literature. However, there isn't a comprehensive explanation or a step-by-step guidance on how to properly construct a Delphi process, which leads to a slew of errors in both assessment and interpretation [134].

With the exception of Policy Delphi processes, the fundamental objective of a Delphi may be regarded the measuring of expert panel consensus. Unfortunately, consensus is one of the most divisive aspects of the Delphi approach, and its measurement differs significantly between studies, due to controversial understanding of the term. As a result, studies have employed a variety of methods to assess when the expert panel has achieved a suitable level of agreement [135].

The definition of consensus in Delphi studies depends on the objectives of the research and can be used to determine if agreement exists or as a stopping guideline [136]. In this study, an *a priori* criterion on number of rounds was set, so the consensus served to determine the existence of agreement between the participants regarding the most preferred data visualization format for indicator in question.

Researchers have employed a variety of statistics to measure consensus, according to the literature review [137]. The application of statistical tests depends on the level of data, which is determined by one of four data measurements scales: nominal, ordinal, interval and ratio [135]. Since each option of this Delphi process corresponds to the data visualization format, the scale used is a nominal one. Nominal scale contains rules for deciding if objects are equivalent or not, by identifying them through labels [138]. In the case of nominal-scale data, simple statistical tests are usually applied, due to non-numerical data limitations.

Many Delphi processes use certain levels of agreement in order to quantify consensus among an expert panel [135]. This quantification is important in order to know which data visualization formats are seen as most preferred in each round to see the final level of agreement concerning each indicator. The percentage of agreement is the simplest measure of level of agreement. It is calculated as the number of times an option was chosen (frequency rate), divided by the total number of units of observation that are rated, multiplied by 100. The determination of consensus by a certain percentage of agreement is particularly meaningful, if nominal scales for the degree of agreement are used [135]. Afterwards, the mode can be used to determine the option most voted by the participants [139]. It is a measure of central tendency that refers to the proportion of experts that choose the item with most popular rate [140]. According to Green, the mode gives the most accurate picture of the views that have been expressed by the experts [141].

Therefore, to analyze the case study results, it was decided to calculate the frequency and percentage of votes for each option regarding all the questions for first and second rounds, and additionally, provide the most common data visualization format voted (mode). Moreover, if a percentage value of mode was above 50% of the total respondents, then consensus was assumed, and the correspondent option was qualified as the one with highest level of agreement.

In Delphi research, quality criteria that ensure a particular scientific standard are crucial to consider [135]. Therefore, it is critical to evaluate the Delphi's reliability and validity throughout its design, facilitation, and reporting in order to maximize research rigor and credibility. The degree to which a study process produces consistent results each time it is repeated is known as reliability [134]. In another words, reliability refers to the consistency and stability of the measurement instrument and means freedom from random error [135]. It is important to produce stable and dependable results, and it has been suggested that agreement is worthless until group stability is achieved [136]. Stability refers to the consistency of responses between successive rounds of a study. It occurs when the responses obtained in two successive rounds are shown to be not significantly different from each other, irrespective of whether a convergence of opinion occurs [142]. There are a number of statistical tests to determine stability, however, tests which are suitable for use on nominal data are highly restricted. The Chi-squared test has been used to evaluate stability, however it is not recommended for Delphi research since it determines “the independence of the rounds from replies found in them” rather than the stability of response between rounds [136]. The calculation of percentage of participants that doesn’t change their response was suggested as a measure of stability in this study. Therefore, in this study a statement was considered stable when 70% or more of the participants do not change their responses [3].

The degree to which a test measures what it claims to measure is known as validity [134]. The validity is connected to the consensus criteria: stricter criteria will provide more validity for the results, but it will be more difficult to assess consensus [134]. The methods to ensure content and external validity were implemented in this study. The extent to which a Delphi research covers all and only the problems relevant to the topic of interest is referred to as the study's content validity [134]. In this study, it was enhanced by ensuring that the panel is a representative sample of relevant participants who are motivated to participate, as well as providing panelists enough chance to comment on the topics that are being evaluated for inclusion in the survey [134]. On another hand, the external validity of a research reflects the extent to which its findings may be applied to a larger population. This study improved external validity by using as large, representative and heterogeneous sample of participants as possible.

5.2.3 Implications of the Delphi results for dashboard design

According to the previous research, public web-based COVID-19 dashboards present various limitations, specifically an overall underuse of known and proven data delivery techniques, e.g., visualization techniques, from the perceptive of final user [36]. The lack of support to guide visualization choices for diverse dataset domains, and absence of participatory methods in the dashboard development contribute to the inadequate transmission of information to the general public, resulting in misinterpretation of data by population.

Therefore, the results of the case study can improve already developed COVID-19 dashboards, which represent same indicators selected for the research. This way, data visualization formats used by dashboards can be substituted by new ones, considered more preferable due to the results of case study, enhancing a perception of general population.

In Portugal, the national health authority official online service is the DGS (Direção-Geral da Saúde – Directorate-General of Health) interactive platform. This platform reproduces the official reports, representing number of confirmed COVID-19 cases daily, cumulatively or aggregated by certain category. Additionally, this dashboard integrates the map visualization to display the number of cases by geographical location. Therefore, this dashboard was chosen to be the one for further improvement with the results of case study.

6 Case study results

This chapter presents the results of implementing the developed approach for the context of COVID-19 pandemic. More specifically, the results of Delphi process are provided for each round, and their analysis is performed. Moreover, the implications of the Delphi results for DGS dashboard design are presented.

6.1 Delphi participation

The Delphi process was developed to select the data visualization formats that are most preferred by the participants for each indicator. As has been already mentioned, participants were recruited according to the ‘snowballing’ method by using social media, with the result of 101 recruiters from different areas of knowledge.

During the first round, the panelists were required to select the data visualization formats that they prefer the most between suggested alternatives, for each indicator in question. In the final section of questionnaire, participants were asked to leave their e-mail in order to be contacted for the second round, with 72 participants from a total of 101 (71,3% completion rate) having decided to fill their e-mail. The average completion time recorded by survey was 05:55 minutes, that was considered to be sufficient to answer all the questions without provoking any fatigue.

The participants’ composition presented certain heterogeneity in terms of gender (female – 59,6%, male – 40,4%), however, the prevalence of participants within age group of 20 to 29 years was encountered. In terms of highest educational level, 17% of participants have completed high school, 34% - bachelor’s degree and a master’s degree was achieved by 44,7% of participants. Additionally, 30% of participants had the professional activity related with health sector, and 61,7% of them were familiar with the concept of COVID-19 dashboard.

For the second round, the response rate was 65%, with 47 panelists decided to participate in the Delphi study. The participants were offered the possibility to change or maintain their responses, considering the distribution of first-round votes provided for each option.

6.2 Delphi results

For first and second rounds, the frequency and percentage of votes for each option regarding all questions were calculated. Afterwards, modes for each round were determined to reflect about central tendency, and finally, were emphasized by using orange color. Additionally, the data visualization format corresponded to the mode of the second round was highlighted by blue rectangle to facilitate the perception. At the same time, to understand if the consensus could be assumed, the data visualization with more than 50% of votes was mentioned if existent.

a) First part

As have already been mentioned, the first part consisted of six questions exploring the data visualization formats of temporal COVID-19 indicators. For the first two questions, the indicator to be represented was “Daily number of new confirmed COVID-19 cases”, considering two different time intervals – 3 months and 7 days.

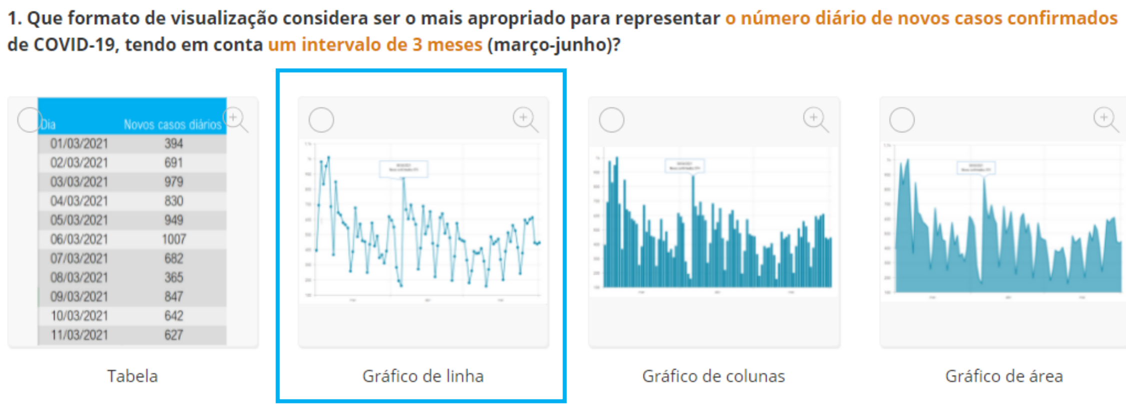


	Table	Line Chart	Column Chart	Area Chart
Round 1	10 (13,9%)	30 (41,7%)	13 (18,1%)	19 (26,4%)
Round 2	4 (8,5%)	22 (46,8%)	12 (25,5%)	9 (19,1%)

Figure 18. Question 1 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering three-month time interval (March-June)?"). Frequency and percentage of votes for first and second rounds.

For the first question (Figure 18), related to three-month interval, in the first round the mode with the value of 41,7% was attributed to line chart. Moreover, the most voted answer in the second round continued to be a line chart, with 46,8% of percentage agreement. It is noteworthy to mention that the percentage value increased (from 41,7% to 46,8%), appointing to evolution towards higher level of agreement across rounds. Additionally, the variations in other categories were slight, however, provoking the switch of preference regarding column and area charts. Although the mode value did not reach the limit of 50% to conclude the consensus between participants, this value still can be considered very close.

	Table	Line Chart	Column Chart	Area Chart
Round 1	12 (16,7%)	27 (37,5%)	28 (38,9%)	5 (6,9%)
Round 2	8 (17,0%)	16 (34,0%)	22 (46,8%)	1 (2,1%)

Figure 19. Question 2 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases, considering one-week interval?"). Frequency and percentage of votes for first and second rounds.

Regarding the next question (Figure 19), this time considering smaller time interval, the column chart was recognized as the most voted one, with the mode of 38,9%. However, it is worth to note that the mode value of column chart was very proximate to the one of the line chart, not allowing to have any definitive conclusion. In the second round, the column chart continued to be most preferred with 46,8% of votes. The increase in mode percentage indicates the evolution of group judgements towards a higher level of agreement along two rounds. Besides that, the value was very proximate to achieve consensus between participants.

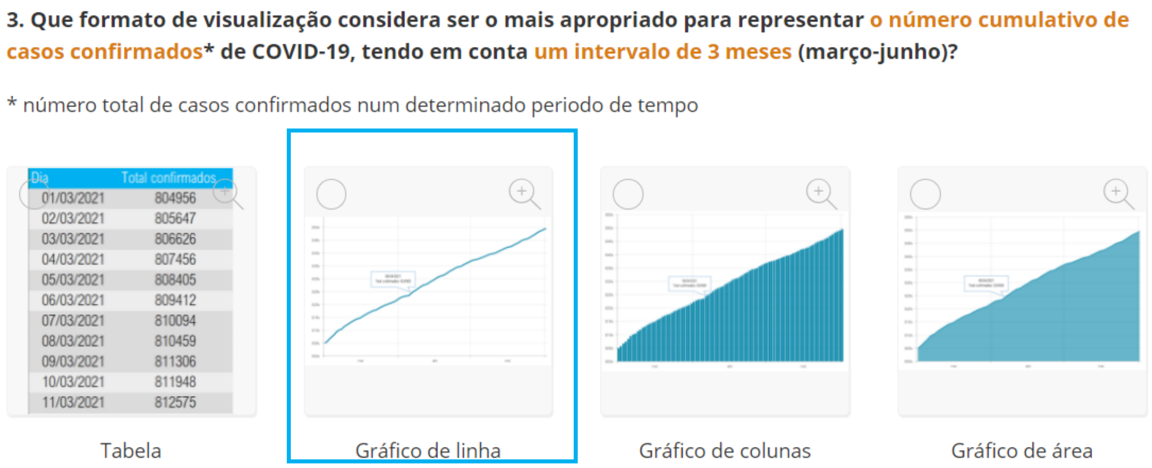


	Table	Line Chart	Column Chart	Area Chart
Round 1	6 (8,3%)	35 (48,6%)	16 (22,2%)	15 (20,8%)
Round 2	3 (6,4%)	30 (63,8%)	6 (12,8%)	8 (17,0%)

Figure 20. Question 3 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a cumulative number of confirmed COVID-19 cases, considering three-month time interval (March-June)?"). Frequency and percentage of votes for first and second rounds.

Next, the preferred visual interpretation of cumulative number of confirmed cases of COVID-19 was studied over two time intervals: three-month and one week. In this case, the majority of votes for the third question (Figure 20), with the longer time interval explored, went for the line chart with percentage of 48,6%. In the second round, the most voted option continued to be line chart with 63,8% of percentage agreement, surpassing the predefined threshold of 50% of achieving consensus. Moreover, the increase in mode value also indicated an evolution towards higher level of agreement.

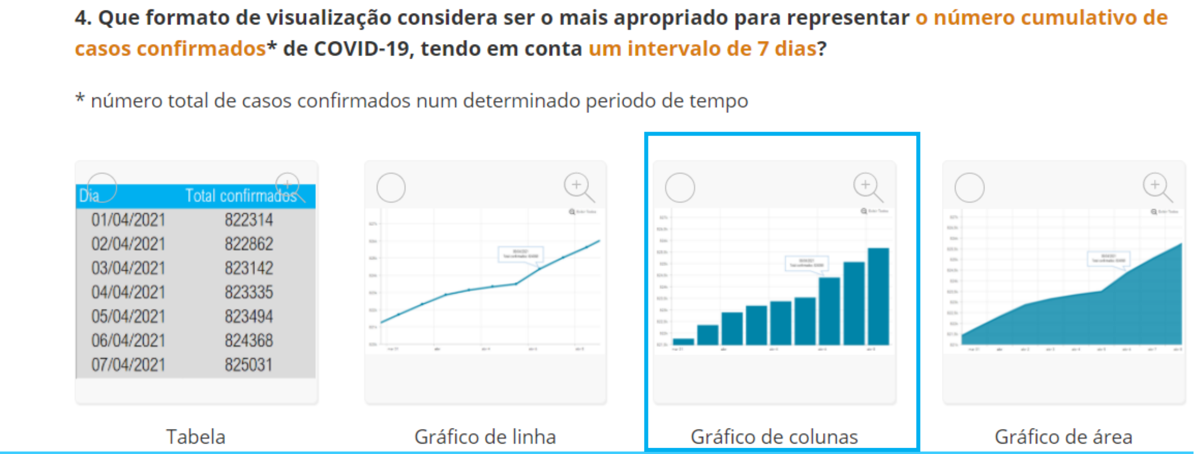


	Table	Line Chart	Column Chart	Area Chart
Round 1	8 (11,1%)	22 (30,6%)	34 (47,2%)	8 (11,1%)
Round 2	5 (10,6%)	10 (21,3%)	29 (61,7%)	3 (6,4%)

Figure 21. Question 4 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a cumulative number of confirmed COVID-19 cases, considering one-week interval?"). Frequency and percentage of votes for first and second rounds.

In the first round, the column chart was the preferred one with 47,2% for question 4 (Figure 21), studying the shorter time interval. In the next round, the most voted visualization for cumulative number of COVID-19 cases continued to be line chart, increasing the percentage agreement and reaching the consensus with 61,7%.

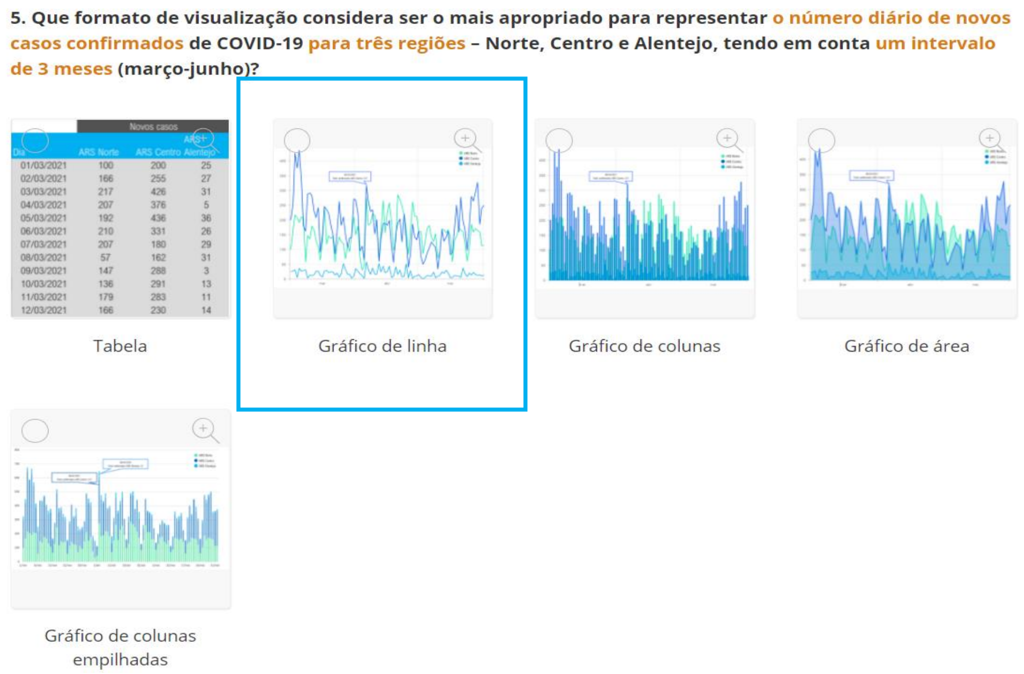
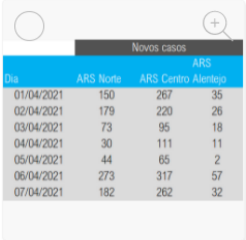


	Table	Line Chart	Column Chart	Area Chart	Stacked Column Chart
Round 1	20 (27,8%)	22 (30,6%)	3 (4,2%)	13 (18,1%)	14 (19,4%)
Round 2	9 (19,1%)	22 (46,8%)	5 (10,6%)	3 (6,4%)	8 (17,0%)

Figure 22. Question 5 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases for three regions – Norte, Centro e Alentejo, considering three-month time interval (March-June)?") Frequency and percentage of votes for first and second rounds.

6. Que formato de visualização considera ser o mais apropriado para representar o número diário de novos casos confirmados de COVID-19 para três regiões – Norte, Centro e Alentejo, tendo em conta um intervalo de 7 dias?



Novos casos			
		ARS	
Dia	ARS Norte	ARS Centro	Alentejo
01/04/2021	150	267	35
02/04/2021	179	220	26
03/04/2021	73	95	18
04/04/2021	30	111	11
05/04/2021	44	65	2
06/04/2021	273	317	57
07/04/2021	182	262	32

Tabela

Gráfico de linha

Gráfico de colunas

Gráfico de área

Gráfico de colunas empilhadas

	Table	Line Chart	Column Chart	Area Chart	Stacked Column Chart
Round 1	10 (13,9%)	18 (25,0%)	25 (34,7%)	6 (8,3%)	13 (18,1%)
Round 2	6 (12,8%)	12 (25,5%)	25 (53,2%)	1 (2,1%)	3 (6,3%)

Figure 23. Question 6 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a daily number of new confirmed COVID-19 cases for three regions – Norte, Centro e Alentejo, considering one-week interval)?"). Frequency and percentage of votes for first and second rounds for Question 6.

Next two questions investigated the daily number of new confirmed cases of COVID-19 for three regions, taking into account two time intervals: 3-month and one week. The objective of these question was to determine if perception changes when different categories are compared on the same chart, within different time intervals.

For these two questions, the answers of Round 1 were inconclusive, since the percentages of agreement were very proximate. Besides that, the mode was attributed to line chart (30,6%), in case of question 5 for longer period of time (Figure 22), and to column chart (34,7%), in case of question 6 for shorter period of time (Figure 23). For these questions, charts with 50% and above were inexistent, contributing for the lack of clarity of results.

In second round, for 3-month period, the most voted option continued to be line chart, with percentage value increased (from 30,6% to 46,8%) (Figure 22), appointing to evolution towards higher level of agreement across rounds. On the other hand, for shorter time interval – one week – the most preferred option continued to be column chart, with 53,2% (Figure 23), reaching the consensus between the participants. The tendency towards higher level of agreement (from 34,7% to 53,2%) was also notable.

It is noteworthy to mention that the most voted options from first two questions are compliant with the results of questions 3 and 4, as of questions 4 and 5, respectively, where the line chart was chosen to interpret the graphs with longer time intervals, and column chart – for events which occur within shorter period of time.

b) Second part

For this section, the data visualization formats corresponding to the total number of confirmed cases since the beginning of the COVID-19 pandemic were investigated, aggregated by region, sex and/or age group. The first three questions investigated aggregation by age group, sex and region, respectively.

	Table	Pie Chart	Column Chart	100% Stacked Bar Chart	Donut Chart	Bar Chart
Round 1	3 (4,2%)	28 (38,9%)	19 (26,4%)	2 (2,8%)	3 (4,2%)	17 (23,6%)
Round 2	0 (0,0%)	14 (29,8%)	19 (40,4%)	0 (0,0%)	2 (4,3%)	12 (25,5%)

Figure 24. Question 7 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by age group?"). Frequency and percentage of votes for first and second rounds.

Considering the indicator "Total number of confirmed cases by age group" in Question 7 (Figure 24), nine age groups were represented by different visualizations, where the preferred data visualization was pie chart, with the value of 38,9%. However, in second round the most voted option changed from pie to column chart, with the percentage of 40,4%, therefore this question showed inconclusive results. This switch can be related with the comment which the participant left during the first round of Delphi, advising not to use pie chart with high number of categories.

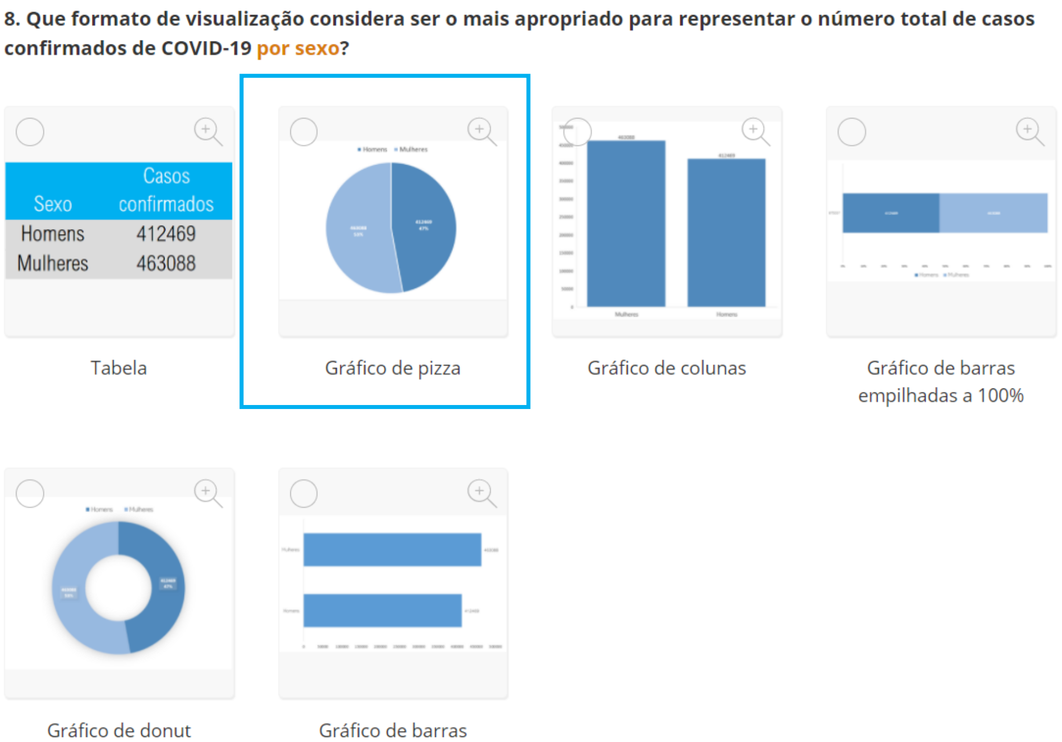


	Table	Pie Chart	Column Chart	100% Stacked Bar Chart	Donut Chart	Bar Chart
Round 1	5 (6,9%)	42 (58,3%)	9 (12,5%)	3 (4,2%)	10 (13,9%)	3 (4,2%)
Round 2	1 (2,1%)	35 (74,5%)	3 (6,4%)	2 (4,3%)	6 (12,8%)	0 (0,0%)

Figure 25. Question 8 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by sex?"). Frequency and percentage of votes for first and second rounds.

For the next question (Figure 25), the pie chart was the most voted one with the value of 58,3%, surpassing 50% threshold. The results from round 2 provided relatively high percentage agreement regarding the option "Pie Chart" – 74,5%, not only confirming the results from the first round but also increasing the level of agreement along rounds and determining the consensus between participants.

	Table	Pie Chart	Column Chart	100% Stacked Bar Chart	Donut Chart	Bar Chart
Round 1	11 (15,3%)	15 (20,8%)	24 (33,3%)	1 (1,4%)	6 (8,3%)	15 (20,8%)
Round 2	3 (6,4%)	9 (19,1%)	24 (51,1%)	0 (0,0%)	3 (6,4%)	8 (17,0%)

Figure 26. Question 9 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by region?"). Frequency and percentage of votes for first and second rounds.

For the question 9 (Figure 26), the preferred visualization corresponded to the cases aggregated by seven ARS were asked. In the round 1, a chart with the maximum level of percent agreement (33,3%) was the column chart, not allowing to have any definitive conclusion. In the next round this visualization format continued to be the most voted option with the percentage agreement of 51,1%, not only evolving toward higher level of agreement, but also reaching consensus between participants.

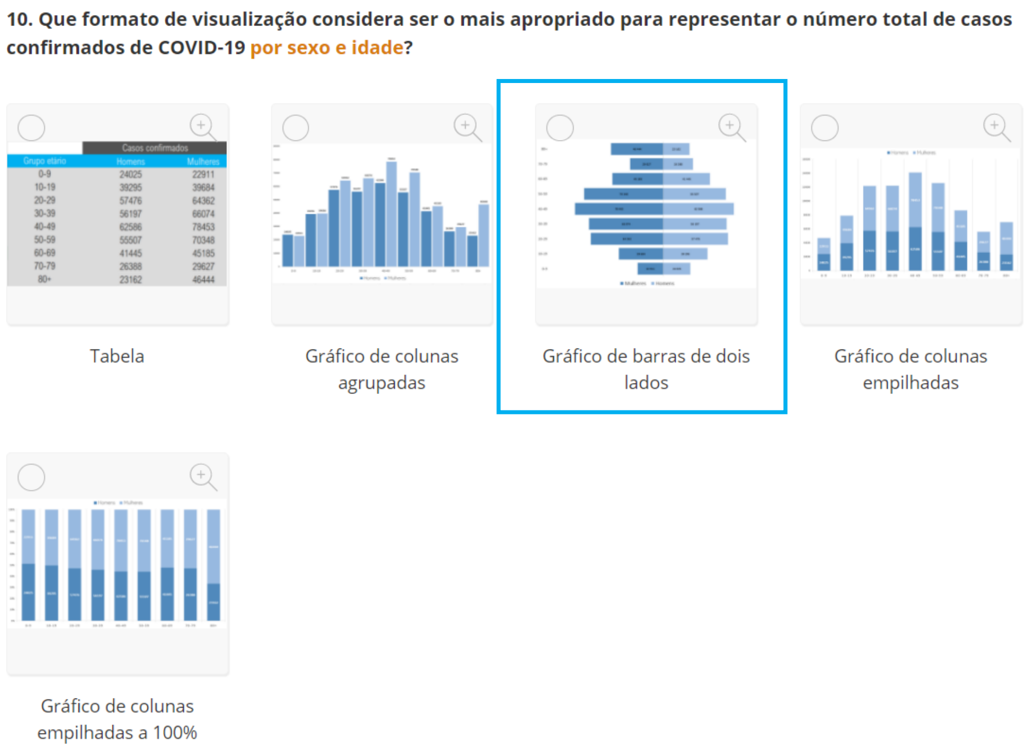


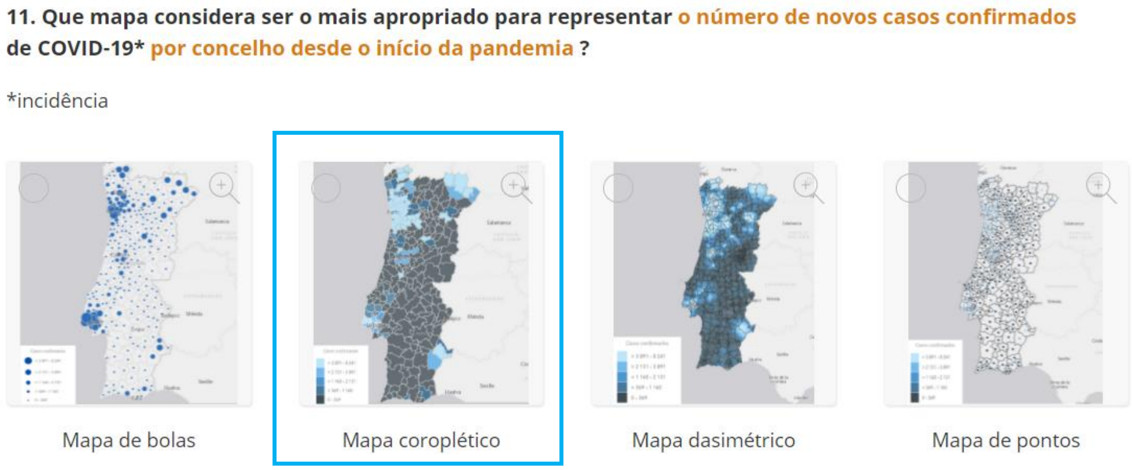
	Table	Grouped Column Chart	Two-sided Bar Chart	Stacked Column Chart	100% Stacked Column Chart
Round 1	7 (9,7%)	13 (18,1%)	45 (62,5%)	4 (5,6%)	3 (4,2%)
Round 2	3 (6,4%)	4 (8,5%)	37 (78,7%)	3 (6,4%)	0 (0,0%)

Figure 27. Question 10 of Delphi questionnaire (trans. "What visualization format do you consider to be the most appropriate to represent a total number of confirmed COVID-19 cases by sex and age group?"). Frequency and percentage of votes for first and second rounds.

Finally, for the last question of the second part of the questionnaire (Figure 27), the indicator "Total number of confirmed cases by age group and sex" was explored, with the objective to investigate the perception of participants regarding the comparation of two distinct categories. In this case, the two-sided bar chart was the most preferred data visualization format in the first round, surpassing the threshold of 50% with the percentage agreement of 62,5%. In the second round the two-sided bar chart continued to be the preferred option selected by participants, representing the percentage of 78,7%, reaching the consensus. It is noteworthy to mention that it was possible to observe an evolution of group judgements towards a higher level of agreement in this question.

c) Third part

The last part of questionnaire corresponded to the map visualizations, integrating the SCOPE project. The question 10 (Figure 28) was regarding the representation of COVID-19 incidence by most common map visualizations used in dashboards, including choropleth, bubble, dasymetric and point maps.



	Bubble map	Choropleth map	Dasymetric map	Point map
Round 1	19 (26,4%)	46 (63,9%)	7 (9,7%)	0 (0,0%)
Round 2	7 (14,9%)	38 (80,9%)	2 (4,3%)	0 (0,0%)

Figure 28. Question 11 of Delphi questionnaire (trans. "What map do you consider to be the most appropriate to represent a number of new confirmed COVID-19 cases per municipality (incidence) since the beginning of pandemic?"). Frequency and percentage of votes for first and second rounds.

It is noteworthy to mention that the point map was not voted at all, suggesting that it was considered as not appropriate for representation of COVID-19 cases, and highlighting its low perception. Additionally, the highest value of percent agreement was assigned to choropleth map (63,9%). It was also considered a preferred one in the second round with the percentage of 80,9%, being consistent with the first-round results, and reaching the consensus regarding this data visualization format. Moreover, the evolution towards a higher level of agreement has been observed during two rounds.

	A risk is constant within administrative unit	A risk is continuous in space
Round 1	51 (70,8%)	21 (29,2%)
Round 2	38 (80,9%)	9 (19,1%)

Figure 29. Question 12 of Delphi questionnaire (trans. "What map do you consider to be the most appropriate to represent a number of new confirmed COVID-19 cases per 100000 inhabitants (cumulative incidence), per municipality since the beginning of pandemic?"). Frequency and percentage of votes for first and second rounds.

For the last question of questionnaire (Figure 29), the two maps suggested by project SCOPE were integrated. The first one corresponded to the map with constant infection risk within the administrative unit, and another represented the SCOPE map with infection risk, which is continuous in the space. The last one has been the focus of the project since it reflects more precise information about COVID-19 risk distribution, however, the perception of public can be important to understand if this map can be well-interpreted by generic population.

However, the participants chose the maps with constant infection risk within the administrative unit with the agreement percentage of 70,8%, confirming that the perception of general population doesn't depend on precision of representation. In the second round, the participants maintained the maps with constant infection risk within the administrative unit with the percentage of 80,9%, increasing the level of agreement along two rounds.

To summarize the second-round results, the next table (Table 6) was created with the mode and percentage agreement values for each question:

Table 6. Summary of second round results of Delphi process.

Question	**1**	**2**	**3**	**4**	**5**	**6**
Mode	Line Chart	Column Chart	Line Chart	Column Chart	Line Chart	Column Chart
Percentage agreement	46,8%	46,8%	63,8%	61,7%	46,8%	53,2%
Question	**7**	**8**	**9**	**10**	**11**	**12**
Mode	Column Chart	Pie Chart	Column Chart	Two-sided Bar Chart	Choropleth Map	A risk is constant within administrative unit
Percentage agreement	40,4%	74,5%	51,1%	78,7%	80,9%	80,9%

Additionally, Table 7 was developed to compare the percentages of how many people changed their opinion throughout the Delphi, in order to discuss whether the statements can be considered stable.

Table 7. The percentages of how many people changed their opinion throughout the Delphi process.

Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Percentage (%)	17,0	25,5	29,8	25,5	23,4	19,1	25,5	25,5	27,7	19,1	12,8	14,9

Regarding the change of opinion by the participants, this was around 20% conferring overall stability to the Delphi, and consequently, reliability of the process.

6.3 Implications of the Delphi results for DGS dashboard design

As has been already mentioned, DGS dashboard was chosen to be the one for further improvement with the results of case study. For that purpose, the analysis of DGS COVID-19 dashboard was performed in order to discover inconsistencies with case study results, and afterwards, the alternative data visualization formats were proposed with the final objective to develop an adjusted DGS COVID-19 Dashboard.

a) *Analysis of DGS COVID-19 Dashboard*

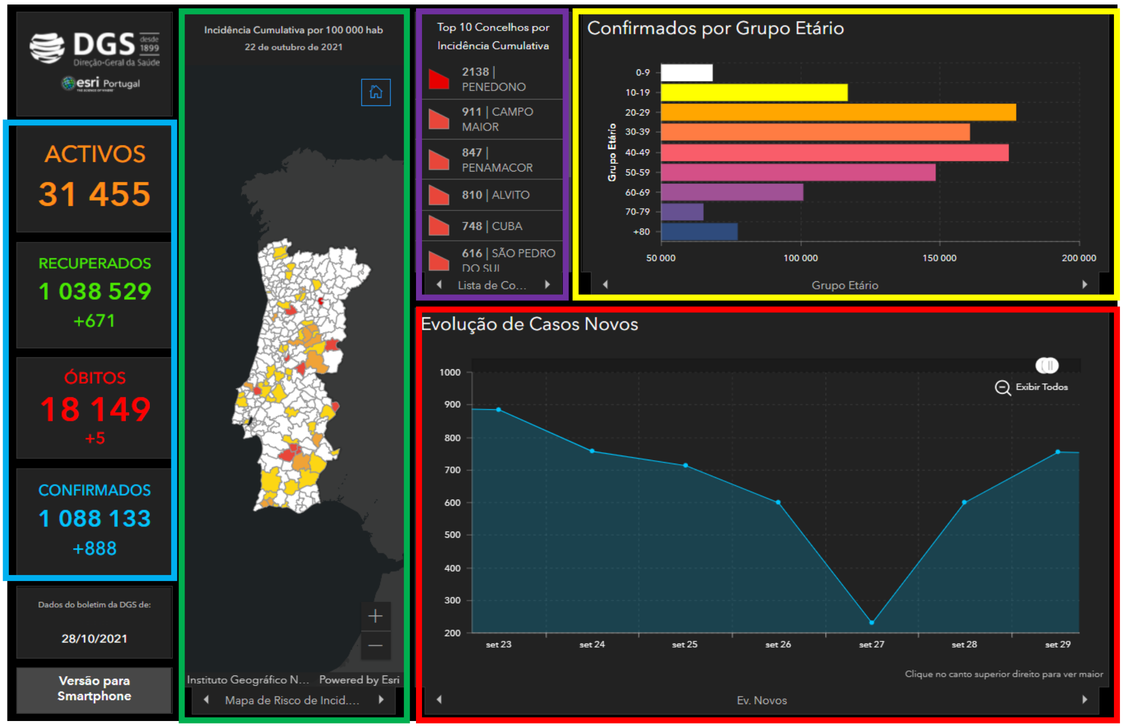


Figure 30. DGS COVID-19 Dashboard [43].

The design of the dashboard demonstrates the simplicity of interface, and the interactions are very user-friendly, considering Visual Information Seeking Mantra. The left part of dashboard represents windows with the current numbers of active, recovered and confirmed cases, as the number of deaths (Figure 30 – Blue Rectangle). The map visualization shows the cumulative incidence, or number of new confirmed cases per municipality since the beginning of pandemic per 100 000 habitants, by using choropleth map with defined administrative limits (Figure 30 – Green Rectangle). The bottom right side displays the evolution of temporal COVID-19 indicators, where the time interval can be chosen using available at the top zoom option, and the indicator is selected by clicking on available options at the bottom (Figure 30 – Red Rectangle). The top right section is divided by two parts: the right one represents the total number of COVID-19 confirmed cases since the beginning of the pandemic, aggregated by certain category where indicators can be selected by using the carrousel at the bottom (Figure 30 – Yellow Rectangle). The left section (Figure 30 – Purple Rectangle) displays some information about distribution of cases by different ARS.

Afterwards, the analysis of data visualization formats utilized in this dashboard for the case study indicators needed to be done in order to evaluate their appropriateness. First of all, the map used for representing cumulative incidence was identified as choropleth, with a constant risk within administrative unit, which is compliant with results of case study for questions 11 and 12. Consequently, this data visualization format can be considerate adequate for the use on DGS dashboard.

Next, the temporal COVID-19 indicators were investigated in order to understand if DGS COVID-19 dashboard utilized data visualizations compatible with case study results. The indicators “Daily number of new confirmed COVID-19 cases” and “Cumulative number of new confirmed COVID-19 cases” were represented by area chart and line chart, respectively, for a longer time interval. The first indicator has been showed not being compliant with case study results (Figures 18 and 20), which suggested line chart as a most adequate data visualization format. Additionally, after the zoom-in action the data visualization didn’t change to column chart, most preferred for the shorter intervals of time. Therefore, this can be corrected by implementing the logic where for the certain data interval after the zoom-in action the chart type changes from the line to column chart, however it might not be supported by all the platforms and can be complicated to perform.

Afterwards, the section regarding the total number of confirmed COVID-19 cases since the beginning of the pandemic was analyzed. This indicator aggregated by sex was correctly represented by the pie chart, preferred by the majority of Delphi participants, however, the bar chart reflected the number of cases by age group, not being consistent with the results of Delphi study (Figure 24). Therefore, this data visualization format should be replaced by column chart, that was the most preferred one by Delphi participants.

The confirmed cases per ARS were represented by the table format, however, the most voted format determined by case study was a column chart, therefore, this alteration should be implemented in the platform. Regarding the indicator “Number of confirmed cases by sex and age group”, the DGS dashboard utilized grouped column chart, while case study results (Figure 27) showed a strong preference (79% of voters) for the two-sided bar chart.

Therefore, some alterations can be performed with DGS dashboard, considering the results of the case study, based on the previously developed approach. These alterations can improve the communicability and efficiency of platform by integrating data visualizations, previously validated not only by acquired knowledge, but also by involvement of final users in their selection. Despite of the fact that the case study didn’t cover all indicators included in DGS dashboard, it was possible to exemplify the possible application of results for some of the indicators.

b) Implementation

As a consequence of the previously performed analysis, it was decided to develop renewed DGS COVID-19 dashboard, that has been adjusted to Delphi results. This dashboard implemented the alternatives suggested above, considering all the possible improvements acquired in this study.

Since most of the indicators used in Delphi process were implemented in the original DGS Dashboard, they were maintained for the further design development. The only indicator of Delphi process, not presented in original DGS dashboard, “Daily number of new confirmed COVID-19 cases by region” was added due to its high frequency in public COVID-19 dashboards, this way increasing the usefulness of the developed platform.

The platform selected for the dashboard development was ArcGIS Online, cloud-based software to create and share interactive web maps. Additionally, it allows to upload data from different sources and represent it using different data visualizations on a single screen, by utilizing ArcGIS Dashboards, one of the ArcGIS extensions. This platform was selected due to its user-friendly interface, easy integration of different data formats, a robust package of data visualizations, and the possibility of map creation and its posterior implementation in dashboard.

A main data source used in the developed dashboard was the repository GitHub named "Dados relativos à pandemia COVID-19 em Portugal", based on the daily updates of COVID-19 data presented by DGS. The data is organized in .csv files, each corresponding to different COVID-19 topics and following certain structure. For this implementation, two files were used: data.csv, which contains the data extracted from DGS dashboard and daily DGS reports, and data_concelhos_incidencia.csv, which contains the data from the accumulated confirmed cases of the 14 days prior to the reporting date proportional to 100k inhabitants.

The first file, data.csv, reflects the information about daily confirmed COVID-19, aggregated by ARS, sex and/or age, and the information about the deaths, vaccinated and tested. However, for this implementation, only the pertinent indicators were extracted from the database, uploaded to ArcGIS and saved for future operations. For map visualization, another file, data_concelhos_incidencia.csv, was analyzed, and the data for the last registered day were extracted. Afterwards, the data upload to ArcGIS Online was performed, and the map layer was created by using ArcGIS Map extension.

The dashboard development using ArcGIS Dashboard was then proceeded. The overall visual appearance was decided to be similar to the original DGS Dashboard, using dark theme with the analogous color palette. Moreover, the localization of blocks was also organized in the identical way (Figure 31), where the section with static indicators were replicated (Figure 31 – Blue Rectangle). The map layer, previously developed in the separate extension, was integrated as a choropleth map with administrative limits, in the middle of the dashboard (Figure 31 - Green Rectangle). Afterwards, the temporal indicators were implemented in the bottom right of the dashboard, using data visualization formats consistent with the Delphi results by integrating the uploaded data from data.csv (Figure 31 - Red Rectangle). The scroll bar was incorporated within visualizations to perform zoom action; however, ArcGIS presented certain limitations due to the impossibility to change data visualization according to the amplified time interval. Therefore, the data visualization formats for shorter period of time (one week) were implemented separately.

For total number of confirmed cases aggregated by ARS, sex and/or age group the data visualizations were implemented at top right of dashboard, using the previously uploaded information (Figure 31 – Yellow Rectangle). However, since ArcGIS doesn't provide two-sided bar chart, the indicator "Total number of confirmed COVID-19 cases by age group and sex" could not be integrated in final dashboard.

The upgraded DGS dashboard can be accessed from the link https://www.arcgis.com/apps/dashboards/4a6782bb9c5a4234bbb251ed9c461471. It is noteworthy to mention that the data presented on dashboard were collected on the day 29 of September of 2021, so the information can be outdated. However, ArcGIS Online provides maintenance of information by offering the possibility of updating old information, despite being time-consuming.

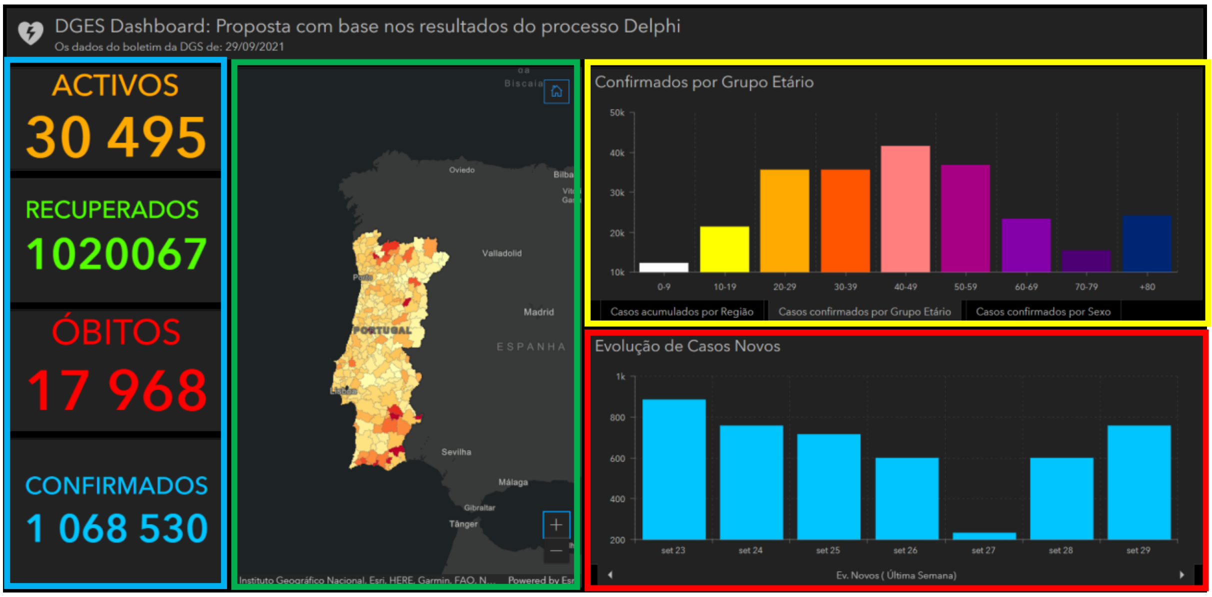


Figure 31. Adjusted DGS COVID-19 Dashboard.

7 Discussion

This section characterizes the methods and results elaborated in this thesis. Namely, the developed novel approach for data visualization's selection is discussed, and the choices made for the implementation are debated as well as its implications. In a later stage the results of case study are analyzed and discussed. To finalize, the contribution of this work is acknowledged, its overall limitations and future developments are considered.

7.1 Novel approach

The proposed approach was developed to tackle the limitations presented in dashboard literature, namely, the lack of well-constructed approach to select the appropriate visualizations, taking into account not only the theory-based evidence, but also the perception of final users. For instance, the developed approach reinvents the selection process in a more competent and effective manner, offering assistance for data visualization selection. This way, the indicators' data structure is taken as a starting point, afterwards mapping it to the corresponding data visualization format, also considering the communication purpose that the final visualization pretends to transmit. For that reason, the decision table was developed to select the set of most appropriate data visualization formats according to the metrics described above.

However, previously described process alone does not provide a unique solution, at the same time not considering user perspectives and perceptions. Henceforward, the Delphi process ensures the usability and acceptability of selected visualizations by capturing heterogeneous users' opinions and reaching an agreement in a legitimate and dynamic way. Delphi was considered more appropriate for the dashboard design, where the heterogeneity of users implies high number of experts, and high integration and inclusion of different types of information. Therefore, in this process the pre-set of data visualization formats is proposed to the participants, and the perceptions and preferences of users are investigated by asking the preferred data visualization format for indicator in question.

7.1.1 Implementation

Developing a novel approach to improve data visualization selection process in dashboards culminated in its direct application, where it was implemented in the context of COVID-19 pandemic. Many web-based dashboards have been created to assist in the dissemination of key pandemic information to the public and its understanding of the COVID-19 data. To surpass challenges presented by public web-based COVID-19 dashboards, the developed approach was implemented in two steps: first, the pre-set of data visualization formats was defined for the list of selected COVID-19 indicators, considering their data structure and communication purpose, and afterwards the Delphi process was developed to select the most preferred data visualization for each indicator in question.

Firstly, in order to apply the developed approach, the set of most common COVID-19 indicators was predefined by performing the analysis of study, which explores the characteristics of 158 public web-based COVID-19 dashboards. However, the legitimacy of this indicators was not verified due to the lack of their validation by applying separate methodology, which was decided to be omitted since the focus of this study is not related with indicator's selection. At the same time, the number of selected indicators was restricted with the purpose of not provoking fatigue in the posterior participatory process.

The selected indicators were classified by identifying their data structure, namely the number of variables and their attribute type, and final communication purpose. Finally, the pre-set of data visualization formats was defined by using decision table, developed in this study. The decision table, mostly oriented to the public health dashboards, presented several limitations since the data visualization formats used were the most common ones, requiring the simplicity of indicator's structure.

This study was developed within the scope of the SCOPE (Spatial Data Science Services for COVID-19 Pandemic) project, which provides daily updates on maps with local averages of infection risk, without considering the administrative limits. Therefore, it was important to understand if this type of maps can be useful to transmit information to general population, comparing it to the usual maps with constant infection risk value within each area.

Afterwards, the Delphi process was implemented to identify the most preferred visualization formats among the public. The web-based format was chosen to reduce fatigue and attrition between rounds, at the same time lowering costs and facilitate data processing. SurveyHero platform was selected to be the one to execute Delphi process, however, some difficulties were encountered during the implementation, since the platform was not adapted for Delphi processes but only for one-round surveys. Therefore, the second-round questionnaires were adjusted for each participant and sent individually, without automatization of process. An *a priori* criterion of two rounds in this Delphi process was established not only to ensure against participant fatigue but also to guarantee results are meaningful. However, one should assume that, in this case, the consensus may be not achieved.

7.2 Results

In Chapter 6, the results from the different rounds of Delphi were presented. In the first round, 72 participants from total 101 totally completed the questionnaire. Specifically, the majority of participants didn't fill their e-mail at the end of questionnaire in order to proceed with the process. This could be explained by participants not feeling comfortable to share their personal information, besides the guarantee of anonymity and of the use of email only for study purpose. 47 participants decided to participate in second round of the Delphi study, with the response rate of 65%. The low response rate can be explained by low interest, increase of fatigue between rounds and the fact that the general population are not familiar with two-round Delphi questionnaires.

The first part of questionnaire explored the data visualization formats of temporal COVID-19 indicators. After two rounds, the preferred data visualization format for number of new confirmed COVID-19 cases, represented daily, cumulatively or categorized by regions for longer time intervals (three months), was a line chart. Additionally, for shorter time interval (one week), the most voted data visualization format for same indicators was a column chart. Therefore, results suggest that a column chart is more informative to depict variations for shorter periods of times, whereas a line chart is better in displaying overall evolution picture. Moreover, these results highlight the importance of using different visualization formats for different screen amplifications.

For the second part of questionnaire total number of confirmed cases since the beginning of the COVID-19 pandemic were investigated, aggregated by region, sex and/or age group. For indicator “Total number of confirmed cases of COVID-19 by age group” the results were inconclusive due to the switch of modes between the rounds, and correspondent percentage agreement of 40,4% for column chart were relatively low to deduce any conclusions. As has been already mentioned, this inconsistency between rounds can be related with the comment of the participant, that advised not to use pie chart with high number of categories, provoking the decrease of the votes for this visualization in the second round. This strong influence of the comment for final results reenforces the usefulness of Delphi due to the interaction of different participants between rounds.

For the total number of confirmed cases of COVID-19 by sex, the most voted data visualization format was a pie chart, meanwhile for aggregation by regions a column chart was the preferred option. Finally, for the last question of the second part of the questionnaire, the indicator “Total number of confirmed cases by age group and sex” was explored, where a two-sided bar chart was the most preferred data visualization format.

Lastly, for map visualization, a choropleth map was the preferred option to represent the number of new confirmed cases of COVID-19 per municipality since the beginning of the pandemic. Moreover, for last question regarding the project SCOPE, participants preferred the maps with constant infection risk within the administrative unit, suggesting that the general population has some difficulties to interpret SCOPE maps, despite their higher spatial resolution.

Regarding overall results, they were considered stable, and it was possible to observe an evolution of group judgements towards a higher level of agreement along the rounds. Considering that a group majority (consensus) was established when at least 50% of the participants selected certain data visualization, 8 from 12 questions achieved consensus regarding the selection of data visualization format, and 3 of the 4 remaining questions reached the percentage value close to 50% (46,8%). As have been already mentioned, only the indicator “Total number of confirmed cases of COVID-19 by age group” showed inconclusive results and switch in modes.

7.2.1 Implications of the Delphi results for dashboard design

The direct application of case study results was performed to DGS COVID-19 Dashboard, the platform developed in Portugal to transmit the general COVID-19 information to the population. The dashboard was analyzed in terms of indicators and correspondent data visualization formats used, and afterwards the inconsistencies with case study results were identified. The alternatives then were proposed, in order to improve transmissibility of relevant information to the population and communicability of platform. Finally, the updated DGS COVID-19 was developed by using ArcGIS software, integrating the case study results. The implementation was performed without major difficulties, only presenting some limitations regarding zoom interaction and absence of two-sided bar chart. This way, the DGS dashboard was improved by including data visualizations that have been previously verified by developed approach.

7.3 Advantages of the proposed approach

The developed approach provides guideline for identification of the most adequate data visualization formats, based on the previous studies in the field of information visualization, now applied to the dashboard design. This approach provides a step-by-step solution for finding the best visualization type for a given data set and communication purpose. Moreover, it covers most common data visualization formats, mostly used in public health dashboards, that are easily perceivable by the general public.

However, the developed approach not only allows to select the set of most appropriate data visualization formats from theoretical point of view, but also test the cognitive perception of possible dashboard users and understand the preferences regarding certain visualizations. The implementation of Delphi process ensured the acceptability of the selected data visualization format by considering not just participants' differing interests, but also establishing a valid and dynamic consensus on various judgments.

Overall, the proposed approach surpasses the challenges presented by the modern public health dashboards by providing guideline for data visualization selection with implemented Delphi process.

7.4 Limitations

During the development of this approach several challenges were encountered. Firstly, the limited number of studies was found reflecting on visualizations, since the research on dashboards and data visualization principles is still in its early stages, with just a few publications starting to appear recently. Therefore, the approach for the visualization selection developed in this study was not previously applied specifically to the dashboard context, and thus, requires validation regarding this point. On another hand, the developed approach required the abstraction from individual graphical expressions, such as shape, color or position, and focused on the underlying data structure. Consequently, it did not cover the aesthetics of visualizations, so this topic requires further investigation.

As has been already mentioned, the indicators' selection stage must precede the proceedings regarding the identification of visual tools. This step requires highly complex process and involves the implementation of separate methodology; therefore, it was not implemented in this approach, only focusing on data visualization selection.

In terms of the Delphi process, it represents several limitations, such as it may be influenced by participant's bias and tendency to eliminate extreme positions to achieve central consensus [143]. Additionally, the lack of empirical rules or guidelines for definition of level of consensus creates additional problems while developing the process [143]. Moreover, the nominal scale, utilized in this process, is less flexible for the statistical analysis, allowing to perform only simple statistical tests.

Several limitations regarding implementation of some visual formats in dashboard platform were encountered, appointing to the constraints imposed by the software.

8 Conclusions

Although the rise of data visualization tools has raised concerns about their potential shortcomings, little is known about the various methodologies involved in making an efficient dashboard. The lack of support regarding visualization's selection contributes to development of dashboards with highly subjective choice of data visualization formats, mostly focused on the developer vision and perspective.

The main goal of this thesis was to explore methods to inform the design of data visualization tools and select the data visualization formats, that can be incorporated in dashboards and used to transmit relevant information about COVID-19 pandemic. With the work presented during the thesis, it is possible to see that this objective was clearly fulfilled by developing a novel approach, that considers not only the theory-based evidence, but also the preferences and views of general population for distinct data visualization formats. For that purpose, the literature research was developed in order to establish guideline for selection of appropriate data visualization formats for certain indicators, and afterwards, Delphi process was applied to choose the preferred visualization format between the options of pre-selected set.

The outcomes of case study using developed approach showed that its implementation is working without any problems, and it can be useful for researchers working within the public health area. A consensus was achieved for the majority of statements, with relatively high stability of the process, and it was possible to observe an evolution of group judgements towards a higher level of agreement along the rounds. The posterior application of case study results to already developed DGS COVID-19 dashboard showed that this type of platforms can be improved by implementing such approach for their development.

However, being a novel approach of this nature, it has a number of aspects in which it may be improved or even maintained to be more complete or autonomous. Therefore, it can be viewed as a model for future advancements, such as the ones described below.

Regarding future work to be developed in the sequence of this thesis: first of all, the suggested future work is related to correction of discovered limitations, namely, the validation of developed approach by experts. Moreover, this approach can be completed by investigating in more detail the graphical expressions of data visualization formats and appropriate user interaction steps for each visual format. On the other hand, the developed approach did not cover all existing data visualization formats, therefore the more profound analysis is needed to explore other visualization formats.

As has been already mentioned, a dashboard development includes different stages, therefore it would be considerable to integrate the developed approach into the more complex one, which covers all stages of dashboard development, e.g., including the methodology for key indicators' selection.

Finally, this study gives the future possibility to develop powerful data visualization tool that integrates data visualization formats of the selected health indicators in a concise, efficient and visually effective way, taking into account the suggested approach. This tool may be used by government officials to transmit relevant information and new public policies and optimize the use of public resources.

References

[1] Q. Li, “Overview of Data Visualization,” in *Embodying Data: Chinese Aesthetics, Interactive Visualization and Gaming Technologies*, Singapore: Springer Singapore, 2020, pp. 17–47.

[2] R. A. Vila, E. Estevez, and P. R. Fillottrani, “The design and use of dashboards for driving decision-making in the public sector,” in *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, 2018, vol. 7, Accessed: Mar. 06, 2021. [Online]. Available: https://doi.org/10.1145/3209415.3209467.

[3] M. Prokofieva, “Using dashboards and data visualizations in teaching accounting,” *Educ. Inf. Technol. 2021*, pp. 1–17, Apr. 2021, doi: 10.1007/S10639-021-10525-Z.

[4] G. Sedrakyan, E. Mannens, and K. Verbert, “Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts,” 2018, doi: 10.1016/j.jvlc.2018.11.002.

[5] L. Albó, J. Barria-Pineda, P. Brusilovsky, and & Davinia Hernández-Leo, “Knowledge-Based Design Analytics for Authoring Courses with Smart Learning Content,” doi: 10.1007/s40593-021-00253-3.

[6] G. A. Miller, “The magical number seven, plus or minus two: some limits on our capacity for processing information,” *Psychol. Rev.*, vol. 63, no. 2, pp. 81–97, Mar. 1956, doi: 10.1037/h0043158.

[7] O. M. Yigitbasioglu and O. Velcu, “A review of dashboards in performance management: Implications for design and research,” *Int. J. Account. Inf. Syst.*, vol. 13, no. 1, pp. 41–59, Mar. 2012, doi: 10.1016/j.accinf.2011.08.002.

[8] A. Cairo, *The Functional Art: An Introduction to Information Graphics and Visualization*, 1st ed. USA: New Riders Publishing, 2012.

[9] A. Kirk, “Data visualization : a successful design process ; a structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively,” *European University Institute*. Packt Pub, Birmingham, UK, 2013.

[10] Stephen Few, *Information dashboard design: The effective visual communication of data*, 1st ed. 2006.

[11] R. Spence, “Information Visualization.” Springer International Publishing, 2014, doi: 10.1007/978-3-319-07341-5.

[12] H. Whitney. and H. Whitney., *Data Insights: New Ways to Visualize and Make Sense of Data*.

Morgan Kaufmann Publishers, 2013.

[13] J. Gulden, “A Description Framework for Data Visualizations in Enterprise Information Systems,” *Proc. - IEEE Int. Enterp. Distrib. Object Comput. Work. EDOCW*, vol. 2015-Novem, pp. 68–73, 2015, doi: 10.1109/EDOC.2015.20.

[14] K. Koskinen and C. Bonina, *Information and Communication Technologies for Development: Strengthening Southern-Driven Cooperation as a Catalyst for ICT4D*, no. April 2020. 2019.

[15] P. Kostkova, S. Garbin, J. Moser, and W. Pan, “Integration and Visualization Public Health Dashboard: The medi+board Pilot Project,” doi: 10.1145/2567948.2579276.

[16] M. Elshehaly *et al.*, “QualDash: Adaptable generation of visualisation dashboards for healthcare quality improvement,” *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 689–699, Feb. 2021, doi: 10.1109/TVCG.2020.3030424.

[17] R. A. Dixit *et al.*, “Rapid development of visualization dashboards to enhance situation awareness of COVID-19 telehealth initiatives at a multihospital healthcare system,” *J. Am. Med. Informatics Assoc.*, vol. 27, no. 9, pp. 1456–1461, Sep. 2020, doi: 10.1093/JAMIA/OCAA161.

[18] P. Ruchikachorn, “Applying uncommon visualizations to government dashboards,” *VISIGRAPP 2021 - Proc. 16th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl.*, vol. 3, no. Visigrapp, pp. 203–209, 2021, doi: 10.5220/0010298102030209.

[19] R. Barcellos, J. Viterbo, L. Miranda, F. Bernardini, C. Maciel, and D. Trevisan, “Transparency in practice: Using visualization to enhance the interpretability of open data,” *ACM Int. Conf. Proceeding Ser.*, vol. Part F1282, pp. 139–148, 2017, doi: 10.1145/3085228.3085294.

[20] M. Dastani, “The role of visual perception in data visualization,” *J. Vis. Lang. Comput.*, vol. 13, no. 6, pp. 601–622, 2002, doi: 10.1006/jvlc.2002.0235.

[21] T. Munzner, *Visualization Analysis & Design Tamara Munzner A K Peters Visualization Series Illustrations by Eamonn Maguire*. 2013.

[22] J. Bertin, *Semiology of graphics : diagrams, networks, maps*. Redlands, Calif.: Redlands, Calif. : ESRI Press : Distributed by Ingram Publisher Services, 2010.

[23] E. Tufte, “The Visual Display of Quantitative Information (1983),” *Diagrammatik-Reader*, pp. 219–230, Oct. 2016, doi: 10.1515/9783050093833-035/HTML.

[24] J. I. Helfman and J. H. Goldberg, “Selecting the Best Graph Based on Data, Tasks, and User Roles,” *Usability Prof. Assoc. UPA Conf. Patterns Blueprints Usability, Austin, TX*, no. July, 2007.

[25] S. S. J. Alhadad, “Visualizing Data to Support Judgement, Inference, and Decision Making in

Learning Analytics: Insights from Cognitive Psychology and Visualization Science," *J. Learn. Anal.*, vol. 5, no. 2, Aug. 2018, doi: 10.18608/jla.2018.52.5.

[26] D. Dowding and J. A. Merrill, "The Development of Heuristics for Evaluation of Dashboard Visualizations," *Appl. Clin. Inform.*, vol. 9, no. 3, pp. 511–518, Jul. 2018, doi: 10.1055/s-0038-1666842.

[27] W. Bibliothek and by Johannes Dostal, "Designing a Dashboard for Student Information Systems Media and Human-Centered Computing." Accessed: Apr. 10, 2021. [Online]. Available: www.tuwien.ac.at.

[28] S. Al-Hajj, I. Pike, and B. Fisher, "Interactive Dashboards: Using Visual Analytics for knowledge Transfer and Decision Support."

[29] S. Jiang *et al.*, "Healthcare Data Visualization: Geospatial and Temporal Integration," doi: 10.5220/0005714002120219.

[30] L. Meloncon and E. Warner, "Data visualizations: A literature review and opportunities for technical and professional communication," Aug. 2017, doi: 10.1109/IPCC.2017.8013960.

[31] D. Gotz and D. Borland, "Data-Driven Healthcare: Challenges and Opportunities for Interactive Visualization," 2016.

[32] D. Dowding *et al.*, "Dashboards for improving patient care: Review of the literature," *Int. J. Med. Inform.*, vol. 84, no. 2, pp. 87–100, 2015, doi: 10.1016/j.ijmedinf.2014.10.001.

[33] S.-H. Zhang, Y. Cai, and J. Li, "Visualization of COVID-19 spread based on spread and extinction indexes," vol. 63, p. 3, 2020, doi: 10.1007/s11432-020-2828-1.

[34] N. Marques Da Costa, N. Mileu, A. Alves, S. Portela, and A. Crooks, "Dashboard COMPRIME_COMPRI_MOv: Multiscalar Spatio-Temporal Monitoring of the COVID-19 Pandemic in Portugal Academic Editors: Carlos Filipe Da," 2021, doi: 10.3390/fi13020045.

[35] "COVID Live Update: 150,336,400 Cases and 3,166,902 Deaths from the Coronavirus - Worldometer." https://www.worldometers.info/coronavirus/ (accessed Apr. 29, 2021).

[36] D. Ivanković *et al.*, "Features constituting actionable COVID-19 dashboards: Descriptive assessment and expert appraisal of 158 public web-based COVID-19 dashboards," *J. Med. Internet Res.*, vol. 23, no. 2, 2021, doi: 10.2196/25682.

[37] C. K. Leung, Y. Chen, C. S. H. Hoi, S. Shang, Y. Wen, and A. Cuzzocrea, "Big Data Visualization and Visual Analytics of COVID-19 Data," no. Iv, pp. 415–420, 2021, doi: 10.1109/iv51561.2020.00073.

[38] M. D. Verhagen, D. M. Brazel, J. B. Dowd, I. Kashnitsky, I. Kashnitsky, and M. C. Mills, "Forecasting spatial, socioeconomic and demographic variation in COVID-19 health care demand in England and Wales," *BMC Med.*, vol. 18, no. 1, p. 203, Jun. 2020, doi: 10.1186/s12916-020-01646-2.

[39] J. Budd *et al.*, "Digital technologies in the public-health response to COVID-19," *Nat. Med.*, vol. 26, no. 8, pp. 1183–1192, Aug. 2020, doi: 10.1038/s41591-020-1011-4.

[40] "WHO Coronavirus (COVID-19) Dashboard | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data." https://covid19.who.int/ (accessed Apr. 29, 2021).

[41] "COVID-19 Map - Johns Hopkins Coronavirus Resource Center." https://coronavirus.jhu.edu/map.html (accessed Apr. 29, 2021).

[42] "COVID-19 | European Centre for Disease Prevention and Control." https://qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html#global-overview-tab (accessed Apr. 29, 2021).

[43] "Ponto de Situação Atual em Portugal - COVID-19." https://covid19.min-saude.pt/ponto-de-situacao-atual-em-portugal/ (accessed Apr. 29, 2021).

[44] C. Engledowl and T. Weiland, "Data (Mis)representation and COVID-19: Leveraging Misleading Data Visualizations For Developing Statistical Literacy Across Grades 6–16," 2021, doi: 10.1080/26939169.2021.1915215.

[45] S. Barone, A. Chakhunashvili, and A. Comelli, "Building a statistical surveillance dashboard for COVID-19 infection worldwide," *Qual. Eng.*, vol. 32, no. 4, pp. 754–763, Oct. 2020, doi: 10.1080/08982112.2020.1770791.

[46] M. Alcíbar, "Information visualisation as a resource for popularising the technical-biomedical aspects of the last Ebola virus epidemic: The case of the Spanish reference press," *Public Underst. Sci.*, vol. 27, no. 3, pp. 365–381, Apr. 2018, doi: 10.1177/0963662517702047.

[47] M. Khan and S. Shah Khan, "Data and Information Visualization Methods, and Interactive Mechanisms: A Survey," *Artic. Int. J. Comput. Appl. Int. J. Comput. Appl.*, vol. 34, no. 1, pp. 975–8887, 2011, [Online]. Available: https://www.researchgate.net/publication/264623537.

[48] C. Akujuobi, M. N. O. Sadiku, A. E. Shadare, S. M. Musa, C. M. Akujuobi, and R. G. Perry, "Data visualization," *Int. J. Eng. Res. Adv. Technol.*, Accessed: Mar. 19, 2021. [Online]. Available: https://www.researchgate.net/publication/311597028.

[49] M. P. Cota, M. D. Rodriguez, M. R. Gonzalez-Castro, and R. M. M. Goncalves, "Massive data visualization analysis analysis of current visualization techniques and main challenges for the future," Jul. 2017, doi: 10.23919/CISTI.2017.7975704.

[50] S. Soma Ajibade and A. Adediran, “An Overview of Big Data Visualization Techniques in Data Mining.” Accessed: Mar. 19, 2021. [Online]. Available: www.researchpublish.com.

[51] J. Gulden, D. Van Der Linden, and B. Aysolmaz, “A Research Agenda on Visualizations in Information Systems Engineering Privacy impact of technology for animals View project A Research Agenda on Visualizations in Information Systems Engineering,” doi: 10.5220/0005898702340240.

[52] J. Gulden and H. A. Reijers, “Toward Advanced Visualization Techniques for Conceptual Modeling.”

[53] T. Soukup and I. Davidson, *Visual Data Mining: Techniques and Tools for Data Visualization and Mining Visual Data Mining: Techniques and Tools for Data Visualization and Mining by Tom Soukup and Ian Davidson John Wiley & Sons ?2002 (382*, vol. 2002. 2002.

[54] O. Ola and K. Sedig, “Beyond simple charts: Design of visualizations for big health data,” *Online J. Public Health Inform.*, vol. 8, no. 3, Dec. 2016, doi: 10.5210/ojphi.v8i3.7100.

[55] M. J. Kraak, “The role of the map in a Web-GIS environment,” *J. Geogr. Syst.*, vol. 6, no. 2, pp. 83–93, 2004, doi: 10.1007/s10109-004-0127-2.

[56] M. J. P. Rytkönen, “International Journal of Circumpolar Health Not all maps are equal: GIS and spatial analysis in epidemiology,” *Int. J. Circumpolar Health*, vol. 63, no. 1, pp. 9–24, 2004, doi: 10.3402/ijch.v63i1.17542.

[57] F. B. Mocnik, P. Raposo, W. Feringa, M. J. Kraak, and B. Köbben, “Epidemics and pandemics in maps–the case of COVID-19,” *J. Maps*, vol. 16, no. 1, pp. 144–152, Jan. 2020, doi: 10.1080/17445647.2020.1776646.

[58] L. A. Waller and C. A. Gotway, *Applied Spatial Statistics for Public Health Data*, vol. 100, no. 470. 2005.

[59] M. Raubal, “Spatial data science for sustainable mobility,” *J. Spat. Inf. Sci.*, vol. 20, pp. 109–114, 2020, doi: 10.3929/ETHZ-B-000423945.

[60] A. Fioravanti, G. Novembri, and F. Rossini, “Improving Proactive Collaborative Design Through the Integration of BIM and Agent-Based Simulations,” 2017.

[61] B. Hansoti, “Business Intelligence Dashboard in Decision Making.” Accessed: Mar. 21, 2021. [Online]. Available: http://docs.lib.purdue.edu/techdirprojhttp://docs.lib.purdue.edu/techdirproj/15.

[62] “Visual Business Intelligence – There’s Nothing Mere About Semantics.” https://www.perceptualedge.com/blog/?p=2793 (accessed Mar. 07, 2021).

[63] S. Wexler, J. Shaffer, and A. Cotgreave, *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios* | *Wiley*. Hoboken:Wiley,2017., 2017.

[64] A. Sarikaya, M. Correll, L. Bartram, M. Tory, and D. Fisher, "What do we talk about when we talk about dashboards?," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 682–692, 2019, doi: 10.1109/TVCG.2018.2864903.

[65] R. Alharbey, "Predictive analytics dashboard for monitoring patients in advanced stages of COPD," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, Mar. 2016, vol. 2016-March, pp. 3455–3461, doi: 10.1109/HICSS.2016.431.

[66] K. Pauwels *et al.*, "Dashboards as a Service," *J. Serv. Res.*, vol. 12, no. 2, pp. 175–189, Nov. 2009, doi: 10.1177/1094670509344213.

[67] B. R. Lea and F. F. H. Nah, "Usability of performance dashboards, usefulness of operational and tactical support, and quality of strategic support: A research framework," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Jul. 2013, vol. 8017 LNCS, no. PART 2, pp. 116–123, doi: 10.1007/978-3-642-39215-3_14.

[68] H. Lempinen, "Constructing a design framework for performance dashboards," in *Lecture Notes in Business Information Processing*, 2012, vol. 124 LNBIP, pp. 109–130, doi: 10.1007/978-3-642-32270-9_7.

[69] M. Khalifa and P. Khalid, "Developing strategic health care key performance indicators: A case study on a tertiary care hospital," in *Procedia Computer Science*, 2015, vol. 63, pp. 459–466, doi: 10.1016/j.procs.2015.08.368.

[70] S. Gray, O. O'Brien, and S. Hügel, "Collecting and visualizing real-time urban data through city dashboards," *Built Environ.*, vol. 42, no. 3, pp. 498–509, Oct. 2016, doi: 10.2148/benv.42.3.498.

[71] F. Purnomo, Meyliana, and H. Prabowo, "Smart city indicators: A systematic literature review," *J. Telecommun. Electron. Comput. Eng.*, vol. 8, no. 3, pp. 161–164, 2016.

[72] C. Jing, M. Du, S. Li, and S. Liu, "Geospatial Dashboards for Monitoring Smart City Performance," *Sustainability*, vol. 11, no. 20, p. 5648, Oct. 2019, doi: 10.3390/su11205648.

[73] M. Riphagen, "Supporting Reflective Learning For Daily Activities Using An Interactive Dashboard." Accessed: Mar. 28, 2021. [Online]. Available: https://jawbone.com/up.

[74] G. B. Guy and C. J. Kibert, "Developing indicators of sustainability: US experience," *Build. Res. Inf.*, vol. 26, no. 1, pp. 39–45, Jan. 1998, doi: 10.1080/096132198370092.

[75] J. J. Mitchell and A. J. Ryder, "Developing and Using Dashboard Indicators in Student Affairs

Assessment," no. 142, pp. 71–81, 2013, doi: 10.1002/ss.

[76] D. Concannon, K. Herbst, ; Ed Manley, and E. Manley, "Developing a Data Dashboard Framework for Population Health Surveillance: Widening Access to Clinical Trial Findings," doi: 10.2196/11342.

[77] T. M. Orlando and W. D. Sunindyo, "Designing Dashboard Visualization for Heterogeneous Stakeholders," *2017 Int. Conf. Data Softw. Eng.*, pp. 1–6, 2017.

[78] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations."

[79] D. Brown, "Good Practice Guidelines for Indicator Development and Reporting." Accessed: Mar. 12, 2021. [Online]. Available: www.stats.govt.nz.

[80] S. Stehle and R. Kitchin, "Real-time and archival data visualisation techniques in city dashboards," *Int. J. Geogr. Inf. Sci.*, vol. 34, no. 2, pp. 344–366, Feb. 2020, doi: 10.1080/13658816.2019.1594823.

[81] M. Derntl, S. Erdtmann, and R. Klamma, "An embeddable dashboard for widget-based visual analytics on scientific communities," in *ACM International Conference Proceeding Series*, 2012, p. 1, doi: 10.1145/2362456.2362486.

[82] M. Abd-Elfattah, T. Alghamdi, and E. Amer, "Dashboard tecnhology based solution to decision making," 2014. Accessed: Mar. 11, 2021. [Online]. Available: www.tjprc.org.

[83] H. J. Watson *et al.*, "BI-based Organizations 4 Advances in Predictive Modeling: How 17 In-Database Analytics Will Evolve to Change the Game Enabling Agile BI with a Compressed 29 Flat Files Architecture BI Experts' Perspective: Pervasive BI 36 BI and Sentiment Analysis 41 Dash," 2010. Accessed: Mar. 21, 2021. [Online]. Available: www.tdwi.org/onsite.

[84] T. Sinuff *et al.*, "Improving end-of-life communication and decision making: The development of a conceptual framework and quality indicators," *J. Pain Symptom Manage.*, vol. 49, no. 6, pp. 1070–1080, Jun. 2015, doi: 10.1016/j.jpainsymman.2014.12.007.

[85] M. Schomaker, "Development of environmental indicators in UNEP," 1997. http://www.fao.org/3/W4745E/w4745e07.htm (accessed Apr. 01, 2021).

[86] T. D. Abel and M. A. Evans, "Cross-disciplinary Participatory & Contextual Design Research: Creating a Teacher Dashboard Application."

[87] "Use of Dashboards in Government." Accessed: Mar. 18, 2021. [Online]. Available: www.businessofgovernment.org.

[88] A. A. Rahman, Y. B. Adamu, and P. Harun, “Review on dashboard application from managerial perspective,” *Int. Conf. Res. Innov. Inf. Syst. ICRIIS*, 2017, doi: 10.1109/ICRIIS.2017.8002461.

[89] S. C. Buttigieg, A. Pace, and C. Rathert, “Hospital performance dashboards : a literature review Abstract,” 2017, doi: 10.1108/JHOM-04-2017-0088.

[90] V. S. Smith, “Data Dashboard as Evaluation and Research Communication Tool,” no. 114, pp. 63–81, 2013, doi: 10.1002/ev.

[91] D. P. Tegarden, “Business Information Visualization,” *Commun. Assoc. Inf. Syst.*, vol. 1, 1999, doi: 10.17705/1CAIS.00104.

[92] M. I. Mashinchi, A. Ojo, and F. J. Sullivan, *Towards a theoretical model of dashboard acceptance and use in healthcare domain*. .

[93] K. Daley, J. Richardson, I. James, A. Chambers, and D. Corbett, “Clinical dashboard: Use in older adult mental health wards,” *Psychiatrist*, vol. 37, no. 3, pp. 85–88, Mar. 2013, doi: 10.1192/pb.bp.111.035899.

[94] M. Isazad Mashinchi, A. Ojo, and F. J. Sullivan, “Investigating Analytics Dashboards’ Support for the Value-based Healthcare Delivery Model,” *Proc. 53rd Hawaii Int. Conf. Syst. Sci.*, pp. 3659–3668, 2020, doi: 10.24251/hicss.2020.448.

[95] K. W. Park, D. Smaltz, D. McFadden, and W. Souba, “The operating room dashboard,” *Journal of Surgical Research*, vol. 164, no. 2. Elsevier, pp. 294–300, Dec. 01, 2010, doi: 10.1016/j.jss.2009.09.011.

[96] B. McLeod *et al.*, “Matching Capacity to Demand: A Regional Dashboard Reduces Ambulance Avoidance and Improves Accessibility of Receiving Hospitals,” *Acad. Emerg. Med.*, vol. 17, no. 12, pp. 1383–1389, Dec. 2010, doi: 10.1111/j.1553-2712.2010.00928.x.

[97] V. Zaydfudim *et al.*, “Implementation of a real-time compliance dashboard to help reduce SICU ventilator-associated pneumonia with the ventilator bundle,” *Arch. Surg.*, vol. 144, no. 7, pp. 656–661, Jul. 2009, doi: 10.1001/archsurg.2009.117.

[98] M. B. Morgan, B. F. Branstetter IV, D. M. Lionetti, J. S. Richardson, and P. J. Chang, “The radiology digital dashboard: Effects on report turnaround time,” *J. Digit. Imaging*, vol. 21, no. 1, pp. 50–58, Mar. 2008, doi: 10.1007/s10278-007-9008-9.

[99] G. Rushton, “Public health, GIS, and spatial analytic tools,” *Annu. Rev. Public Heal.*, vol. 24, pp. 43–56, 2003, doi: 10.1146/annurev.publhealth.24.012902.140843.

[100] P. Elliott, J. Wakefield, N. Best, and D. Briggs, *Spatial Epidemiology*. Oxford University Press, 2009.

[101] R. Bergquist and L. Rinaldi, “Health research based on geospatial tools: A timely approach in a changing environment,” *J. Helminthol.*, vol. 84, no. 1, pp. 1–11, 2010, doi: 10.1017/S0022149X09990484.

[102] F. Conte, P. Vitale, A. Vollero, and A. Siano, “Designing a Data Visualization Dashboard for Managing the Sustainability Communication of Healthcare Organizations on Facebook,” doi: 10.3390/su10124447.

[103] A. L. Hartzler, S. Chaudhuri, B. C. Fey, D. R. Flum, and D. Lavallee, “Integrating Patient-Reported Outcomes into Spine Surgical Care through Visual Dashboards: Lessons Learned from Human-Centered Design,” *eGEMs (Generating Evid. Methods to Improv. patient outcomes)*, vol. 3, no. 2, p. 2, Mar. 2015, doi: 10.13063/2327-9214.1133.

[104] D. Ragouzeos *et al.*, “‘Am I OK?’ using human centered design to empower rheumatoid arthritis patients through patient reported outcomes,” *Patient Educ. Couns.*, vol. 102, no. 3, pp. 503–510, Mar. 2019, doi: 10.1016/j.pec.2018.10.016.

[105] L. Mayoux, “Participatory methods.” Accessed: Apr. 02, 2021. [Online]. Available: http://www.sef.co.za.

[106] P. J. Lavrakas, N. Kim, P. S. Visser, and J. A. Krosnick, “Survey Research,” *Handb. Res. Methods Soc. Personal. Psychol.*, pp. 404–442, 2013, [Online]. Available: https://pprg.stanford.edu/wp-content/uploads/2014-Handbook-of-Research-Methods-in-Social-Psych-Second-Edition.pdf.

[107] E. Martin, “Survey Questionnaire Construction,” *Encycl. Soc. Meas.*, pp. 723–732, 2005, doi: 10.1016/B0-12-369398-5/00433-3.

[108] F. Scheuren and A. S. Association., “What is a Survey.” American Statistical Association, Washington, 2004, Accessed: Mar. 13, 2021. [Online]. Available: http://www.whatisasurvey.info.

[109] J. Rowley, “Conducting research interviews,” *Manag. Res. Rev.*, vol. 35, pp. 260–271, 2012, doi: 10.1108/01409171211210154.

[110] S. Hannabuss, “Research interviews,” *New Libr. World*, vol. 97, no. 5, pp. 22–30, Sep. 1996, doi: 10.1108/03074809610122881.

[111] C. Bezzi, “Evaluational Brainstorming,” *Sociol. Mind*, vol. 01, no. 04, pp. 151–155, Oct. 2011, doi: 10.4236/sm.2011.14019.

[112] K. Holt, “Brainstorming - From classics to electronics,” *J. Eng. Des.*, vol. 7, no. 1, pp. 77–82, Mar. 1996, doi: 10.1080/09544829608907928.

[113] V. L. Putman and P. B. Paulus, “Brainstorming, brainstorming rules and decision making,” *J.*

Creat. Behav., vol. 43, no. 1, pp. 29–40, Mar. 2009, doi: 10.1002/j.2162-6057.2009.tb01304.x.

[114] R. Ørngreen and K. Levinsen, “Workshops as a Research Methodology,” *Electron. J. e-Learning*, vol. 15, p. 70, 2017.

[115] V. Lauttamäki, *Practical guide for facilitating a futures workshop*. Finland Futures Research Centre, 2014.

[116] S. Keeney, F. Hasson, and H. P. McKenna, “A critical review of the Delphi technique as a research methodology for nursing,” *Int. J. Nurs. Stud.*, vol. 38, no. 2, pp. 195–200, Apr. 2001, doi: 10.1016/S0020-7489(00)00044-4.

[117] C. Powell, “The Delphi technique: myths and realities,” *J. Adv. Nurs.*, vol. 41, no. 4, pp. 376–382, Feb. 2003, doi: 10.1046/j.1365-2648.2003.02537.x.

[118] H. Donohoe, M. Stellefson, and B. Tennant, “Advantages and Limitations of the e-Delphi Technique: Implications for Health Education Researchers.”

[119] H. M. Donohoe and R. D. Needham, “Moving best practice forward: Delphi characteristics, advantages, potential problems, and solutions,” *International Journal of Tourism Research*, vol. 11, no. 5. pp. 415–437, Sep. 2009, doi: 10.1002/jtr.709.

[120] C. Helms, A. Gardner, and E. McInnes, “The use of advanced web-based survey design in Delphi research,” *J. Adv. Nurs.*, vol. 73, no. 12, pp. 3168–3177, Dec. 2017, doi: 10.1111/JAN.13381.

[121] M. Karami, M. Langarizadeh, and M. Fatehi, “Evaluation of Effective Dashboards: Key Concepts and Criteria,” *Open Med. Inform. J.*, vol. 11, no. 1, pp. 52–57, Oct. 2017, doi: 10.2174/1874431101711010052.

[122] J. Ahn, F. Campos, M. Hays, and D. Digiacomo, “The Journal of Learning Analytics works under a Creative Commons License,” vol. 6, no. 2, pp. 70–85, doi: 10.18608/jla.2019.62.5.

[123] K. Miettinen and K. Miettinen, “Survey of methods to visualize alternatives in multiple criteria decision making problems,” *OR Spectr.*, vol. 36, pp. 3–37, 2014, doi: 10.1007/s00291-012-0297-0.

[124] A. Kirk, “Data visualization : a successful design process ; a structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively.” Packt Pub, Birmingham, UK, 2013.

[125] H. Wittmer, F. Rauschmayer, and B. Klauer, “How to select instruments for the resolution of environmental conflicts?,” *Land use policy*, vol. 23, no. 1, pp. 1–9, Jan. 2006, doi: 10.1016/j.landusepol.2004.09.003.

[126] I. J. Pérez, F. J. Cabrerizo, S. Alonso, Y. C. Dong, F. Chiclana, and E. Herrera-Viedma, “On dynamic consensus processes in group decision making problems,” *Inf. Sci. (Ny).*, vol. 459, pp. 20–35, Aug. 2018, doi: 10.1016/j.ins.2018.05.017.

[127] W. Vernon, “The Delphi technique: A review,” vol. 16, no. 2, 2009.

[128] P. Shirazi, “Identifying Challenges in Cybersecurity Data Visualization Dashboards.”

[129] N. Martin, J. Bergs, D. Eerdekens, B. Depaire, and S. Verelst, “Developing an emergency department crowding dashboard: A design science approach,” *Int. Emerg. Nurs.*, vol. 39, pp. 68–76, Jul. 2018, doi: 10.1016/j.ienj.2017.08.001.

[130] M. A. Chatti *et al.*, “How to design effective learning analytics indicators? a human-centered design approach,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Sep. 2020, vol. 12315 LNCS, pp. 303–317, doi: 10.1007/978-3-030-57717-9_22.

[131] J. Rowley and F. Slack, “Conducting a literature review,” *Management Research News*, vol. 27, no. 6. pp. 31–39, Jun. 01, 2004, doi: 10.1108/01409170410784185.

[132] E. Zitzler and S. Künzli, “Indicator-based selection in multiobjective search,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3242, pp. 832–842, 2004, doi: 10.1007/978-3-540-30217-9_84.

[133] A. L. Hartzler, J. P. Izard, B. L. Dalkin, S. P. Mikles, and J. L. Gore, “Design and feasibility of integrating personalized PRO dashboards into prostate cancer care,” *J. Am. Med. Informatics Assoc.*, vol. 23, no. 1, pp. 38–47, Jan. 2016, doi: 10.1093/jamia/ocv101.

[134] I. Belton, A. MacDonald, G. Wright, and I. Hamlin, “Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process,” *Technol. Forecast. Soc. Change*, vol. 147, pp. 72–82, Oct. 2019, doi: 10.1016/J.TECHFORE.2019.07.002.

[135] H. A. Von Der Gracht, “The future of logistics: Scenarios for 2025,” *Futur. Logist. Scenar. 2025*, pp. 1–309, 2008, doi: 10.1007/978-3-8349-9764-7.

[136] E. G. Trevelyan and N. Robinson, “Delphi methodology in health research: How to do it?,” *Eur. J. Integr. Med.*, vol. 7, no. 4, pp. 423–428, 2015, doi: 10.1016/j.eujim.2015.07.002.

[137] H. A. Von Der Gracht, “Consensus measurement in Delphi studies Review and implications for future quality assurance,” 2012, doi: 10.1016/j.techfore.2012.04.013.

[138] S. Kalaian and R. M. Kasim, “Terminating Sequential Delphi Survey Data Collection,” *Pract. Assessment, Res. Eval.*, vol. 17, 2012, doi: 10.7275/g48q-je05.

[139] A. K. Chakravarti, B. Vasanta, A. S. A. Krishnan, and R. K. Dubash, "Modified Delphi Methodology for Technology Forecasting Case Study of Electronics and Information Technology in India," *Technol. Forecast. Soc. Change*, vol. 58, no. 1–2, pp. 155–165, May 1998, doi: 10.1016/S0040-1625(97)00081-4.

[140] S. Birko, E. S. Dove, and V. Özdemir, "Evaluation of Nine Consensus Indices in Delphi Foresight Research and Their Dependency on Delphi Survey Characteristics: A Simulation Study and Debate on Delphi Design and Interpretation," *PLoS One*, vol. 10, no. 8, p. e0135162, Aug. 2015, doi: 10.1371/JOURNAL.PONE.0135162.

[141] R. A. Green, "The Delphi Technique in Educational Research:," *https://doi.org/10.1177/2158244014529773*, vol. 4, no. 2, Apr. 2014, doi: 10.1177/2158244014529773.

[142] J. S. Dajani, M. Z. Sincoff, and W. K. Talley, "Stability and Agreement Criteria for the Termination of Delphi Studies."

[143] A. B. Costello and J. W. Osborne, "Practical Assessment, Research & Evaluation," *Pract. Assessment, Res. Eval.*, vol. 104, pp. 1–10, 2005.

[144] Đ. T. N. Quyên, "Developing University Governance Indicators and their Weighting System Using a Modified Delphi Method," *Procedia - Soc. Behav. Sci.*, vol. 141, pp. 828–833, Aug. 2014, doi: 10.1016/j.sbspro.2014.05.144.

[145] A. E. E. Sobaih, C. Ritchie, and E. Jones, "Consulting the oracle?: Applications of modified Delphi technique to qualitative research in the hospitality industry," *Int. J. Contemp. Hosp. Manag.*, vol. 24, no. 6, pp. 886–906, 2012, doi: 10.1108/09596111211247227.

[146] D. E. Cohn *et al.*, "Consensus in controversy: The modified Delphi method applied to Gynecologic Oncology practice," *Gynecol. Oncol.*, vol. 138, no. 3, pp. 712–716, Sep. 2015, doi: 10.1016/j.ygyno.2015.07.014.

[147] K. K. Franklin and J. K. Hart, "Idea generation and exploration: Benefits and limitations of the policy delphi research method," *Innov. High. Educ.*, vol. 31, no. 4, pp. 237–246, Jan. 2007, doi: 10.1007/s10755-006-9022-8.

[148] R. C. de Loë, N. Melnychuk, D. Murray, and R. Plummer, "Advancing the State of Policy Delphi Practice: A Systematic Review Evaluating Methodological Evolution, Innovation, and Opportunities," *Technol. Forecast. Soc. Change*, vol. 104, pp. 78–88, Mar. 2016, doi: 10.1016/j.techfore.2015.12.009.

[149] T. Gnatzy, J. Warth, H. Von Der Gracht, and I.-L. Darkow, "Validating an innovative real-time Delphi approach - A methodological comparison between real-time and conventional Delphi

studies," *Technol. Forecast. Soc. Chang.*, vol. 78, pp. 1681–1694, 2011, doi: 10.1016/j.techfore.2011.04.006.

[150] T. Gordon and A. Pease, "RT Delphi: An efficient, 'round-less' almost real time Delphi method," *Technol. Forecast. Soc. Change*, vol. 73, no. 4, pp. 321–333, 2006, doi: 10.1016/j.techfore.2005.09.005.

[151] J. T. Biehl, M. Czerwinski, G. Smith, and G. G. Robertson, "FASTDash," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*, 2007, pp. 1313–1322, doi: 10.1145/1240624.1240823.

[152] A. Sarcevic, I. Marsic, and R. S. Burd, "Dashboard design for improved team situation awareness in time-critical medical work: Challenges and lessons learned," in *Designing Healthcare That Works: A Sociotechnical Approach*, Elsevier Inc., 2018, pp. 113–131.

[153] H. Tokola, C. Gröger, E. Järvenpää, and E. Niemi, "Designing Manufacturing Dashboards on the Basis of a Key Performance Indicator Survey," in *Procedia CIRP*, Jan. 2016, vol. 57, pp. 619–624, doi: 10.1016/j.procir.2016.11.107.

[154] R. Martinez Maldonado, J. Kay, K. Yacef, and B. Schwendimann, "An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7315 LNCS, pp. 482–492, 2012, doi: 10.1007/978-3-642-30950-2_62.

[155] A. Papadopoulos, "A user centered approach to developing an interactive dashboard that supports task planning."

[156] A. Giordanengo *et al.*, "Design and prestudy assessment of a dashboard for presenting self-collected health data of patients with diabetes to clinicians: Iterative approach and qualitative case study," *JMIR Diabetes*, vol. 4, no. 3, pp. 1–21, 2019, doi: 10.2196/14002.

[157] M. Yoklavich, J. Reynolds, and D. Rosen, "A comparative assessment of underwater visual survey tools: results of a workshop and user questionnaire," 2015, doi: 10.7289/V5/TM-SWFSC-547.

Annexes

Annex A.

Delphi type	Modification	Advantages	Disadvantages
Modified	The first questionnaire integrates a set of thoroughly chosen topics extracted from different sources (summarized reviews of the literature and/or interviews with experts) [144].	Ensure the credibility of researcher by providing reliable scientific background [145]. Reduce the deviation to personal preferences [146].	The restriction of idea generation by predefined set of topics [145].
Policy	The objective is to produce wide variety of distinct opinions on a particular topic rather than reach consensus, in order to estimate impact, acceptability and consequences of any particular option, and to examine [147].	Best suited to examine problems that require inputs from multiple, diverse, and usually contradictory points of view. Identification of opposing positions, agreement or disagreement points and views [148].	High inconsistency due to the heterogeneity of views and fragmented understanding of a problem [148]. Lack of consideration regarding the level of disagreement [147].
Real-time	Round-less method, using online survey which gives immediate feedback and opportunity to re-assess and adjust their estimations during the predefined time period [149].	Reduce redundancies in arguments [149]. High efficiency and response rates [150].	The different feedbacks may have an impact on experts'(final) judgements and opinion convergence [149]. Impossibility to monitor the progress over time [150].
e-Delphi	Uses the internet-based platform to organize, monitor and facilitate interaction between the researcher and the participants [118].	Convenience and flexibility for both researcher and participants imposed by possibility to access information wherever there is Internet access. Time and cost savings. Facilitation of data management, transparency, and reduction of errors. Efficient data collection and analysis [118].	Mandatory requirement of Internet access Platform use complications. [118].

Annex B.

PARTICIPATORY METHOD	ARTICLES	CONTEXT
SURVEY	*D. Dowding and J. A. Merrill, "The Development of Heuristics for Evaluation of Dashboard Visualizations," Appl. Clin. Inform., vol. 9, no. 3, pp. 511–518, Jul. 2018, doi: 10.1055/s-0038-1666842.*[26]	Online survey to identify a candidate set of visualization evaluation heuristics for dashboard.
	J. T. Biehl, M. Czerwinski, G. Smith, and G. G. Robertson, "FASTDash: A visual dashboard for fostering awareness in software teams," in Conference on Human Factors in Computing Systems - Proceedings, 2007, pp. 1313–1322, doi: 10.1145/1240624.1240823.[151]	Electronic survey to identify the existing visualization tools that can be relevant for dashboard in case
	A. Sarcevic, I. Marsic, and R. S. Burd, "Dashboard design for improved team situation awareness in time-critical medical work: Challenges and lessons learned," in Designing Healthcare That Works: A Sociotechnical Approach, Elsevier Inc., 2018, pp. 113–131.[152]	Survey to reflect the design thinking of participants based on actual scenarios for Time-Critical Medical dashboard
	H. Tokola, C. Gröger, E. Järvenpää, and E. Niemi, "Designing Manufacturing Dashboards on the Basis of a Key Performance Indicator Survey," in Procedia CIRP, Jan. 2016, vol. 57, pp. 619–624, doi: 10.1016/j.procir.2016.11.107. [153]	Exploratory survey to identify the preferred type of visualization for certain type of data
INTERVIEWS	*A. L. Hartzler, J. P. Izard, B. L. Dalkin, S. P. Mikles, and J. L. Gore, "Design and feasibility of integrating personalized PRO dashboards into prostate cancer care," J. Am. Med. Informatics Assoc., vol. 23, no. 1, pp. 38–47, Jan. 2016, doi: 10.1093/jamia/ocv101.* [133]	Interviews to identify patients' and providers' preferences regarding visual formats for PRO dashboard
	J. Ahn, F. Campos, M. Hays, and D. Digiacomo, "The Journal of Learning Analytics works under a Creative Commons License," vol. 6, no. 2, pp. 70–85, doi: 10.18608/jla.2019.62.5.[122]	Semi-structured interviews to explore the visual features for learning analytics dashboard.
	J. Piela, "KEY PERFORMANCE INDICATOR ANALYSIS AND DASHBOARD VISUALIZATION IN A LOGISTICS COMPANY."	Semi-structured interview to identify the best measures to illustrate the chosen measures for logistic dashboard
	N. Zaric, M. Gottschlich, R. Roepke, and U. Schroeder, "Supporting gamification with an interactive gamification analytics tool (igat)," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Sep. 2020, vol. 12315 LNCS, pp. 461–466, doi: 10.1007/978-3-030-57717-9_45.	Semi-structured interviews with the objective to improve visualization tools in Gamification dashboard
	R. Martinez Maldonado, J. Kay, K. Yacef, and B. Schwendimann, "An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 7315 LNCS, pp. 482–492, 2012, doi: 10.1007/978-3-642-30950-2_62.[154]	Iterative series of interviews to evaluate both the visualizations and the structure of the learning dashboard
BRAINSTORMING	*P. Shirazi, "Identifying Challenges in Cybersecurity Data Visualization Dashboards."*[128]	Brainstorming to identify design features of security visualization dashboard
	A. Papadopoulos, "A user centered approach to developing an interactive dashboard that supports task planning."[155]	Brainstorming session to identify new interactive visualizations to support the task planning dashboard.

WORKSHOP	*W. Bibliothek and by Johannes Dostal, "Designing a Dashboard for Student Information Systems Media and Human-Centered Computing." Accessed: Apr. 10, 2021. [Online]. Available: www.tuwien.ac.at.*[27]	Workshop to evaluate different design solutions for Students Information Systems
	A. Sarcevic, I. Marsic, and R. S. Burd, "Dashboard design for improved team situation awareness in time-critical medical work: Challenges and lessons learned," in Designing Healthcare That Works: A Sociotechnical Approach, Elsevier Inc., 2018, pp. 113–131.[152]	Participatory design workshops to identify the design elements for Time-Critical Medical dashboard
	A. Giordanengo et al., "Design and prestudy assessment of a dashboard for presenting self-collected health data of patients with diabetes to clinicians: Iterative approach and qualitative case study," JMIR Diabetes, vol. 4, no. 3, pp. 1–21, 2019, doi: 10.2196/14002.[156]	Facilitated workshops to define how information should be displayed on health dashboard.
	M. Yoklavich, J. Reynolds, and D. Rosen, "NOAA Technical Memorandum NMFS A COMPARATIVE ASSESSMENT OF UNDERWATER VISUAL SURVEY TOOLS: RESULTS OF A WORKSHOP AND USER QUESTIONNAIRE," 2015, doi: 10.7289/V5/TM-SWFSC-547.[157]	Workshops to select underwater visual survey tools
DELPHI	*S. Al-Hajj, I. Pike, and B. Fisher, "Interactive Dashboards: Using Visual Analytics for knowledge Transfer and Decision Support."*[28]	Modified Delphi to solicit experts' inputs regarding the design of a visual analytical tools for interactive dashboard
	P. Shirazi, "Identifying Challenges in Cybersecurity Data Visualization Dashboards."[128]	Tailored Delphi to identify design features of security visualization dashboard
	N. Martin, J. Bergs, D. Eerdekens, B. Depaire, and S. Verelst, "Developing an emergency department crowding dashboard: A design science approach," Int. Emerg. Nurs., vol. 39, pp. 68–76, Jul. 2018, doi: 10.1016/j.ienj.2017.08.001. [129]	Modified Delphi to identify visualization tools required for emergency department crowding dashboard.

Annex C.

1. **Daily (or cumulative) number of new confirmed cases**

Number of cases	Day
A_{11}	A_{12}
A_{21}	B_{22}
…	…

Data variables: QQ - 2 x quantitative (Cases, Day)
(considering the day as a categorical value – CQ)
Communication purpose: Showing changes over time
Data visualization formats: Line Chart, Area Chart, Column Chart

2. **Daily number of new confirmed cases by region:**

	Cases		
Day	ARS Norte	ARS Centro	ARS Alentejo
A_{11}	A_{12}	A_{13}	A_{14}
A_{21}	B_{22}	B_{23}	B_{24}
…	…		

Data variables: QQC **-** 2 x quantitative (Cases, Day), 1 x categorical (Regions)
(considering day as a categorical value - CCQ)
Communication purpose: Showing changes over time
Data visualization formats: Line Chart, Area Chart, Grouped Column Chart, Stacked Column Chart

3. **Total number of confirmed cases by age group**

Number of cases	Age Group
A_{11}	0-9
A_{21}	10-19
…	…

Data variables: CQ - 1 x quantitative (Cases), 1 x categorical (Age Group)

Communication purpose: Comparing categorical values + Assessing part-to-whole-relationships

Data visualization formats: Bar Chart, Column Chart, Pie Chart, Donut Chart, Stacked Bar Chart

4. **Total number of confirmed cases by sex**

Number of cases	Sex
A_{11}	Feminine
A_{21}	Masculine

Data variables: CQ - 1 x quantitative (Cases), 1 x categorical (Sex)

Communication purpose: Comparing categorical values + Assessing part-to-whole-relationships

Data visualization formats: Bar Chart, Column Chart, Pie Chart, Donut Chart, Stacked Bar Chart

5. **Total number of confirmed cases by region**

Number of cases	Region
A_{11}	ARS Norte
A_{21}	ARS Centro
…	…

Data variables: CQ - 1 x quantitative (Cases), 1 x categorical (Region)

Communication purpose: Comparing categorical values + Assessing part-to-whole-relationships

Data visualization formats: Bar Chart, Column Chart, Pie Chart, Donut Chart, Stacked Bar Chart or, in case if A_{xy} are geographic coordinates: Bubble Map, Choropleth Map, Dasymetric Map, Point Map.

6. **Total number of confirmed cases by sex and age group:**

	Cases	
Age Group	Feminine	Masculine
A_{11}	A_{12}	A_{13}
A_{21}	B_{22}	B_{23}
…	…	…

Data variables: CCQ **-** 2 x categorical (Age Group, Sex), 1 x quantitative (Cases)

Communication purpose: Comparing categorical values + Assessing part-to-whole-relationships

Data visualization formats: Two-sided Bar Chart, Grouped Column Chart, Stacked Column Chart, 100% Stacked Column Chart

Annex D.

Available at: https://surveyhero.com/c/ad39c641

Annex E.

Available at: https://surveyhero.com/c/cwxq49r9