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

Ekaterina Ignatenko

Master Science Degree in Biomedical Engineering, Instituto Superior Técnico

ekaterina.ignatenko@tecnico.ulisboa.pt

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 visualization 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 execute 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; COVID-19; Data visualization tools; Delphi process; Portugal

1. INTRODUCTION

Nowadays, due to the massive data generation in all areas, new challenges exist to the selection and visualization of relevant information. In healthcare, evolving data-related concerns present analogous challenges of information integration, putting public health practitioners under more stress [1]. As a consequence, new data visualization tools have emerged, allowing the representation of data that stimulates human´s perceptual and cognitive abilities of problem solving [2]. 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. Data visualization tool allows to use visual elements, with the purpose of representing large amounts of information in the way that facilitates data interpretation. Through data visualization tools, 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 [3].

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. However, deciding on the right visualization methods is not easy at all since the effective visualization tool must be perceptible and render the main characteristics of the metric in question. Otherwise, the selected visualization can be misleading and guide to wrong conclusions [4].

The evolution of visualization methods culminated in the creation of dashboards, data visualization tools which incorporate different visualization formats on a single screen to improve efficiency of information transmission [5]. 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 [6], [7].

Despite the growing popularity of data visualization tools, little is known about the procedures and methods used to support their formats [8]. Not only there is a few number of studies regarding the construction of data visualizations tools, but also existing ones do not provide recommendations for theorizing the selection of data visualization formats. Moreover, there are significant limitations regarding approaches used to select data visualizations [9]–

[12]: theoretical reflections on visualizations are rarely performed in information systems science, which leaves a wide-open research question as to how to depict and select visualizations independently from their widely visual appearance [13].

The end-user's perception is emphasized as an important contribute for visualization selection [14]. The participation of stakeholders or users in dashboard design is considered as 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, the participatory methods have been adopted, which improve the usability of interactive systems by collecting and analysing a direct input from users [15], [16]. The integration of stakeholders 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 [17] However, despite the highlighted importance of stakeholder's opinion for dashboard design, the number of studies that integrate participatory methods to explore visualizations options of dashboards is very limited.

In this paper, attention is given to the development of approach, which integrates participatory methods to inform the selection of data visualization formats in the context of dashboard design for the COVID-19 pandemic. The research aims to provide a rationale for selecting appropriate visualizations, either considering theory-based evidence and the perspective of dashboard end users.

1.1. 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) [18]. 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 [19], [20].

Therefore, different policies for detection and containment of clusters of infection were established to control the propagation of pandemic through community transmission. For that reason, many data visualization tools emerged for decision support - for instance, web-based dashboards have been developed to facilitate the transmission of relevant information to the general population and to promote an understanding of the COVID-19 data by the public. Therefore, public web-based COVID-19 dashboards ultimately share a common objective: to serve as both a communication tool and call for individual and collective action to respond to the COVID-19 pandemic [21].

The majority of existing web-based public COVID-19 dashboards is focused on epidemiological indicators, such as number of confirmed cases and mortality. These indicators can be reported at different geographic levels, representing spatial variations, or represented over time, predominantly by day, to show the evolution of the pandemic and the effects of implemented policies [22], [23]. In addition to geographic and temporal breakdowns, dashboards are used to analyse data by other breakdowns, the most common of which are age and sex [21].

Visualization techniques have been front-and-center in the efforts to communicate the science around COVID-19 to the general population. Public web-based COVID-19 dashboards utilize data visualization formats to transmit indicators usually in the form of graphs or charts, maps and tables [21]. However, an overall underuse of known and proven delivery data techniques and a lack of insights from users is a common issue for COVID-19 dashboards, that can mislead both unintentionally and intentionally if the data visualizations are not accurately selected and represented [24].

1.2. SCOPE - Spatial Data Science Services for COVID-19 Pandemic

The research developed in this study has been developed within the scope of the SCOPE project. Specifically, it contributes to one of the objectives of the project that aims to bridge the gap between the creation of risk maps and dashboards related to the pandemic, and the use of such maps to support decision-making and the design of policies. The COVID-19 maps produced by the SCOPE project 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 analyse 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.

2. LITERATURE REVIEW

2.1. Data visualizations

The selection of appropriate visualization formats to satisfy specific objective is essential for effective data visualization tool. The traditional view holds that visualizations can be selected by mapping conceptual elements (e.g., data sets) and visual representations (e.g., charts). This way, the essential properties of a visual representation are abstracted from its individual graphical expressions, such as shape, colour or position, and focus on the underlying data structure [25]. Several studies ([13], [27], [24]) suggest the identification of number of variables and their type, categorical or quantitative, in order to proceed with the refinement of data visualization choice. This way, the nature of the variables in question reduces the range of suitable visualizations.

However, Kirk [3] suggests that the choice of visualization format must be firstly driven by its desired purpose, and afterwards, the accommodation of data properties must be performed. In this respect, the taxonomy suggested by Kirk categorizes most common data visualization formats by the primary communication purpose, focusing on variety of possible outcomes:

- Comparing categorical values
- Assessing hierarchies and part-of-a-whole relationships
- Showing changes over time
- Charting and graphing relationships
- Mapping geo-spatial data

This categorization is flexible, since there are examples of charts that span across two method classifications.

2.2. Dashboards

The analysis of output of multiple data visualization formats requires navigating between different platforms, which is a time-consuming and inefficient process. Therefore, a dashboard was invented to provide the capability to compare and analyse different visual formats at the same time, which is essential to establish relations between objectives and variables and to improve efficiency of decision-making [28]. A dashboard provides a rich user interface that exhibits the data in a graphical form using a range of visualization formats, e.g., charts, tables and maps [29].

Dashboards are used across different domains due to their beneficial contribution for organizations. Their adoption in healthcare domain allows healthcare professionals to examine and determine trends in data and assist them in better identification of anomalies and their interpretation. There are numerous examples of application of dashboards to real-life problems in healthcare fields [30], [31], [32], [33]. Lately, the healthcare organizations are also forced to communicate their efforts, initiatives and activities not only to stakeholders, but also to the public though internet and social media. Therefore, dashboards also provide the solution as an efficient tool to transmit the needed information to a generic population in a simple and transparent way [33].

The information that should be included on the dashboard is usually reflected by indicators. They are considered one of the most vital components of dashboard, since they provide the capability not only to assess the 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 to evaluate a complex social, economic, or physical context” [34].

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 type and purpose of dashboard. Data selection and visualization design is considered the most important one, since it involves the design and construction of the dashboard, including the identification of indicators and selection of corresponding data visualization formats, the posterior extraction of selected indicators from database and their implementation within dashboard design.

2.3. Participatory methods in dashboard design

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, such as survey, interview, brainstorming, workshop and Delphi, for data visualization tools’ design. After examination, a set of 18 articles were selected for their posterior analysis.

Despite the highlighted importance of stakeholder’s opinion for dashboard design, the number of discovered studies that integrates participatory methods to explore visualization of dashboard was severely 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 format for certain indicator. The participatory methods mostly explored the feedback of stakeholders through questions about the certain design details, such as elements’ position in dashboard framework and their interactivity. In other cases, the participatory methods were used to evaluate and validate the pre-developed dashboard, reducing the stakeholders’ level of involvement in dashboard design.

3. PROPOSED APPROACH AND IMPLEMENTATION

To understand which visualization formats should be incorporated within dashboards for the COVID-19 pandemic, a specifically designed approach, based on a Delphi process, was developed to understand the preferences and views of the general population regarding distinct data visualization formats. The developed approach was implemented to select appropriate data visualization formats for presenting information commonly presented in public web-based COVID-19 dashboards. First, the indicators mostly used in COVID-19 were established, and finally, the implementation of the developed approach was described. Specifically, the selection of pre-set of data visualization formats was performed, taking in account the theory-based evidence regarding the data to be considered. Finally, the Delphi process was described and implemented. 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.

3.1. Indicators' selection

The selection of appropriate and well-designed indicators to integrate dashboards is vital for dashboard effectiveness and performance, and so it usually complex and requires specific attention.

Since the focus of this study is to explore which visualization formats are appropriate – according to views of the public – to display dashboard information, the starting point is a set of indicators commonly used for web-based public COVID-19 dashboards. Specifically, the study entitled "Features Constituting Actionable COVID-19 Dashboards: Descriptive Assessment and Expert Appraisal of 158 Public Web-Based COVID-19 Dashboards" [21] was analysed, which explores characteristics of 158 public web-based COVID-19 dashboards by assessing their features, namely the key performance indicators and their frequency; and summarizes the types of analysis and presentation of data in such dashboards, for instance the frequency of considerations such as time trends, geographic levels and disaggregation options (sex, age, etc.).

Accordingly, one has selected the indicators related to the theme "Cases", which comprise the most common representation types included in dashboards, such as age, sex and regions. Indicators of that theme have the highest frequency in public COVID-19 dashboards, while the inclusion of most frequent representation 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 per 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.

3.2. Selection of pre-set of data visualization formats

This step transforms the input data into structured data, and it was decided to combine the data visualization process from Dastani [35] with Kirk's approach [3], where the first step covers the identification of the data structure. Hence, each indicator is structured into $m \times n$ data table, and afterwards, an attribute-based classification of identified variables is performed. The data attributes can comprise data values that can be classified as quantitative or categorical. Alternatively stated, the quantitative variables contain numerical values, whereas categorical ones have individual values (e.g., geographic regions, names, or products). For further simplification, one can use the designation used by Helfman [36], 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, categorical (C) or quantitative (Q). However, Kirk [3] highlights the necessity to "telling stories" with data visualizations and 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 with indicator was considered to be important.

The second step in data visualization process [35], is to 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. Therefore, the decision table was developed (Table 1), using information from literature, to enumerate the most common 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 [37]. It allows to predefine the set of most appropriate data visualization formats for certain indicator but does not provide a unique answer and requires the refinement from stakeholders.

Afterwards, the following procedure was adopted:

Table 1– 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 map, Dasymetric map	-	-

1.The data structure of previously selected indicators was identified, i.e., the number of variables and their attribute type were determined.

2.The communication purpose of data visualization formats for each indicator was discovered, considering the categorization provided by Kirk.

3.The set of data visualization formats was defined by mapping previously determined data structure and communication purpose using a developed decision table (Table 1).

Through this sequence, data visualization formats to be included in the Delphi processes were selected, leading to the formats presented in Table 2.

Table 2 – Indicators and correspondent data visualization formats.

Indicator	Data visualization formats
Daily number of new confirmed cases	Line Chart, Area Chart, Column Chart.
Cumulative number of confirmed cases	
Daily number of new confirmed cases per region	Line Chart, Area Chart, Grouped Column Chart, Stacked Column Chart
Total number of confirmed cases by:	
region	Bar Chart, Column Chart, Pie Chart, Donut Chart, 100% Stacked Bar Chart.
age group	
sex	
sex and age group	Two-sided Bar Chart, Grouped Column Chart, Stacked Column Chart, 100% Stacked Column Chart.

Additionally, choropleth, bubble, dasymetric and point maps were selected to represent indicators with associated geographic location.

Since suggested process offers the set of alternative data visualization formats for the 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.

Delphi process

The results of an evaluation of different participatory methods have demonstrated that the Delphi is considered more appropriate for the public dashboard design, where the heterogeneity of population implies high number of participants, and high integration and inclusion of different types of information. Moreover, the Delphi approach facilitates convergence between participants by constant learning during the process and flexibility to change certain behaviour by analysing the other participants' opinions. It is also noteworthy to mention that the Delphi becomes the method of choice where there is little previous research, which is the case of this study [38].

The Delphi is a structured communication process which uses a number of questionnaires or rounds with controlled feedback to collect and deliver information with the objective to achieve a group consensus [39]. The Delphi process aims to identify, forecast, and investigate group attitudes, needs, and priorities through a series of rounds in which participants' viewpoints are gathered through their individual responses to the same questionnaire. Thus, while maintaining anonymity, a summary of the responses is provided back to the participants, who may change their minds in following rounds as a result of this collective knowledge [39], [40].

Drawing on the outcome of previous step, one can distinguish a gross list of potential indicators for further analysis regarding their visual representation. Furthermore, the Web-Delphi process was used to refine the set of data visualizations corresponding to each indicator. The Delphi process used in this study can be considered as modified, since the credibility of the questions elaborated for first round is ensured by scientific background provided from the previous step. Namely, the main objective of this process was 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.

As attrition is likely to increase with each round, to ensure against participant fatigue, but also to guarantee results are meaningful, two rounds were selected, therefore an *a priori* criterion of two rounds in this Delphi study was established.

First round of Delphi took the form of a structured questionnaire with a total of 12 questions, including statements generated from the Table 2, 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 and cumulative number of new confirmed cases. Moreover, the analysis of perceptions captured from observing different time intervals (one week versus three months) in certain data visualization formats was performed to understand the way cognition evaluates the variance in the COVID-19 cases. The third part explored the data visualization formats corresponding to total number of confirmed cases since the beginning of the COVID-19 pandemic, aggregated by region, age group and/or sex. Finally, the last part of questionnaire corresponded to the map visualizations. The first question offered the set of map visualizations regarding the indicator of COVID-19 incidence, including choropleth, bubble, dasymetric and point maps. Finally, 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 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.

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' votes for distinct data visualization formats. 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 survey tool selected to design and execute the Delphi process was the one designated as SurveyHero, due to the possibility to integrate the image as a multiple-choice option. This platform presents the "Image Choice" question type, which allows to create a multiple-choice question using images as the possible answers. The user can also zoom in on the image to see all the details.

3.3. Analysis of Delphi results

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, 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 [41].

Many Delphi processes use certain levels of agreement in order to quantify consensus among an expert panel [41]. This quantification is important 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), 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 [41], as in the case of this Delphi process.

Therefore, to analyze the case study results, it was decided to calculate the frequency and percentage of votes (percentage of agreement) 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.

The degree to which a study process produces consistent results each time it is repeated is known as stability. It occurs when responses obtained in two successive rounds are shown to be not significantly different from each other, irrespective of whether a convergence of opinion occurs [42]. There are number of statistical tests for determine stability, however, tests which are suitable for use on nominal data are highly restricted. 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.

3.4. 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 visualization techniques from perspective of final users [21]. The lack of support to guide visualization choices for diverse dataset domains, and need 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.

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 official reports, representing the 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.

4. CASE STUDY RESULTS AND DISCUSSION

4.1. Delphi participation

During the first round, panellists 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, the participants were asked to leave their e-mail in order to be contacted for the second round, however, only 72 participants from total 101 (71.3% completion rate) decided to fill their e-mail. This could be explained by participants feeling more threatened to share their personal information, besides the guarantee of anonymity and the use of email only for study purpose. 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 a concept of dashboard.

For the second round, the response rate was 65%, where 47 panellists decided to participate in second round of the Delphi process. The participants were offered to change or maintain their responses, considering the percentage of first-round votes provided for each option. 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.

4.2. Delphi results

To summarize the second-round results, the next table (Table 3) was created with the mode and correspondent percentage agreement values for each question.

The first part of questionnaire explored the data visualization formats for 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. Moreover, 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. Consequently, these results highlight the importance of using different visualization formats for different screen amplifications.

For the second part of questionnaire the total number of confirmed cases since the beginning of the COVID-19 pandemic was 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 was relatively low to deduce any conclusions. This inconsistency between rounds can be related with the comment of the participant in the first round of Delphi process, that advised not to use a pie chart with high number of categories, provoking the decrease of the votes for this visualization format in the second round. This strong influence of the comment for final results reenforce the usefulness of Delphi due to the interaction of different participants between rounds.

Table 3 – Delphi process: Second-round results.

Which visualization format do you consider to be the most appropriate to represent …			
		Mode	**Percentage agreement**
	a daily number of new confirmed COVID-19 cases, considering:		
1	three-month interval	Line Chart	46,8%
2	one-week interval	Column Chart	46,8%
	a cumulative number of confirmed COVID-19 cases, considering:		
3	three-month interval	Line Chart	63,8%
4	one-week interval	Column Chart	61,7%
	a daily number of new confirmed COVID-19 cases for three regions – Norte, Centro e Alentejo, considering:		
5	three-month interval	Line Chart	46,8%
6	one-week interval	Column Chart	53,2%
	a total number of confirmed COVID-19 cases		
7	per age group	Column Chart	40,4%
8	per sex	Pie Chart	74,5%
9	per region	Column Chart	51,1%
10	per sex and age group	Two-sided Bar Chart	78,7%
11	**a number of new confirmed COVID-19 cases per municipality since the beginning of pandemic (incidence)**		
		Choropleth map	80,9%
12	**a number of new confirmed COVID-19 cases per 100k inhabitants, per municipality since the beginning of pandemic (cumulative incidence)**		
		A risk is constant within administrative limit	80,9%

For the total number of confirmed COVID-19 cases by sex, the most voted data visualization format was a pie chart (74,5%), meanwhile for aggregation by regions a column chart (51,1%) 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 (78,7%) was the most preferred data visualization format.

Lastly, for map visualizations, a choropleth map (80,9%) 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 the last question regarding the project SCOPE, the participants preferred the map with constant infection risk within the administrative unit (80,9%), confirming the fact that the general population has some difficulties to interpret SCOPE map, despite its high precision.

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 format, 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.

4.3. Implications of the Delphi results for dashboard design

The DGS COVID-19 dashboard was analysed in terms of indicators and correspondent data visualization formats used, and afterwards the inconsistencies with case study results were identified, from the perspective of Delphi process. The alternative visualization formats 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 (Figure 1).

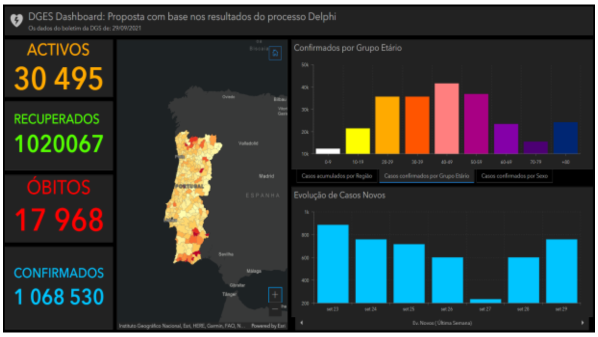

Figure 1 – Adjusted DGS COVID-19 Dashboard.

The implementation was performed without major difficulties, only presenting some limitations regarding zoom function 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.

5. CONCLUSIONS AND FUTURE WORK

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.

BIBLIOGRAPHY

[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] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[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] 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.

[15] 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.

[16] 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.

[17] L. Mayoux, "Participatory methods." Accessed: Apr. 02, 2021. [Online].

Available: http://www.sef.co.za.
[18] 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.
[19] 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.
[20] "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).
[21] 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.
[22] 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.
[23] 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.
[24] 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.
[25] J. Gulden and H. A. Reijers, "Toward Advanced Visualization Techniques for Conceptual Modeling."
[26] T. Munzner, *Visualization Analysis & Design Tamara Munzner A K Peters Visualization Series Illustrations by Eamonn Maguire*. 2013.
[27] J. Bertin, *Semiology of graphics : diagrams, networks, maps*. Redlands, Calif.: Redlands, Calif. : ESRI Press : Distributed by Ingram Publisher Services, 2010.
[28] A. Fioravanti, G. Novembri, and F. Rossini, "Improving Proactive Collaborative Design Through the Integration of BIM and Agent-Based Simulations," 2017.
[29] 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.
[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] 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.
[38] W. Vernon, "The Delphi technique: A review," vol. 16, no. 2, 2009.
[39] 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.
[40] 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.
[41] 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.
[42] J. S. Dajani, M. Z. Sincoff, and W. K. Talley, "Stability and Agreement Criteria for the Termination of Delphi Studies."