Computers in Human Behavior 155 (2024) 108162

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Data visualization and domain knowledge: Insights through focus groups of researchers in Spain

Andrea Vázquez-Ingelmo [a,*], Alicia García-Holgado [a], Sonia Verdugo-Castro [b], Roberto Therón [a,c], Francisco José García-Peñalvo [a]

[a] GRIAL Research Group, Department of Computer Science, Research Institute for Educational Sciences, Universidad de Salamanca, Salamanca, Spain
[b] GRIAL Research Group, Department of Didactics, Organization and Research Methods, Research Institute for Educational Sciences, Universidad de Salamanca, Salamanca, Spain
[c] VisUSAL Research Group, Department of Computer Science, Research Institute for Educational Sciences, Universidad de Salamanca, Salamanca, Spain

ARTICLE INFO

Handling editor: Matthieu Guitton

Keywords:
Visualization
Perception
Domain knowledge
Spain
Qualitative analysis

ABSTRACT

Nowadays, data visualizations are an intrinsic element of decision-making processes. But they can also lead us to wrong conclusions under different circumstances: lack of data visualization expertise, lack of domain knowledge, biases, polarization, backgrounds, etc. In this sense, it is important to consider potential drawbacks during the design and development processes of these tools and address them before delivering the final product to the user. Following this idea, this work describes the results obtained from the analysis of two focus groups (with different backgrounds and expertise) about data visualizations to validate a previously developed dashboard meta-model and analyze the differences between the two groups regarding the interpretation and thoughts on these tools. The results show differences between the discourses of the two groups, and the analysis of the interventions proves that the meta-model is robust and captures the most relevant elements of data visualizations and information dashboards.

1. Introduction

The use of data visualizations as a means to convey information is not new. Over the years, data visualization techniques have been crucial for decision-making processes, enabling the identification of interesting data points, patterns, trends, etc., that could be almost impossible to discern from raw data (Franconeri, Padilla, Shah, Zacks, & Hullman, 2021; Sarikaya, Correll, Bartram, Tory, & Fisher, 2018; Schwabish, 2021; Tufte, 1985; Tufte, Goeler, & Benson, 1990).

Nowadays, the use of these techniques has popularized even more with the increasing availability of data. Data visualizations are everywhere, assisting our day-to-day decision-making (Hullman, Adar, & Shah, 2011; Kay, Kola, Hullman, & Munson, 2016; Kim, Reinecke, & Hullman, 2018), increasing the understandability of large datasets (Ware, 2004; Wilke, 2019), and supporting scientists, journalists, and many other professionals, in the process of communicating complex information (Cairo, 2012, 2016, 2019).

However, the spread of data visualizations throughout several fields has also increased their potential flaws. While these tools can ease the understanding of patterns and insights hidden in data, they can also lead their audience to wrong conclusions under different circumstances: improper design decisions, lack of data visualization expertise, lack of domain knowledge, biases, polarization, backgrounds, etc. (Dimara, Franconeri, Plaisant, Bezerianos, & Dragicevic, 2018; Lin & Thornton, 2021; Mansoor & Harrison, 2018; Szafir, 2018).

In this sense, it is important to consider potential drawbacks during the design and development processes of data visualizations and address them before delivering the final product to the user. Several works in the literature have tackled this issue by researching how different aspects of data visualizations and the characteristics of the audience can affect the effectiveness and usefulness of these tools (Börner, Maltese, Balliet, & Heimlich, 2016; Correll, Bertini, & Franconeri, 2020; Hullman et al., 2011; Hullman, Qiao, Correll, Kale, & Kay, 2019; Kim et al., 2018; Kim, Walls, Krafft, & Hullman, 2019; Lee, Kim, & Kwon, 2017; Wanzer, Azzam, Jones, & Skousen, 2021).

In our previous research, we created a dashboard meta-model aimed at encapsulating the fundamental and technology-agnostic characteristics of dashboards and data visualizations (Vázquez-Ingelmo,

* Corresponding author.
E-mail address: andreavazquez@usal.es (A. Vázquez-Ingelmo).

https://doi.org/10.1016/j.chb.2024.108162
Received 20 May 2023; Received in revised form 26 January 2024; Accepted 30 January 2024
Available online 10 February 2024
0747-5632/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

García-Holgado, García-Peñalvo, & Therón, 2021; Vázquez-Ingelmo, García-Peñalvo, & Therón, 2019; Vázquez-Ingelmo, García-Peñalvo, Therón, & Conde, 2020; Vázquez-Ingelmo, García-Peñalvo, Therón, & Conde-González, 2019; Vázquez-Ingelmo, García-Peñalvo, Therón, & García-Holgado, 2020). This was done to construct an understanding of the key elements that both creators and users should consider for effectively transmitting and comprehending information via data visualizations.

This meta-model evolved through various iterations, leading to its present structure (refer to Section 2). These gradual enhancements focused not just on specific areas within the domain, but also on identifying and resolving potential challenges in subsequent revisions.

Initial iterations concentrated on the physical attributes of dashboards, like layout, resources, and visual elements. Subsequent versions expanded on defining the intended users of the dashboard and their objectives. Later updates explored more abstract concepts, including interaction patterns — dynamic, intangible behaviors — and the core qualities of the data domain and contextual information, which are linked to the practical comprehension of data sets.

The meta-model serves three primary purposes. Firstly, it acts as a guide for designing data visualizations, emphasizing crucial factors such as audience, interaction methods, data transformation, and contextual details. Secondly, it breaks down intricate tools like data visualizations into their most basic components, enhancing the traceability of design choices made. Lastly, it enables the automatic generation of custom data visualizations and dashboards by combining the fine-grained features captured in the meta-model through approaches such as the Software Product Lines paradigm (Vázquez-Ingelmo, García-Peñalvo, & Therón, 2019).

While existing literature over the years has offered numerous best practices in the design of data visualizations and dashboards (Few, 2006; Munzner, 2014; Wexler, Shaffer, & Cotgreave, 2017), our meta-modelling approach organizes all these concepts and their associated practices into a structured tool, also fostering an educative perspective to understand data visualizations from their most abstract and primitive elements.

Nevertheless, certain abstract concepts, particularly those related to the data domain and data context, require additional examination to confirm their accurate depiction in the meta-model. This step is crucial to ensure that these elements are effectively incorporated as concrete components in the creation of data visualizations.

Through this work, we want to accomplish **two objectives**.

- **O1.** Conduct qualitative validation of the entities within the dashboard meta-model.
- **O2.** Examine the variances in interpretation and perspectives on data visualizations among groups with diverse backgrounds, focusing on how introducing new contextual data into different types of data visualizations affects the interpretation of the displayed information.

While the first goal is centered on ensuring that the dashboard meta-model's contents are reflective of practical experiences, the second aims to understand the nuances of how different backgrounds can shape the understanding and analysis of data visualizations, thereby enriching the meta-model with insights that transcend conventional audience considerations.

Following this research line, the present work describes the results obtained from the analysis of **two focus groups** about data visualizations involving participants with different backgrounds and expertise. Due to the complexity and particularities of the data visualization context, it is crucial to select a proper methodology to retrieve meaningful data from the participants. While quantitative approaches can provide significant quantities of data, they can lack depth, which is critical to understanding the point of view of users. For these reasons, we have chosen to follow a qualitative approach through the focus group technique, as it is one of the most useful techniques for identifying and analyzing reasoning processes, interpretations, and insights (Mazza & Berre, 2007; Wilkinson, 1998).

As previously introduced, the dashboard meta-model aims to provide a skeleton and a framework to understand how data visualizations and information dashboards are composed by abstracting their most primitive elements -both theoretical (goals, tasks, data domain, visualization literacy, etc.) and tangible elements (visual marks, encoding strategies, scales, axes, etc.)-, being easier to identify their low-level components (Huron, Jansen, & Carpendale, 2014). By carrying out focus groups with different backgrounds and analyzing the concepts that arise from the interventions of each participant, it is possible to test if the meta-model is robust and complete enough to represent the main aspects involved in data visualizations.

To sum up, the **major contributions** of this paper are:

1. A qualitative validation of the categories involved in the dashboard meta-model developed in (Vázquez-Ingelmo, García-Holgado, et al., 2021; Vázquez-Ingelmo et al., 2019; Vázquez-Ingelmo et al., García-Peñalvo, Therón, & Conde, 2020; Vázquez-Ingelmo, et al., 2019; Vázquez-Ingelmo, García-Peñalvo, Therón, & García-Holgado, 2020).
2. An investigation into how the addition of contextual information to real data visualizations impacts their interpretation, specifically focusing on the variations in this impact among participants with different backgrounds and varying levels of expertise in data visualization.

The rest of this paper is organized as follows. Section 2 details the theoretical framework and the dashboard meta-model that support this work. Section 3 provides an overview of the methods employed to carry out the study. Section 4 presents and discusses the results derived from the analysis of the focus groups. Finally, Section 5 concludes the work with the main insights and future research lines.

2. Theoretical framework

The present study relies on a dashboard meta-model developed in previous research (Vázquez-Ingelmo, García-Holgado, et al., 2021; Vázquez-Ingelmo et al., 2019; Vázquez-Ingelmo, García-Peñalvo, Therón, & Conde, 2020; Vázquez-Ingelmo, García-Peñalvo, Therón, et al., 2019; Vázquez-Ingelmo, García-Peñalvo, Therón, & García-Holgado, 2020). This meta-model is a modelling artefact developed through domain engineering that supports the definition and development of information dashboards and data visualizations by decomposing these complex tools into lower-level elements.

Fig. 1 displays the current version of the dashboard meta-model, in which four main sections can be identified: the **user**, the **layout**, the **data**, and the **specification of the visualizations**.

Each section of the meta-model addresses different -tangible and conceptual-elements of data visualizations and information dashboards. On the one hand, the **user** section focuses on representing the information about the audience of the dashboard or visualization, that is, the user characteristics, preferences, literacy, domain knowledge, biases, etc., as well as their informational goals when using the tool.

The **layout** section captures the structure of the dashboard, the containers that will hold the selected data visualizations, and other technical attributes like the size and position of the containers.

The **data visualization specification** section provides information about how the data visualizations are built through a combination of elements: visual marks (circles, bars, lines, etc.), axis, annotations, visual channels (color, size, position, opacity, etc.), scales, resources (images, text, etc.), and even interactivity.

Finally, the **dataset** specification is also represented in the meta-model through three entities: "Dataset", "Variable" and "DomainVariable" (Fig. 2). The "Dataset" entity represents the data source of the dashboard and visualizations. This entity is composed of a set of

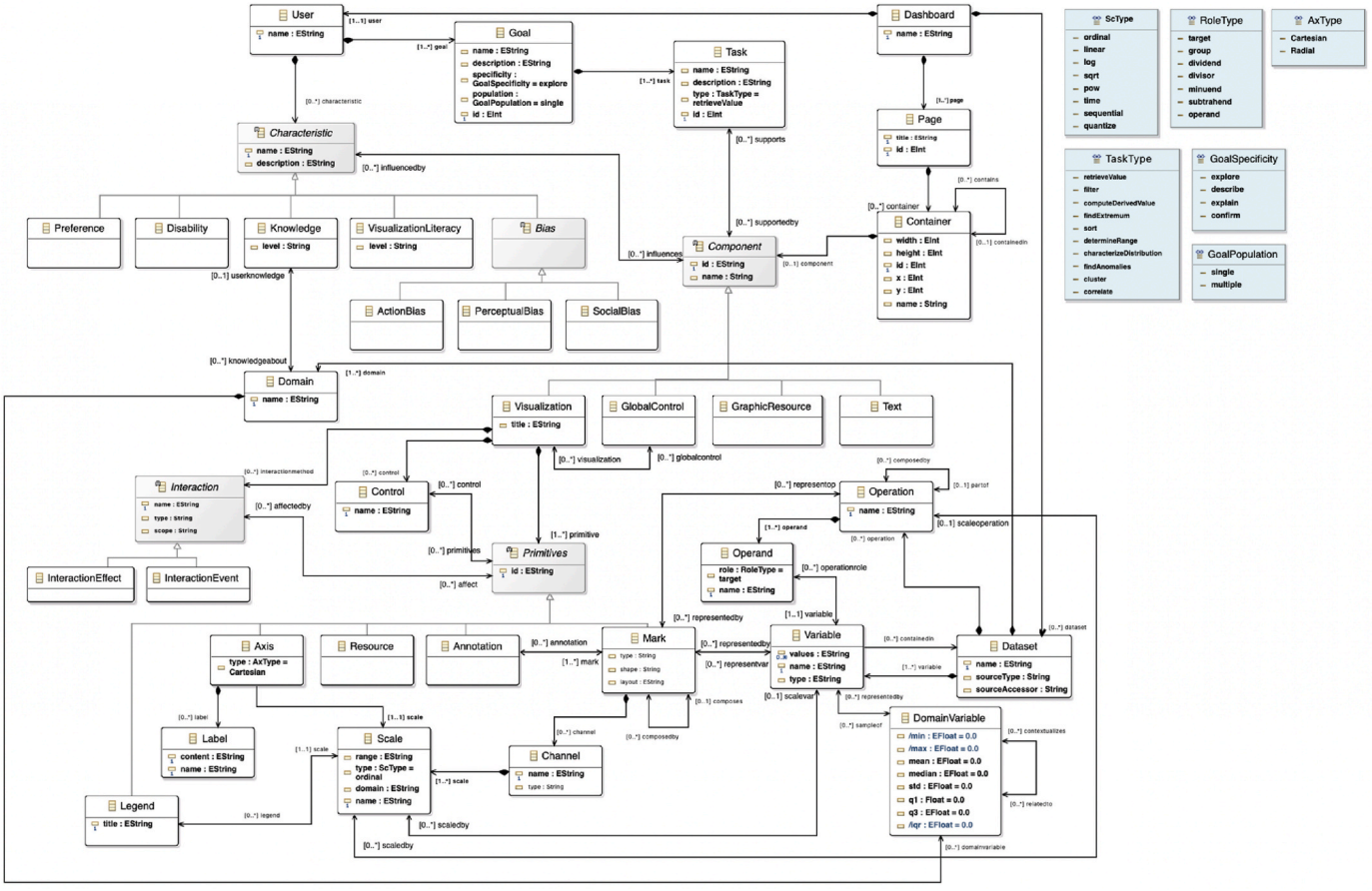


Fig. 1. Current version of the dashboard meta-model (Vázquez-Ingelmo, García-Peñalvo, & Therón, 2021). A high-resolution version can be consulted at https://zenodo.org/record/7037624.

variables (columns) that contain the values (observations) that will be represented through visual elements.

The "DomainVariable" entity represents data variables within the data domain. It is linked to both the 'Domain' class (representing the broader domain in the meta-model) and the 'Variable' class, which denotes a variable from the dataset that will be visualized. Here, the 'Variable' entity is viewed as an instance of the 'DomainVariable' entity (Vázquez-Ingelmo, García-Holgado, et al., 2021).

The introduction of the domain variable is instrumental for conducting analytical tasks on the visualizations, enabling users to gain deeper insights. This includes understanding typical value ranges, identifying outliers, and recognizing emerging trends.

The comprehension of the domain is essential not just for users interpreting the visualizations, but also for practitioners designing them. Without a thorough understanding of the domain, users might draw incorrect conclusions from the visualizations. Similarly, practitioners lacking in-depth knowledge of the data domain might create visualizations that are inaccurate or misleading.

We further expanded the meta-model by introducing an additional association with the 'DomainVariable' entity, enabling the representation of context within a visualization. Here, context is defined as supplementary information associated with a variable. For instance, household income might be linked to the COVID-19 incidence rate (Finch & Hernández Finch, 2020), and incorporating this information in a visualization about COVID-19 can provide a contextual layer to the presented data.

This enhancement is achieved through a reflexive association on the 'DomainVariable' class. This association offers the capability to identify and embody relationships between variables, whether they belong to the same domain or different ones, such as in cases where they are correlated.

The integration of this relationship to encapsulate the concept of context facilitates the consideration of additional, potentially significant variables in a visualization. This step is crucial before deciding on the technical attributes of the visualization, ensuring a more comprehensive and meaningful representation of data.

A key application of the dashboard meta-model is the MetaViz graphical instantiator (Vázquez-Ingelmo, García-Peñalvo, & Therón, 2022), which allows users to experiment with how various attributes, instantiated via the meta-model, affect the final visualization. Figs. 3 and 4 present an instance of a data visualization created using MetaViz. Users can adjust each element of the meta-model, with the resulting visualization automatically updating to reflect changes made to the workspace. This includes the instantiation of not only scales and visual marks, as shown in Fig. 3, but also interaction patterns, demonstrated in Fig. 4.

The dashboard meta-model provides a set of abstract categories and elements related to the design and conceptualization of data visualizations. For these reasons and to validate its content, this meta-model has been selected as the reference model for the present research. In this sense, the main entities of the meta-model have been extracted and adapted to carry out the analysis of the two focus groups.

3. Materials and methods

The study has been planned from the interpretative paradigm (Sarrado, Cléries, Ferrer, & Kronfly, 2004), characterized by the pursuit of understanding the phenomenon to achieve the general objective of determining which visualization characteristics impact its interpretation. Consequently, the qualitative methodology (Colás, 1994;

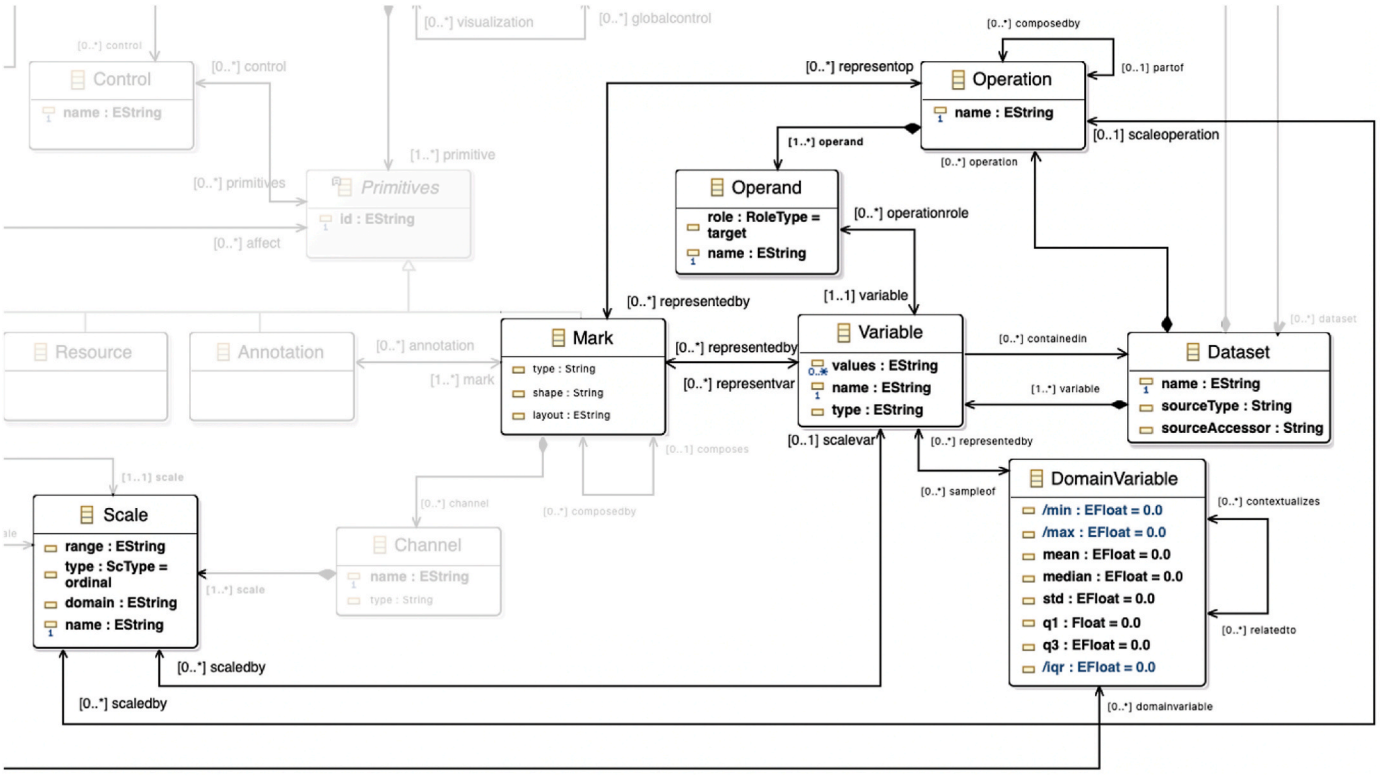


Fig. 2. Section of the meta-model that accounts for the relationship between the variables being visualized, their domain and their context. A high-resolution version can be consulted at https://zenodo.org/record/7037624.

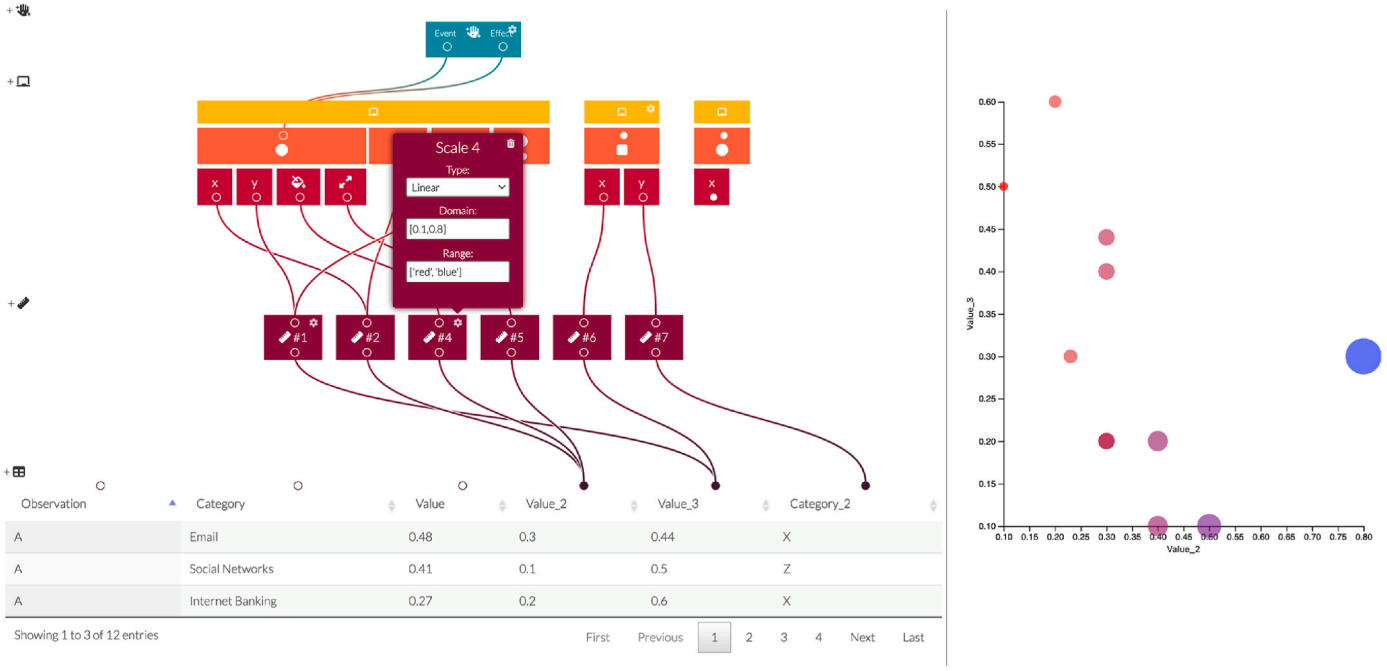
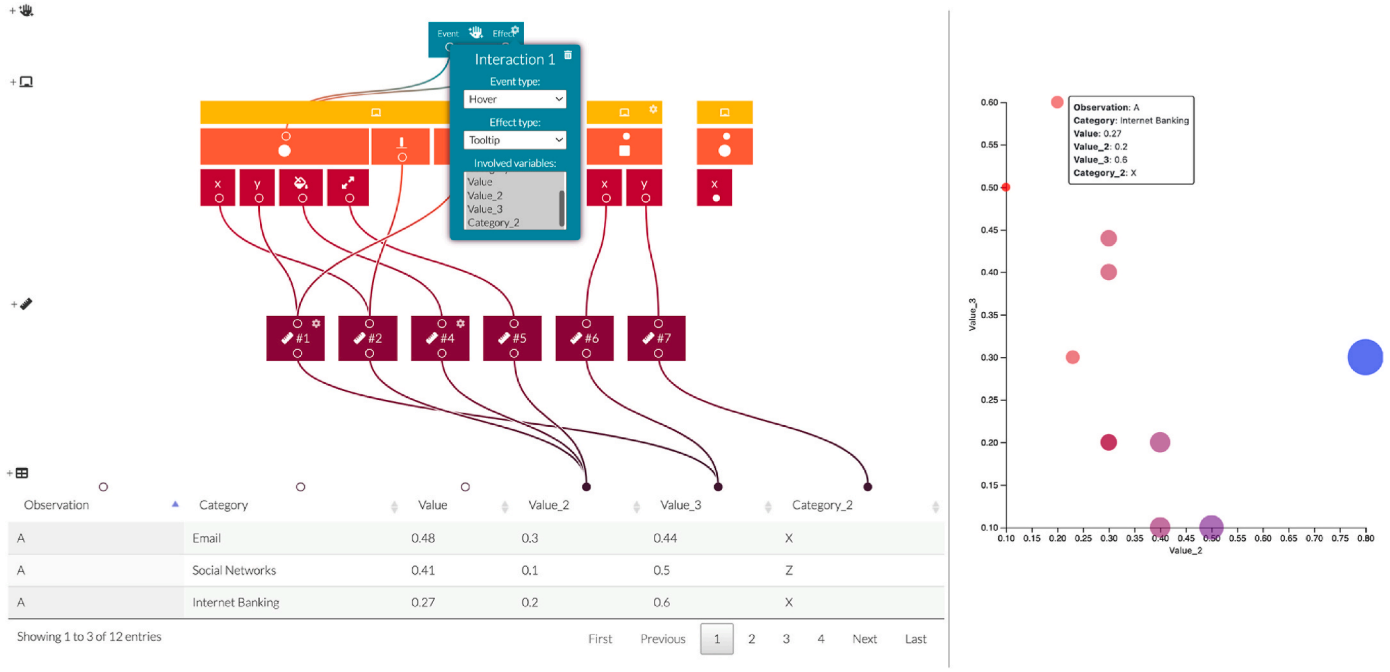


Fig. 3. Detail of a color scale connected to a data visualization channel. The scale functions by employing color encoding, where different colors correspond to different values of "Value_2." As a result, viewers can easily discern variations in "Value_2" through color differences.

Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2014) and the phenomenological method (Rodríguez Gómez & Gómez Ruiz, 2010) were used.

This methodological approach aims to answer the following research questions:

- **RQ1.** To what extent do the data visualization background and expertise of users affect the interpretation of data visualizations?
- **RQ2.** Which characteristics (scales, visual marks, textual elements, encodings, etc.) of data visualizations have an impact on their interpretation?



Observation	Category	Value	Value_2	Value_3	Category_2
A	Email	0.48	0.3	0.44	X
A	Social Networks	0.41	0.1	0.5	Z
A	Internet Banking	0.27	0.2	0.6	X

Showing 1 to 3 of 12 entries First Previous 1 2 3 4 Next Last

Fig. 4. Detail of an interaction pattern designed around a hover event coupled with a tooltip effect. Specifically, it is linked to the circle visual mark within the data visualization. When a user moves their cursor over any of these circle marks, the interaction is triggered, resulting in a tooltip appearing. This tooltip displays selected attributes of the data point associated with the hovered circle, providing users with additional, detailed information.

These two research questions are instrumental in achieving objectives **O1** and **O2**. As introduced, objective O1 aims to conduct a qualitative validation of the entities within the dashboard meta-model, while objective O2 seeks to examine the differences in interpretation and perspectives regarding data visualizations among groups with varied backgrounds.

This will be accomplished by gaining insights from participants' discussions in two ways. Firstly, by identifying the elements they reference during discussions and correlating these to the entities in the existing meta-model, as well as pinpointing new elements that may need to be incorporated into the meta-model. Secondly, through an analysis of the focus group transcriptions, we aim to discern the differences in the elements being discussed, as well as the nature of these discussions. This involves assessing the frequency and content of conversations, thereby highlighting how discussion topics and dynamics vary among participants with different backgrounds and levels of expertise.

3.1. Selection of data visualizations

The selection of data visualizations for this study was a key process, informed by the relevance and representativeness of the visualizations in conveying COVID-related information.

While any dataset could have been suitable for our study, we specifically selected one that would be familiar to all participants, irrespective of the profile for which they were selected. Our objective was to ensure that the problem domain of the visualized data did not unduly influence the discussion. Given the period in which the focus groups were carried out, COVID-19 was a prevalent topic, offering a common ground of understanding for all participants. This approach minimizes the variations in domain-specific knowledge among participants, allowing us to focus more on their interaction with the visualization tools and methods, rather than their understanding of the dataset's subject matter. It's important to note that while we assume a basic level of pandemic literacy, our primary interest is not in the depth of participants' knowledge about the pandemic, but in how they engage with and interpret the visualized data.

The primary criteria for the selection of data visualizations included:

1. **Relevance to the current epidemiological context**: Each visualization was chosen for its direct relation to the ongoing COVID-19 pandemic. This criterion ensured that the visualizations were not only timely but also of immediate interest to the participants.
2. **Source credibility**: We opted for visualizations from the Spanish mainstream media and the Spanish Ministry of Health.

In this sense, we selected and recreated the three following data visualizations. To enhance clarity, each data visualization was assigned a code that indicates the type of chart and whether it has been enriched or not, facilitating reference in later sections of the study.

- Line chart of daily confirmed cases of COVID-19 in Spain (Fig. 5; **Code:** V_LINE). This figure shows the daily confirmed cases of COVID-19 in Spain through a time series chart, using a line to encode the values.
- Bar chart of daily confirmed cases of COVID-19 in Spain (Fig. 6; **Code:** V_BAR). This figure presents the same data as the previous one, however, it utilizes bar graphs to encode the values of the daily confirmed cases of COVID-19.
- Map of COVID-19 cumulative incidence rate (as of September 15, 2020) in Madrid regions (Fig. 7; **Code:** V_MAP). This map employs a color-coded scale to represent the cumulative incidence rate of COVID-19 across the 21 regions of Madrid. The cumulative incidence rate is determined "by dividing the number of new cases in a specific period by the population initially free from the disease at the start of that period," as explained by the Spanish National Epidemiology Centre in their documentation.[1]

To test the influence of including context, we used supplementary data and included them in the previous data visualizations through different approaches, such as annotations (Kong & Agrawala, 2012;

[1] https://www.isciii.es/InformacionCiudadanos/DivulgacionCulturaCientifica/DivulgacionISCIII/Paginas/Divulgacion/DivulgacionInformeTerminosEpidemiologiaCoronavirus.aspx.

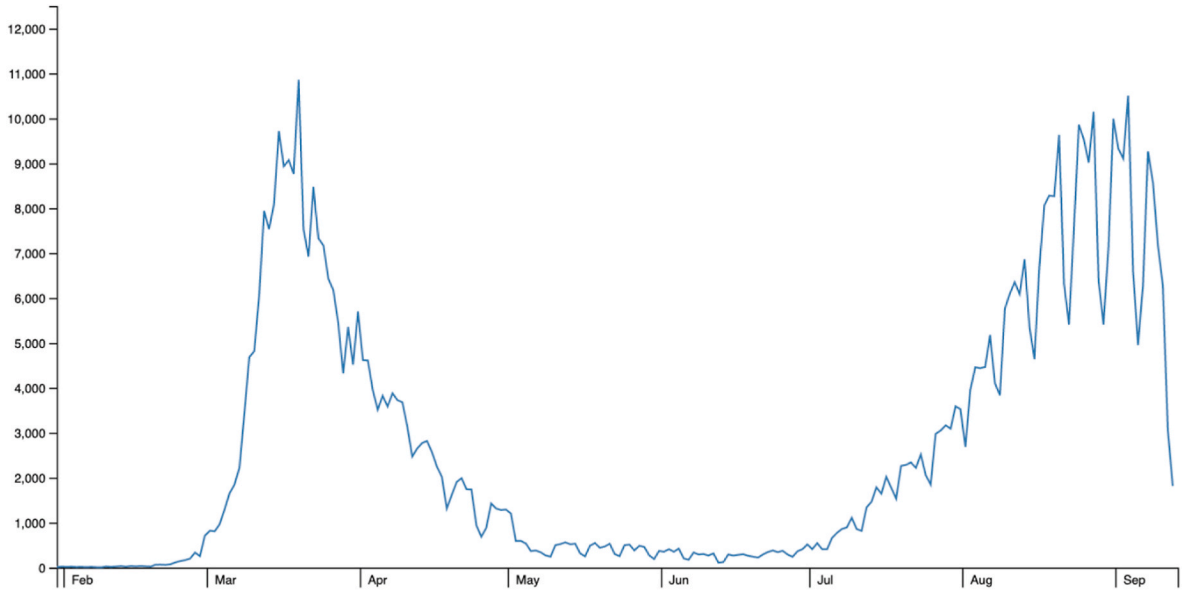


Fig. 5. Line chart of daily confirmed cases over time in 2020. Source: https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Actualizacion_252_COVID-19.pdf.

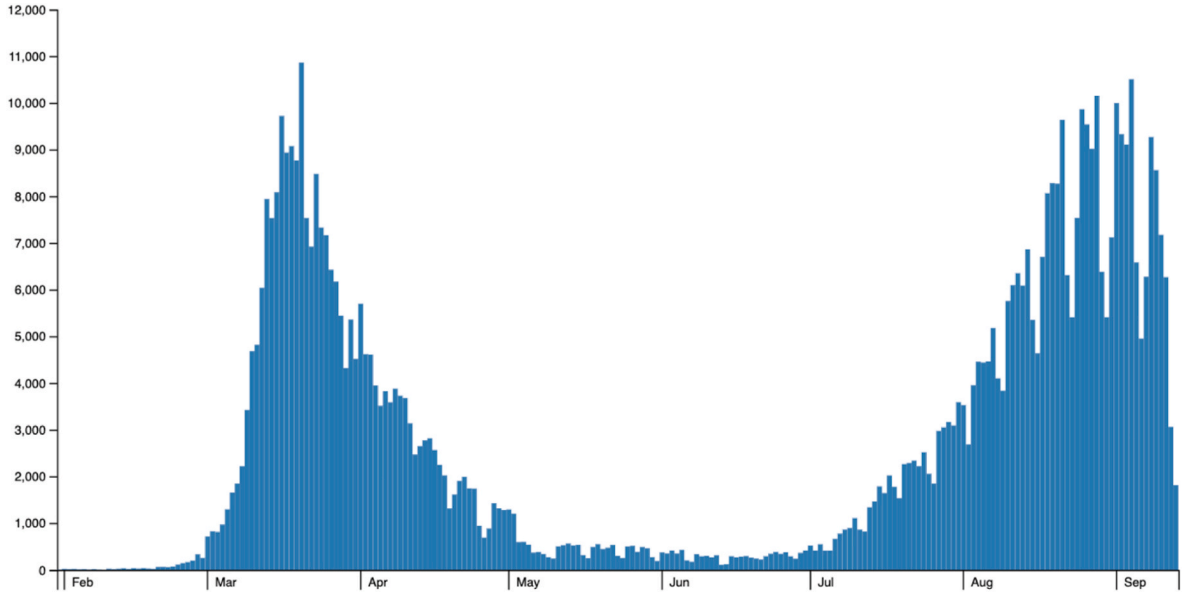


Fig. 6. Bar chart of daily confirmed cases over time in 2020. Source: https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Actualizacion_252_COVID-19.pdf.

Kong, Liu, & Karahalios, 2017) or composition (Javed & Elmqvist, 2012).

For the first visualization (Fig. 5), we incorporated contextual details regarding the Spanish Ministry of Health's approach to defining positive COVID-19 cases using various diagnostic methods.

Initially, up to May 10, both confirmed and probable cases were accounted for, encompassing all available laboratory techniques (PCR, ELISA, rapid tests, or antigen tests). Subsequently, from May 11, 2020, the criteria were refined to include only confirmed cases identified through PCR testing, and in certain instances, those diagnosed via ELISA.

In Fig. 8 (**Code:** V_LINE_ENR), we included this contextual information by employing a color coding and delineation technique (Kong & Agrawala, 2012; Kong et al., 2017). Consequently, the line in the visualization changes color to signify the shift in the case definition methodology.

In the case of Fig. 6, we enriched the displayed information by integrating details about the data collection methodology, particularly focusing on the reporting timelines of new positive COVID-19 cases. This aspect became crucial from July 4, 2020, onwards, as data updates were then only released on Mondays, encompassing all data collected over the weekend.

To visually represent this change in reporting frequency, we employed a color encoding technique in Fig. 6. Weekends were highlighted with a lighter shade to subtly indicate this period (Kong & Agrawala, 2012; Kong et al., 2017), thereby informing the viewer that weekend data, due to the lack of updates during this period, might not provide an accurate representation of the daily reporting trends (Fig. 9; **Code:** V_BAR_ENR).

Finally, we enriched Fig. 7 by using a juxtaposition strategy (Javed &

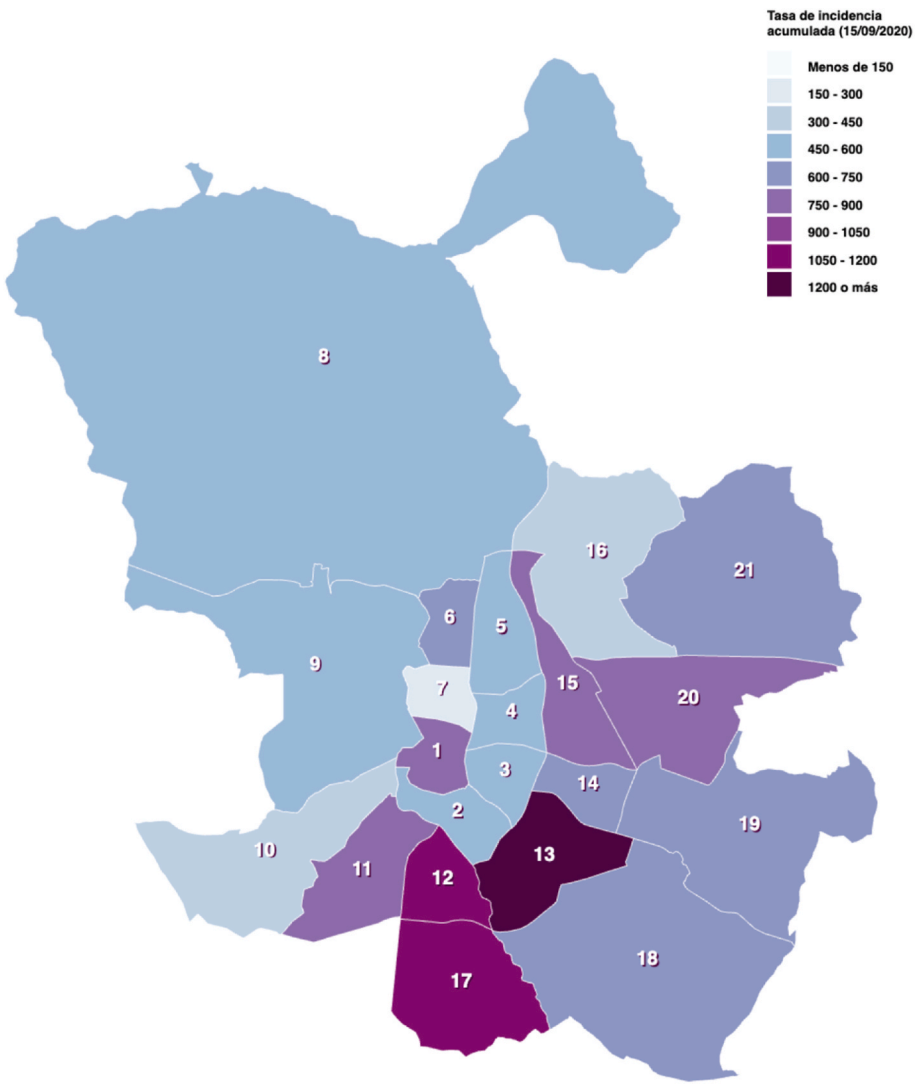


Fig. 7. Cumulative incidence rate of COVID-19 cases as of September 15, 2020 in Madrid regions. For our study, we assigned numerical values to each district in Madrid since the actual names of the regions were not relevant to our research objectives. This method made it easier for participants to refer to specific districts in the visualizations. Source: https://www.elmundo.es/madrid/2020/05/31/5ed0f630fc6c83e3498b4570.html.

Elmqvist, 2012) to visualize an additional contextual variable: the household income of the same Madrid regions, based on the most recent data available from 2017 at the time of conducting our study (Fig. 10; **Code:** V_MAP_ENR). The inclusion of this variable was inspired by various reports in Spanish media[2] that drew comparisons to highlight the influence of socioeconomic factors on the incidence of COVID-19.

3.2. Data collection techniques

The study was approached by applying focus groups. In addition, an adapted version of the IL-HUMASS questionnaire (García Llorente, Abad, & Conde, 2019) was distributed to complement the information, an instrument for measuring self-perceived information competencies. We selected this instrument because its dimensions and items have been validated across different studies in the Spanish context, with young people and adults with different profiles. According to Pinto (2009), the IL-HUMASS questionnaire was developed because of the need for data related to information literacy in higher education.

During the implementation of the focus groups, no personal data were collected to ensure data protection, thus guaranteeing the anonymity of the participants. The sociodemographic data of the participants were collected from the previously distributed survey (García Llorente et al., 2019).

One researcher pseudo-anonymized transcribed data along with the sociodemographic information of each participant. She altered all direct identifiers in the transcribed data and replaced them with non-identifiable codes, ensuring no link could be made between the transcripts and individual identities. Additionally, potentially identifying information was stored separately from the main data, so the researchers in charge of the analysis did not access this information.

Furthermore, an author who is not associated with the development of the meta-model coordinated the analysis without involving the person who performed the pseudo-anonymization. This was done to avoid any potential biases that could arise from the pseudo-anonymizer's familiarity with the data and from other authors' familiarity with the meta-model.

3.2.1. Development of the focus groups

First, the potential participants (researchers in Spain who are experts in data visualization and researchers in Spain who are pursuing a PhD in

[2] https://www.elsaltodiario.com/madrid/discriminaciones-distritos-desigualdad-pobreza-vivienda-transporte-coronavirus-segregacion-mata.

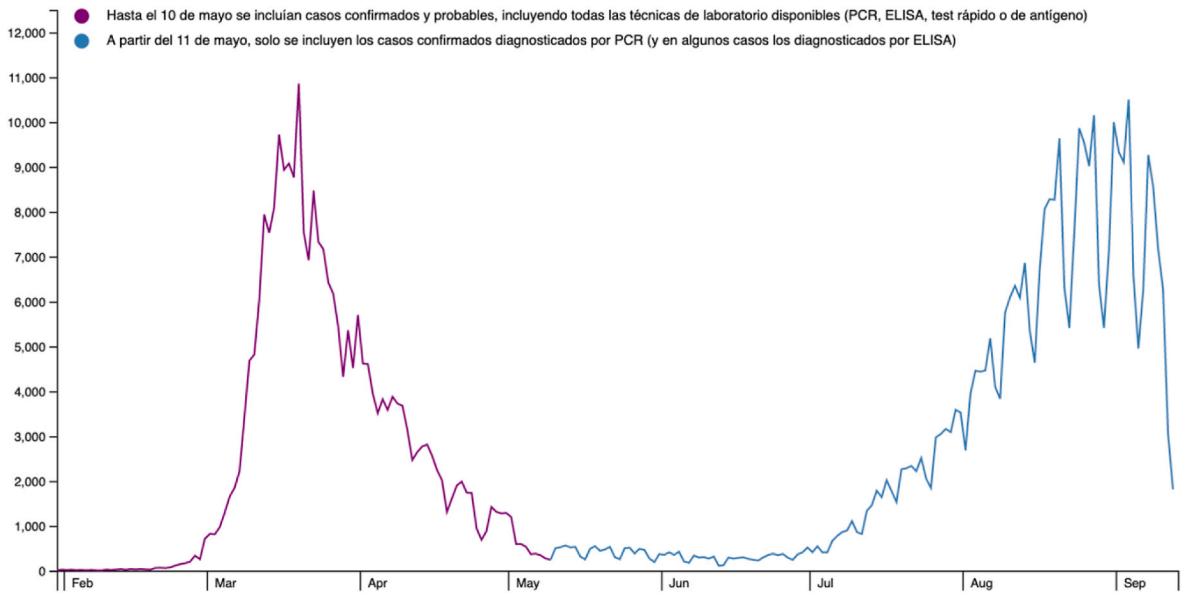


Fig. 8. The enriched version of the data visualization on COVID-19 daily reported positives. The legend reads: "*Until 10 May, confirmed and probable cases were included, including all available laboratory techniques (PCR, ELISA, rapid or antigen test)*", encoded in purple color, and: "*As of 11 May, only confirmed cases diagnosed by PCR (and in some cases those diagnosed by ELISA) are included*", encoded in blue color.

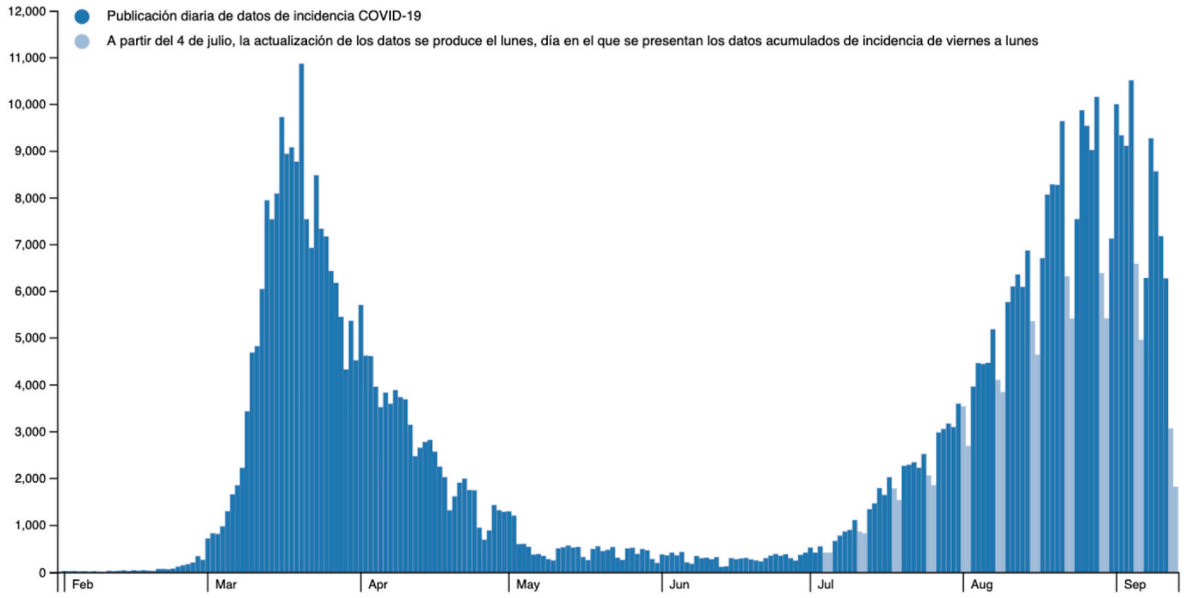


Fig. 9. The second enriched version of the data visualization on COVID-19 daily reported positives. The legend reads: *"Daily publication of COVID-19 incidence data"*, in opaque blue, and: *"From 4 July onwards, data are updated on Mondays, when the cumulative incidence data from Friday to Monday are presented"*, in a lighter shade of blue.

communication/information literacy) were contacted by email to invite them to participate in the study. This invitation contained information about the focus group study. These two population groups were selected to contrast the results concerning their background and expertise.

Seven data visualization experts and six PhD students from the University of Salamanca accepted the invitation. Subsequently, the questionnaire to measure self-perceived information literacy was shared among the selected participants, together with the information sheet and the informed consent form for the study. In addition, the email specified the day of the focus group and the format (online via Zoom).

Finally, two focus groups were held, one for each population group. They took place on **November 18, 2021** (with a duration of **1 h and 20 min**), and **December 21, 2021** (with a duration of **1 h and 43 min**). The focus groups consisted of group discussions (in Spanish) about three sets of data visualizations.

Initially, the visualizations did not include any context or specific data details. Moreover, it is noteworthy that **all participants lacked prior knowledge about the dashboard meta-model and its respective entities**. This lack of pre-existing awareness is critical, as familiarity with the meta-model entities could have introduced bias into the validation process.

The first set of data visualizations were displayed sequentially (Figs. 5–7). To start the conversation, the following question was asked regarding each displayed graphic: "What conclusions can you draw from this data visualization?".

Subsequently, after discussing the first set of visualizations, the

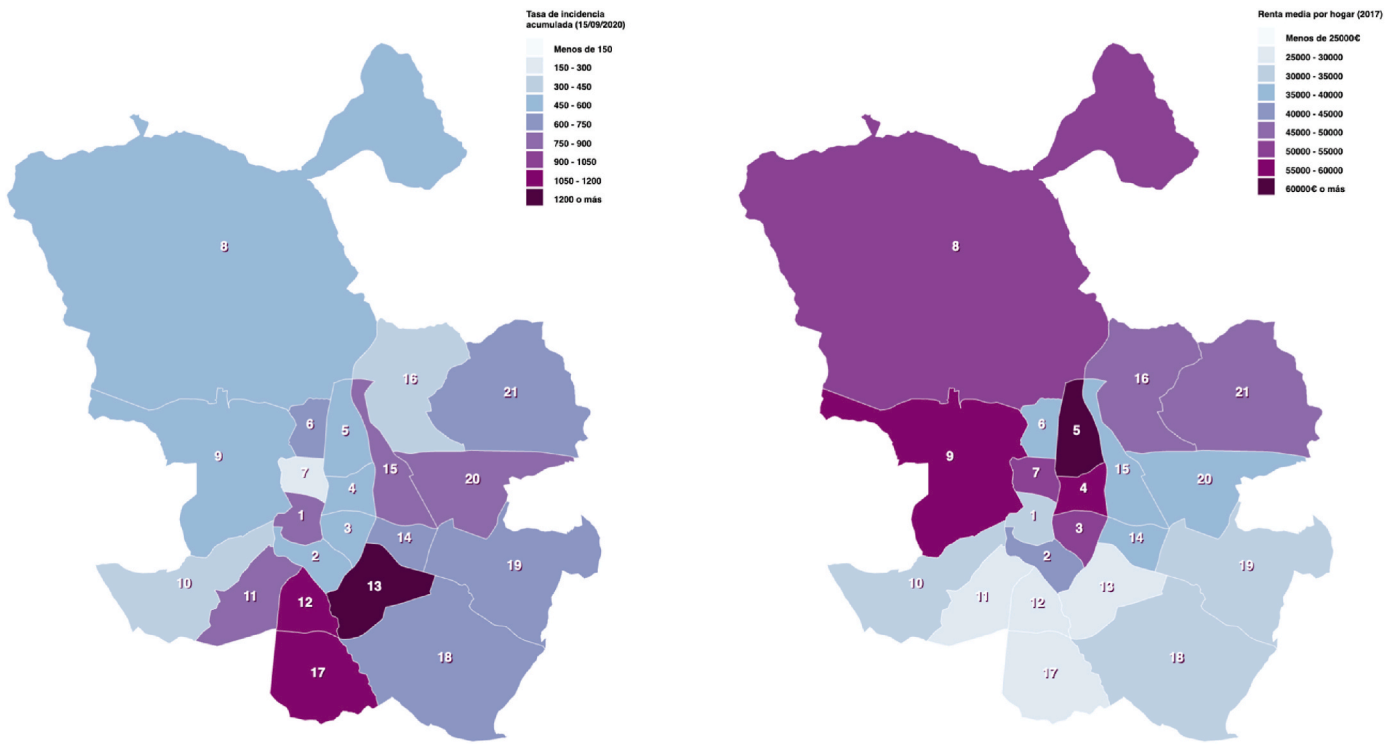


Fig. 10. The enriched version of the data visualization on COVID-19 cumulative incidence rate in Madrid regions. This version shows both the incidence rate and the average household income geographically.

enriched version of each one was shown to the participants sequentially (Figs. 8–10). The following question was asked regarding each enriched visualization: "Would you change your conclusions by looking at this new way of representing the same data?". Finally, participants were asked to select one of two versions of the same visualization and justify the decision made, with the option to choose neither of them.

3.3. Data analysis

After the conduction of the focus groups, the speeches were transcribed verbatim and a content analysis was carried out (Andréu Abela, 2000; Rodríguez Gómez & Gómez Ruiz, 2010). For this purpose, the focus groups were recorded after compiling the signed consent forms. It is important to remark that the study obtained a favorable report from the Bioethics Committee of the University of Salamanca, with registration number 553.

To reduce the data, firstly, the presented dashboard meta-model was taken as a reference and adapted to the content collected, following the categorical system. The adaptation process was carried out by consensus among the different authors of the study to achieve a rigorous and bias-free procedure.

Having established the meta-category (visualization), the categories (user characteristics, dataset, interaction, primitives, context, operation, and task), and the corresponding subcategories, these were then dumped onto a **frequency analysis** matrix, which can be consulted at https://zenodo.org/records/10556271 (Verdugo-Castro, García-Holgado, Vázquez Ingelmo, García-Peñalvo, & Therón, 2024). In this matrix, the categories were identified on the horizontal axis, using different colors according to the hierarchy of the categorical system (meta-category, categories, and subcategories) to facilitate the subsequent axial coding of the content.

Once the units of analysis were divided, the content of the focus groups was matched with the categories and numerical values were assigned according to each category's presence in the analysis unit. Two possible values were used, zero when the unit did not contain content about the corresponding category and one if the unit did contain content about the category. This information reports frequencies for each category to understand which categories were emphasized most.

After the frequency analysis, the **content analysis** was carried out, which can also be consulted at https://zenodo.org/records/10556271 (Verdugo-Castro et al., 2024). Following data reduction, the ideas, or units of analysis about each category were grouped and coded. In this way, the content analysis was completed with the coding of the units of analysis about the respective categories using the axial coding procedure (Anguera Argilaga, 1986). As with the reduction of the meta-model, an inter-judge procedure involving all authors was followed for the coding process to reduce biases and guarantee that the procedure would be consistent and scientifically rigorous. The profile of the authors of the work is visualization, human-computer interaction, and research methodology.

Additionally, in order to provide context to the qualitative findings, a detailed analysis was conducted on responses from a questionnaire designed to assess self-perceived information literacy. This questionnaire consists of eighteen items, each rated on a five-point odd Likert scale ranging from one (strongly disagree) to five (strongly agree). The analysis not only included the extraction of descriptive statistics for the entire sample but also segmented the data based on groups (experts and doctoral students) and gender (male and female).

4. Results and discussion

4.1. Participants

A non-probabilistic sampling technique was followed to obtain the sample of the study (Hernández Sampieri et al., 2014). The study aimed to have a group of researchers in Spain with expertise in data visualization, human-computer interaction, or data science. In addition, the other group focused on including data users at an expert level who also belonged to the research sector in Spain. In particular, PhD students working on doctoral thesis related to communication and/or

information literacy.

The sample is comprised of 13 participants, seven experts considered senior regarding their professional experience and research background (of the seven who were invited), and six doctoral students that considered junior researchers (of the thirteen who were invited) finally took part in the focus groups. Typically, focus groups comprise between six and ten participants, with an average of eight. This range is recommended because it balances the depth and variety of discussion without overwhelming the group dynamics or limiting individual participation (Ryan, Gandha, Culbertson, & Carlson, 2014).

In qualitative research, the emphasis is often on the depth and quality of data rather than statistical representativeness. Therefore, participants were chosen based on their background, aiming for a purposive sample that brings insightful perspectives and experiences to the discussion (Boddy, 2016). Additionally, the number is also strategically aligned with the principle of data saturation. In qualitative studies, data saturation is reached when no new information or themes are observed in the data. With 13 participants across two focus groups, this sample size is considered adequate to reach saturation, ensuring that the data collected are comprehensive and robust for answering the research questions (Saunders et al., 2018).

The resulting sample consisted of seven experts (3 women and 4 men) and six doctoral students (4 women and 2 men). Regarding age, two participants were between 20 and 29 years old, five between 30 and 39 years old, three between 40 and 49 years old, and three between 50 and 59 years old. The experts work on data visualization applied to different areas and research problems. On the other hand, selected doctoral students are data users; they have experience in data analysis, including the use of graphical representations, but they are not experts in data visualization. The aim of involving these two mutual exclusivity groups is to ensure that the meta-model is not only for experts in data visualization but also for experts who use data visualization in their context.

On the other hand, the sample was distributed throughout Spain: five people from Salamanca, two from Madrid, and the remaining participants from Granada, San Cristóbal de La Laguna, Málaga, Valladolid, and Zaragoza, respectively. There was also one Spanish person participating from London. Table 1 provides complete information on the profile of the participants.

4.2. Contextual analysis of the sample

The answers obtained by the participants in the IL-HUMASS questionnaire to measure self-perceived information literacy show favorable results, thus showing high self-perceived information competences, as seen in Fig. 11. The items measured by the questionnaire are the following (García Llorente et al., 2019):

Q01. I can identify various internet search engines

Q02. I am able to access information through links or hyperlinks

Q03. I am able to filter and personalize my internet searches

Q04. I am able to modify and configure different search methods in Internet search engines

Q05. I can judge the reliability of content I find on the Internet

Q06. I know that there are websites where false information can be found

Q07. I know that the sources of information must be verified

Q08. I know how to transform information into knowledge

Q09. I am able to evaluate whether information is useful or useless

Q10. I can list different data storage devices and choose the most appropriate one

Q11. I am able to download and upload files to the internet

Q12. I am aware of the importance of backing up the data that I store

Q13. I am aware of the consequences of storing downloaded content privately and publicly

Q14. I can name different forms of virtual communication

Q15. I know that when I browse the internet, I leave traces of my personal information

Q16. I can filter the information I receive or decide who follows me on social networks

Q17. I am able to report verbal abuse on social networks or forums

Table 1
Profile of participants. Source: Own elaboration.

Code	Group	Gender	Age	Experience	Employed	What studies have you studied or are you currently studying?
GFEXP_01	Expert	Woman	40 to 49	Full professor	Yes	Bachelor's Degree in Computer Science. PhD in Computer Science
GFEXP_02	Expert	Woman	50 to 59	Professor	Yes	Bachelor's Degree in Computer Science. PhD in Computer Science
GFEXP_03	Expert	Man	30 to 39	Postdoctoral Researcher	Yes	Bachelor's Degree in Computer Science. Master's Degree in Intelligent Systems. PhD in Computer Science
GFEXP_04	Expert	Man	30 to 39	Clinical Data Scientist	Yes	Bachelor's Degree in Industrial Engineering. Master's Degree in Statistics. PhD in Statistics
GFEXP_05	Expert	Man	30 to 39	Software Engineering Manager	Yes	Bachelor's Degree in Computer Science. Master's Degree in Intelligent Systems. PhD in Computer Science
GFEXP_06	Expert	Man	40 to 49	Professor	Yes	Bachelor's Degree in Telecommunication Engineering. PhD in Management Information Systems
GFEXP_07	Expert	Woman	40 to 49	Full Professor	Yes	Bachelor's Degree in Computer Engineering. PhD in Computer Engineering
GFDOC_01	PhD student	Man	30 to 39	Fourth year in the PhD. Photographer and community manager	No	Bachelor's Degree in Communication. Master's Degree in Communication
GFDOC_02	PhD student	Woman	50 to 59	Second year in the PhD. Methodological and e-learning manager	Yes	Bachelor's Degree in Marketing and Communication. Master's Degree in e-learning
GFDOC_03	PhD student	Woman	30 to 39	Fifth year in the PhD. Bioinformatician	Yes	Bachelor's Degree in Biotechnology. Master's Degree in Bioinformatics and Genetics
GFDOC_04	PhD student	Woman	20 to 29	Second year in the PhD. ICT coordinator	Yes	Bachelor's Degree in Education. Master's Degree in.Technology in Education
GFDOC_05	PhD student	Man	20 to 29	Third year in the PhD. Journalist	Yes	Bachelor's Degree in Business Administration and Journalism. Master's Degree in Communication
GFDOC_06	PhD student	Woman	50 to 59	Sixth year in the PhD. Librarian	Yes	Bachelor's degree in Documentation

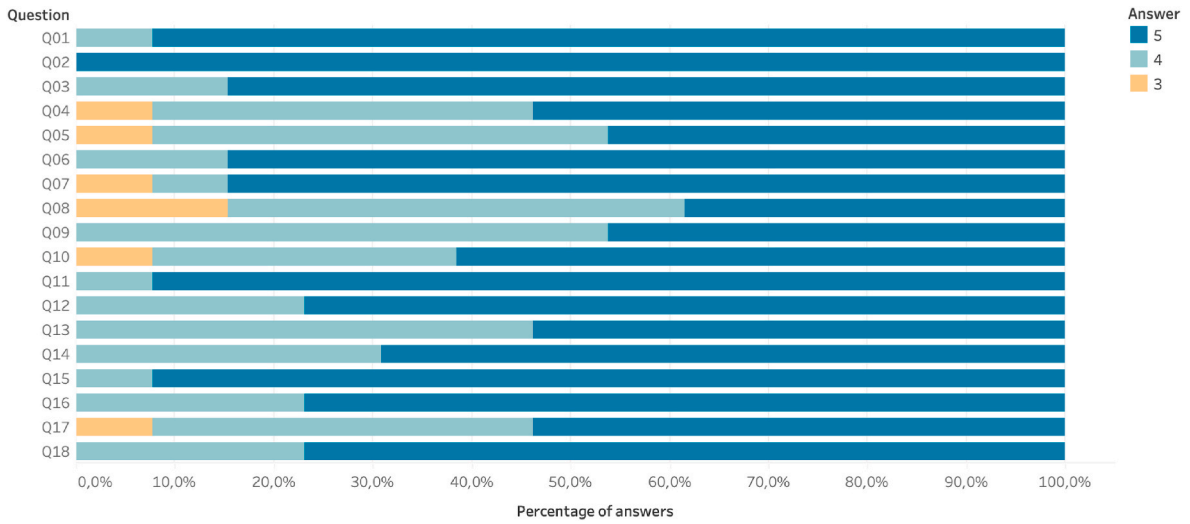


Fig. 11. Questionnaire results for the entire sample. Source: Own elaboration.

Q18. I am aware that there are risks and benefits to exposing my virtual identity

The item with the best result is: "I am able to access information through links or hyperlinks", while the item with the lowest result is: "I know how to transform information into knowledge".

Regarding the slight differences between groups, those items linked to the management of social networks have obtained better average results for the group of doctoral students (Fig. 12). This could be explained by age, as they are younger than the experts, and it is assumed that they are able to spend more time on social networks. Meanwhile, for the items aimed at Internet searches and devices, the higher average results have been for the visualization experts.

As it can be perceived, the differences between the two groups regarding their information literacy is very low, indicating that the diverse perspectives in the discussion were likely influenced more by their varied backgrounds rather than by differences in their information literacy skills.

4.3. Frequency and content analysis

Going deeper into the content analysis of the group of visualization experts, the categories that show the most coding are "visualization", the meta-category of the categorical system, "user characteristics" and "primitives". Certain subcategories, such as "preferences" and "axes," have not been coded because they were not mentioned by the experts during the discussions. Comprehensive details regarding the frequency of category occurrences throughout the experts' focus group interventions are presented in Fig. 13.

Following the results, the main concepts discussed by the experts involved the whole visualization, the user characteristics, the primitives of the visualization, and the selected scales to encode the information, which are mostly technical aspects. In general, experts focused on the technical aspects of the visualizations. For instance, discussions regarding the selected scales, colors, paddings, etc. in the figures were recurrent.

Interventions such as "the scale is a bit problematic" (*GFEXP_06* about V_LINE), "using bars gives me more of a feeling that they are real cases" (*GFEXP_02* about V_BAR), "I believe this (scale) is color-blind safe" (*GFEXP_03* about V_MAP), among other, suggest that experts were more concerned about the design of the lower-level elements of the visualizations. The interactivity of the figures was also alluded to. Expert *GFEXP_06* pointed out that "we have become so used to digitally consulting interactive graphics that these things fall a bit short".

Although technical aspects were discussed mainly, non-technical aspects were also mentioned. The V_MAP visualization (Fig. 7), for

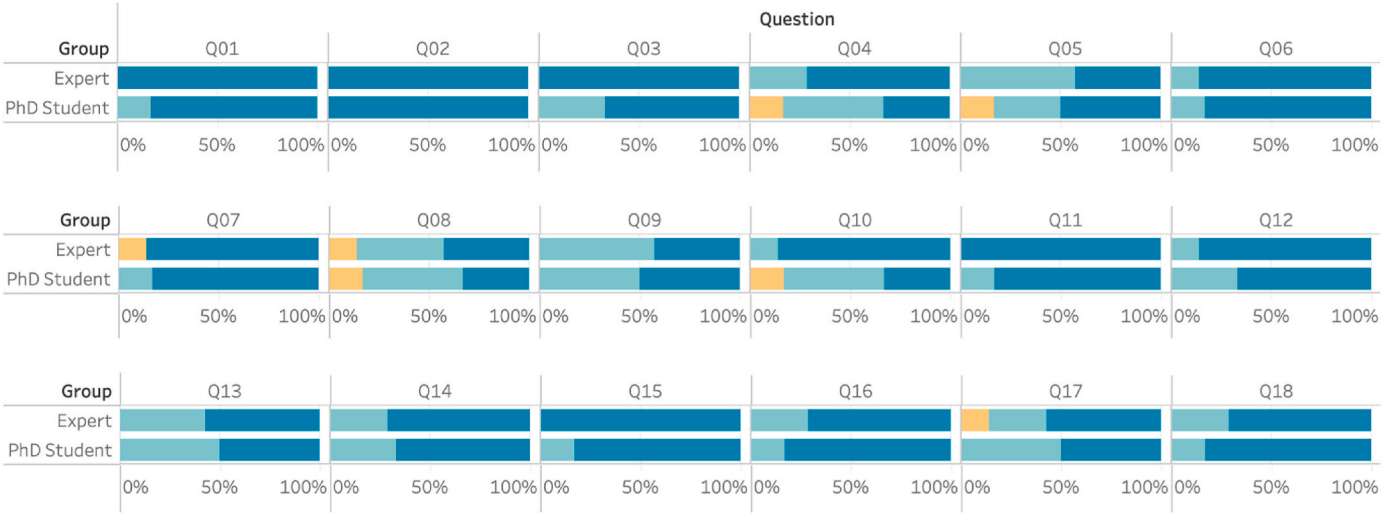


Fig. 12. Questionnaire results by group. Source: Own elaboration.

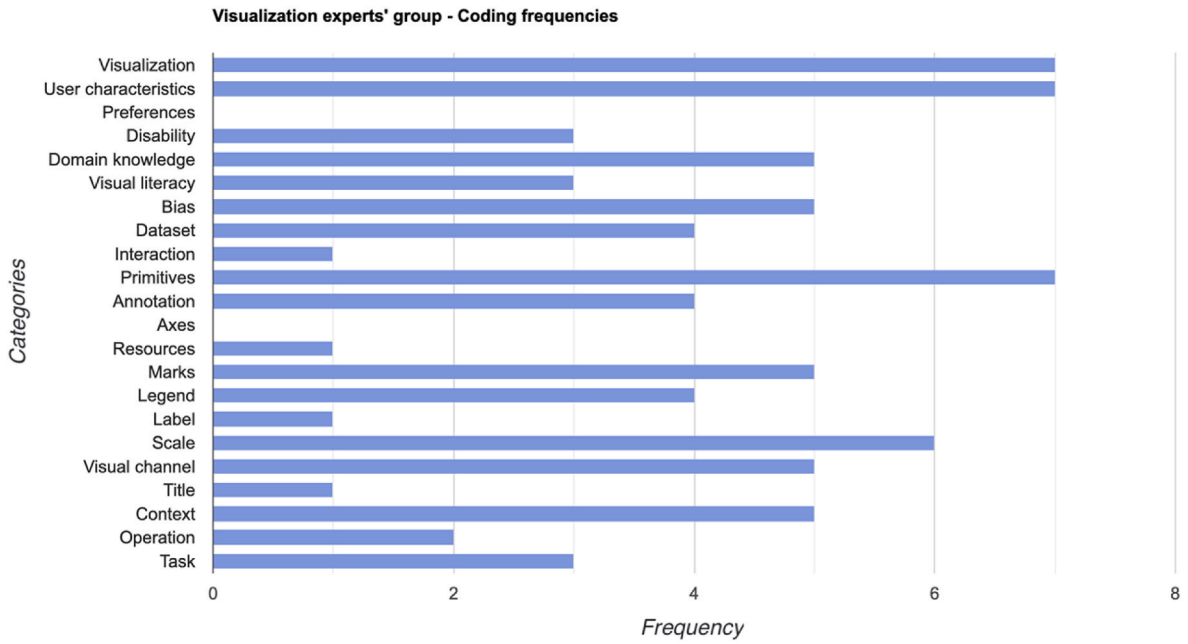


Fig. 13. Frequencies of coding on the categories for the visualization experts' group. Source: Own elaboration.

example, generated controversy related to the lack of context. *GFEXP_06* indicated that they "know that this is the map of districts of Madrid, but also because I am from Madrid. We don't really know where this map is from, right?". One expert who was not originally from Spain was confused about the information shown.

Finally, it is relevant to remark on how many participants in the group preferred one variant of the presented data visualizations rather than the other to convey information. It is important to highlight that they had the option of selecting none of them. Table 2 shows the number of votes in favor of each figure.

In this sense, the group had a clear preference during the last two blocks of images, and although it is true that in the first block the distribution was more evenly distributed, there was a preference for the first figure.

While the second version of each block of figures added more context, experts preferred the first version (with fewer details) in two of the three sets of visualizations. The majority justified this selection due to the design of the "enhanced" version, which they pointed out as "misleading", "confusing" and "cluttered". Following the results of the experts' group, it is concluded that more context does not produce better visualizations by default and that it is necessary to design proper elements to enhance this contextualization.

On the other hand, for the group of doctoral students, the categories with the most coding frequencies are also "visualization", the meta-category of the categorical system, "user characteristics", and "primitives". In this sense, this result coincides with those obtained for the group of experts. However, the category "interaction" and the sub-categories "preferences", "disability", and "label" are not coded, as the content of the discourses was not related to the units of analysis. Comprehensive details regarding the frequency of category occurrences throughout the doctoral students' focus group interventions are presented in Fig. 14.

Table 2
Selection of figures for each block for the group of visualization experts. Source: Own elaboration.

Selection of figures	Experts
V_LINE – V_LINE_ENR	**V_LINE: 5 participants** V_LINE_ENR: 2 participants
V_BAR – V_BAR_ENR	V_BAR: 0 participants **V_BAR_ENR: 7 participants**
V_MAP – V_MAP_ENR	**V_MAP: 6 participants** V_MAP_ENR: 0 participants No answer: 1 participant

We can see that the main concepts discussed in this focus group were also related to the whole visualization, the user characteristics, and the primitives of the visualization, in addition to the legend, and more conceptual factors such as the context.

In this instance, the interventions generally highlighted concerns regarding the absence of context. For example, as *GFDOC_06* indicated about V_BAR, "a decontextualized graphic means nothing", so they were not able to draw robust conclusions from the figures. In general, the participants of this focus group considered these visualizations as a method of explanation that need to be accompanied with resources (texts, images, etc.) to add more information and context, because on the contrary, "one graph could potentially be leveraged as a means to escalate tensions and to attack (someone)'' (*GFDOC_02* about V_BAR_ENR).

The V_MAP visualization (Fig. 7) generated discussions related to the lack of information about the data displayed (in this case, the districts of Madrid). Following the same ideas of the experts' group, participants pointed out that they knew the map was from Madrid because of their previous knowledge, but foreign people could be confused about the visualization if they do not know Spain's geography. In fact, their existing domain knowledge was alluded to during the discussion of most figures; regarding V_LINE, for example, *GFDOC_03* indicated that "we think about the peaks at weekends because we know how the data has been given".

Doctoral students were also asked to choose the visualization that they preferred to convey information. Table 3 shows the number of votes in favor of each figure.

Comparing the results with the group of experts, doctoral students had more defined preferences. However, for the group of doctoral students, the distribution of responses was more evenly distributed, with a greater number of participants not leaning toward either option for each block of images.

Contrary to the experts, doctoral students preferred the enriched version of every visualization. They justified the selection because they valued the inclusion of more information into the visualizations, no matter the design decisions that were taken to include it. In fact, the

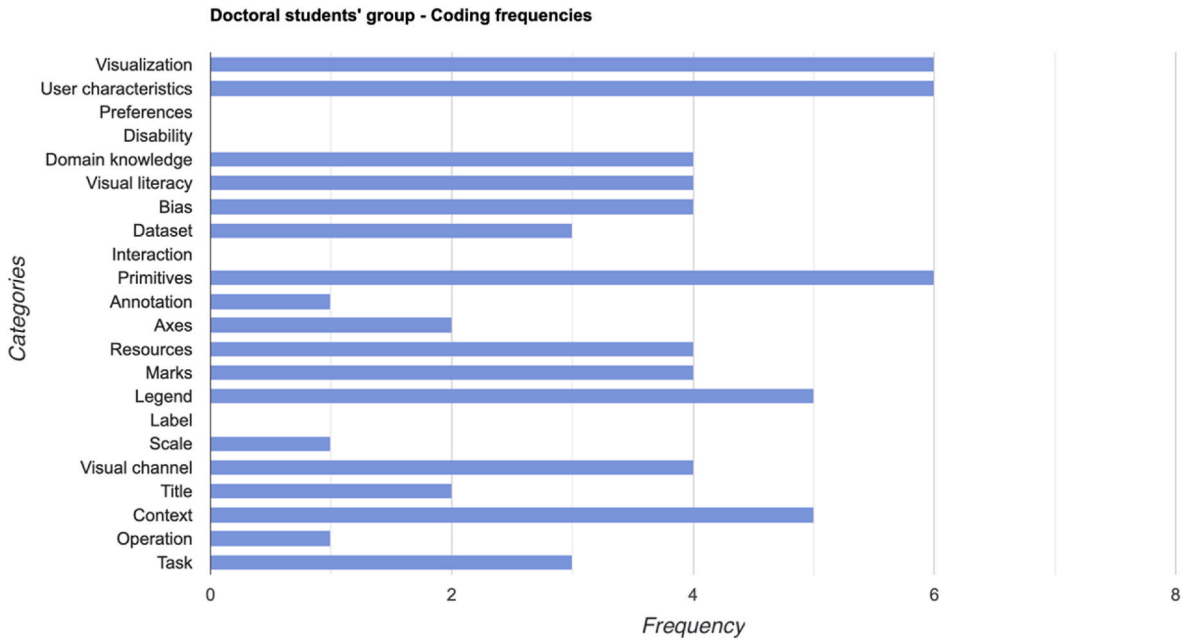


Fig. 14. Coding frequencies of the categories for the group of doctoral students. Source: Own elaboration.

Table 3
Selection of the figures in each block for the group of doctoral students. Source: Own elaboration.

Selection of figures	Ph.D. students
V_LINE – V_LINE_ENR	V_LINE: 0 participants **V_LINE_ENR: 5 participants** No answer: 1 participant
V_BAR – V_BAR_ENR	V_BAR: 0 participants **V_BAR_ENR: 3 participants** No answer: 3 participants
V_MAP – V_MAP_ENR	V_MAP: 0 participants **V_MAP_ENR: 4 participants** No answer: 2 participants

weight that doctoral students gave to context during the discussions aligns with this decision.

To conclude the analysis, it is evident that the "visualization" meta-category achieves the highest frequency in terms of coding across both groups (Fig. 15). The same is true for the categories of "user characteristics" and "primitives". Other categories have lower frequencies, such as "context", "dataset", "task", "operation", and "interaction". Also, the subcategory of "preferences" has not been coded during the analysis, which means that the content and units of analysis obtained are not associated with this meta-model concept.

The original transcription in Spanish, along with its English automatic translation, frequency, and content matrices, are available for detailed consultation at this repository: https://zenodo.org/records/10556271.

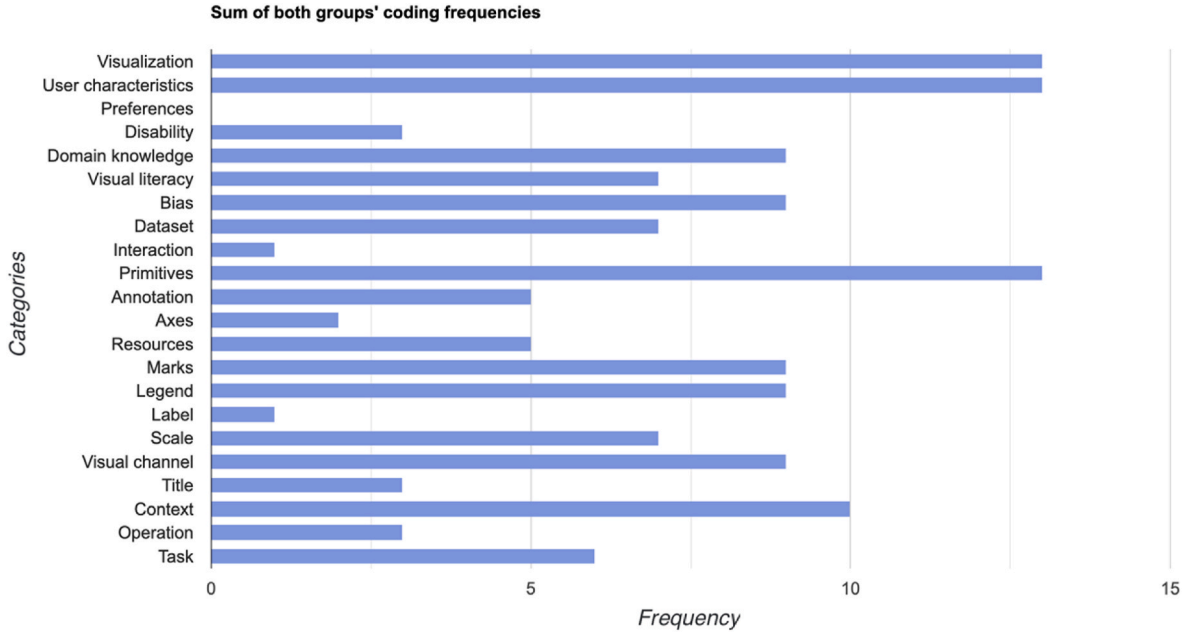


Fig. 15. Coding frequencies of the categories for the two groups. Source: Own elaboration.

4.4. General discussion

Following the analysis and the discussion of the results, it is possible to answer the two research questions posed at the beginning of the study. On the one hand, regarding **RQ1** (To what extent do the data visualization background and expertise of users affect the interpretation of data visualizations?) the analysis showed some differences between the two groups.

These differences can be observed in the frequency analysis. For instance, "interaction" (a technical aspect) and "disability" (related to technical aspects such as the color palette) were only discussed in the experts' group, and "resource" (a category related to giving context to data visualization through text or images) was more discussed in the doctoral students' group.

Moreover, there were also differences when participants were asked to choose one of the two figures from each block. While experts rejected two of the three enriched versions (V_LINE_ENR and figure V_MAP_ENR), doctoral students voted them in favor, which suggests that doctoral students preferred figures that displayed more information (although they could lack effectiveness due to their design, as noted by the experts).

In this sense, one of the primary revelations of our study is the profound impact that background and expertise have on the interpretation of data visualizations. We observed that experts in data visualization focus more on technical aspects such as design, scales, and color choices. In contrast, doctoral students, perhaps less entrenched in technical norms, emphasized the contextual aspects and narrative potential of the visualizations. This distinction is crucial for practitioners and educators in the field, suggesting that the design and presentation of data visualizations should be tailored to the specific background and expertise of the target audience. This approach can lead to more effective communication and understanding of complex data.

On the other hand, **RQ2** (which characteristics -scales, visual marks, textual elements, encodings, etc.- of data visualizations have an impact on their interpretation?) is answered through the frequency analysis carried out in the experiment (Fig. 15). In this sense, the main discussed concepts were the visualizations (as a whole), their primitives, and the characteristics of the audience. This outcome makes sense as these categories are high-level elements, and they are pointed out in general during the discussions.

However, both tangible and conceptual elements were also identified regarding their impact on the interpretation of data visualizations. Specifically, context, domain knowledge, bias, and visual literacy were recurrent in the interventions of the participants from both groups, suggesting that these factors are crucial when designing effective and useful visualizations.

The emphasis on high-level elements such as "visualization", "primitives", and "audience characteristics" reveals a broad-based strategy for interpreting data visualizations. However, the recurrent mention of context, domain knowledge, bias, and visual literacy in discussions by both groups underline these as critical factors influencing interpretation. These elements highlight the multifaceted nature of data visualizations, where understanding extends beyond mere graphical comprehension to include contextual awareness, domain-specific knowledge, and the ability to discern potential biases. The importance of visual literacy also suggests a need for education and training in reading and understanding visual data, which is increasingly becoming a crucial skill in the information age.

Our findings carry significant insights for the design of data visualizations. The preference of experts for less detailed but technically precise visualizations, and doctoral students for more contextually rich versions, underscores the need for a balanced approach in visualization design. Designers must strive to marry technical accuracy with contextual information, ensuring that visualizations are not only accurate but also narrative-rich and engaging for a broader audience.

The study highlights visual literacy as a critical skill in interpreting data visualizations. This has important educational implications, suggesting a need for incorporating visual literacy training in educational curricula. Enhancing visual literacy can equip students and professionals with the necessary skills to effectively interpret and create meaningful visual representations of data.

Finally, the emergence of practically every meta-model entity in the interventions of the participants from the two focus groups, and the fact that no new categories emerged from this experiment, proves that the dashboard meta-model is comprehensive and captures useful concepts about this discipline. The only not mentioned category in both focus groups was "(user) preferences". This can be explained by the fact that the group discussions were continuously focused on objective aspects of the figures, and this may have caused participants to set aside user preferences, which are more subjective and related to customization capabilities (Vázquez-Ingelmo, García-Peñalvo, et al., 2019).

The dashboard meta-model, as validated by our study, emerges as an effective roadmap for guiding the design of data visualizations. It encapsulates essential elements that span both technical and contextual aspects, highlighting its utility in creating visualizations that are not only technically sound but also contextually rich. This comprehensive approach is crucial in addressing the diverse needs and preferences of different user groups, as identified in our study.

Our findings indicate a dichotomy in the focus of experts and doctoral students: while experts tend to prioritize technical aspects, doctoral students are more inclined towards the contextual elements of data visualizations. However, the meta-model encourages a holistic approach, integrating both these aspects into the design process. This integrated approach is vital in creating visualizations that are not only accurate and easy to interpret for technical experts but also informative and engaging for audiences with a more contextual orientation, like doctoral students.

In practical terms, the meta-model serves as a comprehensive checklist and guideline for designers. It prompts them to consider a wide array of factors ranging from color schemes and scale (appealing to technical experts) to narrative elements and context (resonating with non-experts). For instance, while designing a data visualization, one might start with ensuring technical accuracy and clarity, and then layer in contextual information and storytelling elements to make the data more relatable and understandable for a broader audience.

The application of the meta-model in the design process can bridge the gap between users with varying levels of expertise and backgrounds. By offering a structured approach that encompasses both technical and contextual elements, designers can create visualizations that cater to the nuanced needs of different audiences, thereby enhancing the overall effectiveness and reach of their visual communications.

5. Limitations

The limitations identified in this study, while intrinsic to its design and scope, should be acknowledged and justified for a comprehensive understanding of the research context and implications.

Firstly, the participant diversity and sample size, primarily consisting of researchers and PhD students from Spain, present a limitation in terms of representativeness. The rationale behind choosing this specific group was to contrast the perspectives of data visualization experts and those in the initial stages of their academic careers. However, this selection may not capture the wide range of experiences and interpretations prevalent in a more diverse audience. Furthermore, the small sample size, although sufficient for qualitative depth, limits the generalizability of the findings to a broader population.

The focus on COVID-19 data visualizations was a deliberate choice, given the contemporaneous relevance and universal familiarity with the topic among participants. This context provided a common ground for discussion, enabling a more focused analysis of data visualization interpretation. Nonetheless, this specificity could lead to biased insights, as participants' pre-existing knowledge and opinions about the

pandemic might influence their perceptions.

The order of visualization presentation is a critical factor to consider. Participants viewed the line and bar charts in a sequence, which could potentially influence their evaluations. This sequential exposure might result in an anchoring effect, where the first visualization sets a reference point for subsequent ones. The decision to present visualizations in a particular order was guided by the aim to progressively build on the complexity and type of data representation, and to avoid the effects of these limitations, the same order was followed for the two focus groups.

The potential for confirmation bias arises from the study's intent to validate the pre-established dashboard meta-model. The research team was aware of this risk and took measures to maintain objectivity. Understanding the importance of an unbiased approach, we specifically chose a focus group methodology in which our role was strictly limited to moderation without intervention. This approach was essential to ensure that the participants' discussions and insights remained uninfluenced by our perspectives or preconceptions about the meta-model.

Furthermore, to reinforce the integrity of our methodology and the subsequent analysis, we engaged an author specifically responsible for ensuring methodological correctness. This author, who played a pivotal role in coordinating the analysis, had no prior involvement in the definition or validation of the meta-model. This separation of roles was a deliberate decision to reduce the potential influence of any pre-existing biases or expectations regarding the meta-model.

Finally, the limited scope of technical aspects examined is a reflection of the study's focus. It aimed to balance depth with breadth, exploring a range of technical and contextual elements without becoming overly exhaustive. Future research can build on this foundation by delving into other design elements and user interaction aspects.

6. Conclusions

This work describes the results obtained from the analysis of two focus groups about data visualizations. Both groups had different backgrounds to, additionally, analyze differences between the interventions of the participants; one group was composed of data visualization experts, while the other was composed of PhD students.

The analysis of the interventions provided interesting insights. On the one hand, experts were more conscious of the technical aspects of data visualizations and more critical of the potential enhancements applied to the different figures employed in the experiment, which consisted of including more context through legends and complementary information.

On the other hand, doctoral students focused more on the lack of context and the necessity of including more information through textual elements. In this sense, doctoral students preferred the "enhanced version" of each data visualization, as they included more context, although the visual elements included were not the most effective or useful option.

Finally, this work has also served to validate the dashboard meta-model developed in (Vázquez-Ingelmo, García-Holgado, et al., 2021; Vázquez-Ingelmo et al., 2019; Vázquez-Ingelmo, García-Peñalvo, Therón, & Conde, 2020; Vázquez-Ingelmo, García-Peñalvo, Therón, et al., 2019; Vázquez-Ingelmo, García-Peñalvo, Therón, & García--Holgado, 2020). The analysis proved that the meta-model is comprehensive and contains the main concepts to consider when developing and designing data visualizations and information dashboards. It also underscores the model's utility as a comprehensive guide, capable of capturing the multi-dimensional nature of visual data representation.

While the findings of this study give valuable insights into the perspectives and thoughts of data visualization specialists and Ph.D. students, some limitations must be acknowledged, mostly related to the demographics of the two focus groups. The findings of this study are constrained to the characteristics of the participants of the two focus groups, meaning that these results are not generalizable for users with other backgrounds. To address these limitations, future work will focus on conducting similar research with participants from other contexts and domains, e.g., people from other disciplines such as humanities or arts, lay users, etc.

Funding

This research was partially funded by the Spanish Ministry of Science and Innovation through the AVisSA project grant number (PID2020-118345RB-I00). The authors also want to thank the participants of the study.

CRediT authorship contribution statement

Andrea Vázquez-Ingelmo: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Visualization, Writing – original draft. **Alicia García-Holgado:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft. **Sonia Verdugo-Castro:** Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Writing – original draft. **Roberto Therón:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing. **Francisco José García-Peñalvo:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset generated and analyzed for this study can be found at https://zenodo.org/records/10556271

References

Andréu Abela, J. (2000). Las técnicas de Análisis de Contenido: Una revisión actualizada. *Fundación Centro Estudios Andaluces, Universidad de Granada, 10*(2), 1–34.

Anguera Argilaga, M. T. (1986). La Investigación cualitativa. *Educar*, 23–50. https://raco.cat/index.php/Educar/article/view/42171.

Boddy, C. R. (2016). Sample size for qualitative research. *Qualitative Market Research: An International Journal, 19*(4), 426–432.

Börner, K., Maltese, A., Balliet, R. N., & Heimlich, J. (2016). Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization, 15*(3), 198–213. https://doi.org/10.1177/1473871615594652

Cairo, A. (2012). The Functional Art: An introduction to information graphics and visualization. *New Riders*.

Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. *New Riders*.

Cairo, A. (2019). *How charts lie: Getting smarter about visual information. WW Norton & Company*.

Colás, P. (1994). La Metodología Cualitativa en España. Aportaciones científicas a la educación. *Revista Bordón, 46*(4), 407–423, 1994.

Correll, M., Bertini, E., & Franconeri, S. (2020). In *Truncating the y-axis: Threat or menace? Proceedings of the 2020 CHI conference on human factors in computing systems*. Honolulu, HI, USA.

Dimara, E., Franconeri, S., Plaisant, C., Bezerianos, A., & Dragicevic, P. (2018). A task-based taxonomy of cognitive biases for information visualization. *IEEE Transactions on Visualization and Computer Graphics, 26*(2), 1413–1432.

Few, S. (2006). *Information dashboard design*. O'Reilly Media.

Finch, W. H., & Hernández Finch, M. E. (2020). Poverty and covid-19: Rates of incidence and deaths in the United States during the first 10 weeks of the pandemic. *Frontiers in Sociology, 5*, 47.

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest, 22*(3), 110–161. https://doi.org/10.1177/15291006211051956

García Llorente, H. J. G., Abad, F. M., & Conde, M. J. R. (2019). Validación de un instrumento de evaluación de competencias informacionales autopercibidas en educación secundaria obligatoria. *Anales de Documentación, 22*(1). https://doi.org/10.6018/analesdoc.22.1.305641

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2014). McGraw-hill education. In *Metodología de la investigación* (6).

Hullman, J., Adar, E., & Shah, P. (2011). The impact of social information on visual judgments. In *Proceedings of the SIGCHI conference on human factors in computing systems*. https://doi.org/10.1145/1978942.1979157. Vancouver, BC, Canada.

Hullman, J., Qiao, X., Correll, M., Kale, A., & Kay, M. (2019). Pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics, 25*(1), 903–913. https://doi.org/10.1109/TVCG.2018.2864889

Huron, S., Jansen, Y., & Carpendale, S. (2014). Constructing visual representations: Investigating the use of tangible tokens. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 2102–2111. https://doi.org/10.1109/TVCG.2014.2346292

Javed, W., & Elmqvist, N. (2012). Exploring the design space of composite visualization. In *2012 IEEE Pacific Visualization Symposium*.

Kay, M., Kola, T., Hullman, J. R., & Munson, S. A. (2016). When (ish) is my bus? User-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proceedings of the 2016 chi conference on human factors in computing systems*.

Kim, Y.-S., Reinecke, L., & Hullman, J. (2018). Data through others' eyes: The impact of visualizing others' expectations on visualization interpretation. *IEEE Transactions on Visualization and Computer Graphics, 24*(1), 760–769. https://doi.org/10.1109/TVCG.2017.2745240

Kim, Y.-S., Walls, L. A., Krafft, P., & Hullman, J. (2019). A Bayesian cognition approach to improve data visualization. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3290605.3300912. Glasgow, Scotland UK.

Kong, N., & Agrawala, M. (2012). Graphical overlays: Using layered elements to aid chart reading. *IEEE Transactions on Visualization and Computer Graphics, 18*(12), 2631–2638. https://doi.org/10.1109/TVCG.2012.229

Kong, H.-K., Liu, Z., & Karahalios, K. (2017). Internal and external visual cue preferences for visualizations in presentations. *Computer Graphics Forum, 36*(3), 515–525. https://doi.org/10.1111/cgf.13207

Lee, S., Kim, S. H., & Kwon, B. C. (2017). VLAT: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics, 23*(1), 551–560. https://doi.org/10.1109/TVCG.2016.2598920

Lin, C., & Thornton, M. A. (2021). *Fooled by beautiful data: Visualization aesthetics bias trust in science, news, and social media*.

Mansoor, H., & Harrison, L. (2018). Data visualization literacy and visualization biases: Cases for merging parallel threads. In *Cognitive biases in visualizations* (pp. 87–96). Springer.

Mazza, R., & Berre, A. (2007). Focus group methodology for evaluating information visualization techniques and tools. In *2007 11th international conference information visualization (IV '07)*.

Munzner, T. (2014). *Visualization Analysis and Design*. AK Peters/CRC Press.

Pinto, M. (2009). Design of the IL-HUMASS survey on information literacy in higher education: A self-assessment approach. *Journal of Information Science, 36*(1), 86–103. https://doi.org/10.1177/0165551509351198

Rodríguez Gómez, G., & Gómez Ruiz, M. A. (2010). Análisis de contenido y textual de datos cualitativos. In *Principios, métodos y técnicas esenciales para la investigación educativa*.

Ryan, K. E., Gandha, T., Culbertson, M. J., & Carlson, C. (2014). Focus group evidence: Implications for design and analysis. *American Journal of Evaluation, 35*(3), 328–345.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2018). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics, 25*(1), 682–692. https://doi.org/10.1109/TVCG.2018.2864903

Sarrado, J. J., Cléries, X., Ferrer, M., & Kronfly, E. (2004). Evidencia científica en medicina: ¿única alternativa? *Gaceta Sanitaria, 18*(3), 13, 2004.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., et al. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. *Quality and Quantity, 52*, 1893–1907.

Schwabish, J. (2021). *Better data visualizations: A guide for scholars, researchers, and wonks*. Columbia University Press.

Szafir, D. A. (2018). The good, the bad, and the biased: Five ways visualizations can mislead (and how to fix them). *Interactions, 25*(4), 26–33. https://doi.org/10.1145/3231772

Tufte, E. R. (1985). The visual display of quantitative information. *Journal for Healthcare Quality, 7*(3), 15.

Tufte, E. R., Goeler, N. H., & Benson, R. (1990). In *Envisioning information* (Vol. 126). Graphics press Cheshire, CT.

Vázquez-Ingelmo, A., García-Holgado, A., García-Peñalvo, F. J., & Therón, R. (2021). A meta-modeling approach to take into account data domain characteristics and relationships in information visualizations. In *Advances in intelligent systems and computing series series, no. 1366 trends and innovations in information systems and technologies*. WorldCIST 2021, Azores, Portugal.

Vázquez-Ingelmo, A., García-Peñalvo, F. J., & Therón, R. (2019). Capturing high-level requirements of information dashboards' components through meta-modeling. In *TEEM'19 proceedings of the seventh international conference on technological ecosystems for enhancing multiculturality*. León, Spain.

Vázquez-Ingelmo, A., García-Peñalvo, F. J., & Therón, R. (2019). Information dashboards and tailoring capabilities - a systematic literature review. *IEEE Access, 7*, 109673–109688. https://doi.org/10.1109/ACCESS.2019.2933472

Vázquez-Ingelmo, A., García-Peñalvo, F. J., & Therón, R. (2019). Taking advantage of the software product line paradigm to generate customized user interfaces for decision-making processes: A case study on university employability. *PeerJ Computer Science, 5*. https://doi.org/10.7717/peerj-cs.203

Vázquez-Ingelmo, A., García-Peñalvo, F. J., & Therón, R. (2021). Ecore version of the meta-model for information dashboards (v3). *Zenodo*. https://doi.org/10.5281/zenodo.5788981

Vázquez-Ingelmo, A., García-Peñalvo, F. J., & Therón, R. (2022). MetaViz – a graphical meta-model instantiator for generating information dashboards and visualizations. *Journal of King Saud University - Computer and Information Sciences, 34*(10, Part B), 9977–9990. https://doi.org/10.1016/j.jksuci.2022.09.015

Vázquez-Ingelmo, A., García-Peñalvo, F. J., Therón, R., & Conde, M.Á. (2020). Representing data visualization goals and tasks through meta-modeling to tailor information dashboards. *Applied Sciences, 10*(7), 2306. https://www.mdpi.com/2076-3417/10/7/2306.

Vázquez-Ingelmo, A., García-Peñalvo, F. J., Therón, R., & Conde-González, M.Á. (2019). Learning analytics summer institute Spain 2019: Learning analytics in higher education. In *Extending a dashboard meta-model to account for users' characteristics and goals for enhancing personalization. Proceedings of LASI-Spain 2019*. Vigo, Spain: CEUR Workshop Proceedings Series. June 27-28, *2019* http://hdl.handle.net/10366/139803.

Vázquez-Ingelmo, A., García-Peñalvo, F. J., Therón, R., & García-Holgado, A. (2020). Specifying information dashboards' interactive features through meta-model instantiation. In *CEUR Workshop Proceedings Series, no. 2671 LASI-SPAIN 2020. Learning Analytics Summer Institute Spain 2020: Learning Analytics. Time for Adoption?*. Valladolid, Spain.

Verdugo-Castro, S., García-Holgado, A., Vázquez Ingelmo, A., García-Peñalvo, F. J., & Therón, R. (2024). Analysis of two focus groups about data visualization. https://doi.org/10.5281/zenodo.10556271.

Wanzer, D. L., Azzam, T., Jones, N. D., & Skousen, D. (2021). The role of titles in enhancing data visualization. *Evaluation and Program Planning, 84*(101896). https://doi.org/10.1016/j.evalprogplan.2020.101896

Ware, C. (2004). *Information visualization: Perception for design*. Elsevier. https://doi.org/10.1016/B978-0-12-812875-6.01001-X

Wexler, S., Shaffer, J., & Cotgreave, A. (2017). *The big book of dashboards: Visualizing your data using real-world business scenarios*. John Wiley & Sons.

Wilke, C. O. (2019). Fundamentals of data visualization: A primer on making informative and compelling figures. *O'Reilly Media*.

Wilkinson, S. (1998). Focus group methodology: A review. *International Journal of Social Research Methodology, 1*(3), 181–203. https://doi.org/10.1080/13645579.1998.10846874, 1998/01/01.