



Development of Visualization Metaphors for Monitoring Student Engagement in Distance Learning Environments

Pamella L. S. de Oliveira¹, Rodrigo L. Rodrigues², Alexandre M. A. Maciel¹

¹Departamento de Engenharia da Computação – Universidade de Pernambuco (UPE)
Rua Benfica, 455, Madalena — 50720-001— Recife – PE – Brazil

²Departamento de Educação – Universidade Rural de Pernambuco (UFRPE)
Rua Dom Manoel de Medeiros, s/n, Dois Irmãos – 52171-900 – Recife – PE – Brazil

{pamella.lesilvao, rodrigomuribec}@gmail.com.br, alexandre.maciel@upe.br

Abstract. *Given the low percentage of graduates in Distance Education courses, the need to visualize student data is perceived, as it makes it easier to monitor educational phenomena, such as engagement. The literature shows some gaps, such as the absence of analysis of visualization metaphors. This work aims to develop visualizations with educational data using the stages of understanding the problem, construction, and evaluation. The results show the choice of five visualization metaphors to present engagement variables. In addition, the evaluation results point to the absence of a tool for monitoring and the importance of viewing groups.*

1. Introduction

According to data from the 2021 Higher Education Census [Inep 2021], in Brazil, the number of graduates of courses followed by the Distance Education modality (EaD) is only 36.55%. To understand these values, it is necessary to analyze educational phenomenon. Engagement is the educational phenomenon considered one of the critical components in effective online teaching and essential for retention and improvement in the quality of the overall student experience [Haron et al. 2017].

To analyze engagement in EaD environments, digital “footprints” are collected, which students leave when interacting with the environment [Matcha et al. 2019]. Furthermore, according to Silva *et al.* (2018) [Silva et al. 2018], one of the most effective ways of providing educational information to students and instructors to assist in the understanding and optimization of the teaching-learning process is through interactive data visualization interfaces.

Despite the benefits of these visualizations, some studies in the literature point out the challenges of this type of visualization interface. In [Sedrakyan et al. 2019], the authors mention that the graphics are presented without providing support mechanisms that facilitate their interpretation. In addition, in the systematic mapping carried out by Dourado *et al.* (2018) [Dourado et al. 2018], the authors found few works that validated their proposals for visualization models with real users involved in the teaching-learning process.

Therefore, the objective of this research is to develop a visualization interface for educational data and validate it with specialists in the area of distance education. For this, the methodology of *Design Science Research (DSR)* is being used considering some

adaptations. In the stage of understanding the problem, a Systematic Literature Review (SLR) was carried out. Then, the metaphors of visualizations and the database were analyzed in constructing the artifact step. Finally, a structured interview was carried out with specialists in the evaluation step, as presented in section 3.

2. Theoretical Foundation

2.1. Visualization Metaphors

The literature presents an extensive collection of different visualization metaphors, which makes choosing "good visualizations" complex – (1) clearly illustrate a point; (2) are tailored to the appropriate audience; (3) are adapted to the presentation medium; (4) are memorable and increase understanding of the subject [S. Stoltzman 2018].

In Munzner (2014) [T. Munzner 2014], a structure composed of three questions is presented as a guide to help think about design choices: (1) What - does the user see?; (2) Why - does the user intend to use?; (3) How - will the coding be built?. Corroborating this, in Sedrakyan *et al.* (2019) [Sedrakyan et al. 2019], the authors also mention this structure and add one more question: Who - is it adapted for the target audience? In these questions, some specific points are considered. In the first: data, datasets, and attributes (quantitative, ordinal, categorical). In the second: actions (compare, identify, summarize) and goals (trends and discrepancies). For the third: encode, manipulate, facet, and reduce. Finally, in the fourth, memory and understanding.

2.2. Engagement in Distance Learning Environments

Engagement is the most common educational phenomenon in the [de Oliveira et al. 2022] literature. In addition, it is considered one of the main components to ensure consistency in learning activities, effective online teaching, and student permanence. It improves the quality of the overall experience and interaction, which is crucial to ensuring students' success in learning [Haron et al. 2017].

The literature presents several definitions and many types of engagement, however, there are three more common ones: (1) motivational engagement, which includes interest, affection, and values perceived by students while performing learning tasks; (2) behavioral engagement, which refers to observable student actions while performing a learning task; and (3) cognitive engagement, which is related to the implementation of learning strategies [Carrillo et al. 2017].

3. Materials and methods

In this research, the Design Science Research (DSR) methodology was adapted, which underlies and operationalizes the conduction of research and search based on understanding the problem, constructing the artifact, and evaluating the artifact [Lacerda et al. 2012].

3.1. Understanding the Problem

For the first phase of the DSR, understanding the problem, an RSL was performed. The protocol of this RSL had the following characteristics: four research sources - IEEE xplorer, Springer Link, Science Direct, and Scopus; the filter - 2016 to 2022; the search

string - "dashboard" OR "visualization" AND "engagement" AND "education"; exclusion criteria - works that are not written in English or Portuguese due to possible bias in the translation.

After these applications, thirty-two studies were returned, where the following gaps were identified: 1) the lack of integration between *machine learning* models with visualization interfaces; 2) the use of visualization metaphors little sophisticated; and 3) the lack of analysis of the best visualization metaphors for the nature of the data studied [de Oliveira et al. 2022].

3.2. Artifact Construction

The artifact was constructed in two steps: the definition of the visualization metaphors and the definition of the database used to construct the metaphors.

3.2.1. Definition of Visualization Metaphors

For the choice of visualization metaphors, twenty-seven articles were selected that addressed the grouping. With this, it was possible to analyze the five visualization metaphors most used in them: heat map, scatter, line, bar, and area. Next, Table 1 was elaborated, where it is possible to analyze the analysis structure of these metaphors according to [T. Munzner 2014] and [Sedrakyan et al. 2019].

Table 1. Analysis framework of the five visualization metaphors

Orde	Visualization Metaphors	Type of Attribute (What)	Purpose/Task (Why)
1rd	Heat map	1 quantitative 2 categorical	Find groups, outlier, summarize
2nd	Scatter	2 quantitative	View trends, outliers, distribution, correlation, find groups, compare, relate
3rd	Line	1 quantitative, 1 ordinal	Visualize trends, compare, relate
4th	Bar	1 quantitative, 1 categorical	Search, view trends, and compare values, composition
5th	Area	1 quantitative, 1 categorical, 1 ordinal	To compare

3.2.2. Database Used

The database used was extracted from Moodle referring to five courses at the Universidade de Pernambuco of 2014. This base has a total of 1771 students, however, initially, we are using only one course with 96 students to generate the graphs. In addition, the study [H. R. Macêdo et al. 2021] was considered a basis for choosing the variables corresponding to the student's engagement. They are 1) the Number of replies in forum posts;

2) the number of views to activities; 3) the Environment access number; 4) the Frequency of weekly access by the student; 5) Notes.

3.3. Artifact Assessment

The artifact assessment stage was carried out through a structured interview with four specialists in distance education. Six questions related to student engagement and monitoring were asked. In addition, the interface prototype with the visualization metaphors presenting the five engagement variables was also presented.

4. Analysis and Discussion of Results

4.1. Constructed Visualization Metaphors

As a demonstration, the figures below show the five engagement variables, their characteristics, and recommended visualization metaphors for each one, considering the information in Table 1. The data used refer to a subject taught from August to December from the database mentioned in section 3.

Figure 1 presents the choice of visualization metaphor for the variable "environment access number," which has the quantitative and objective attribute of comparing. Therefore, when analyzing the characteristics of this variable with the information in Table 1, it was noticed that the area chart visualization metaphor could be recommended for this variable, forming the structure: what does the user see? - a quantitative variable (Number of accesses) on the vertical axis and an ordinal variable (weekly periods) on the horizontal axis, and a categorical variable (groups); and why does the user intend to use it? - to compare the number of views of the environment by groups over time.

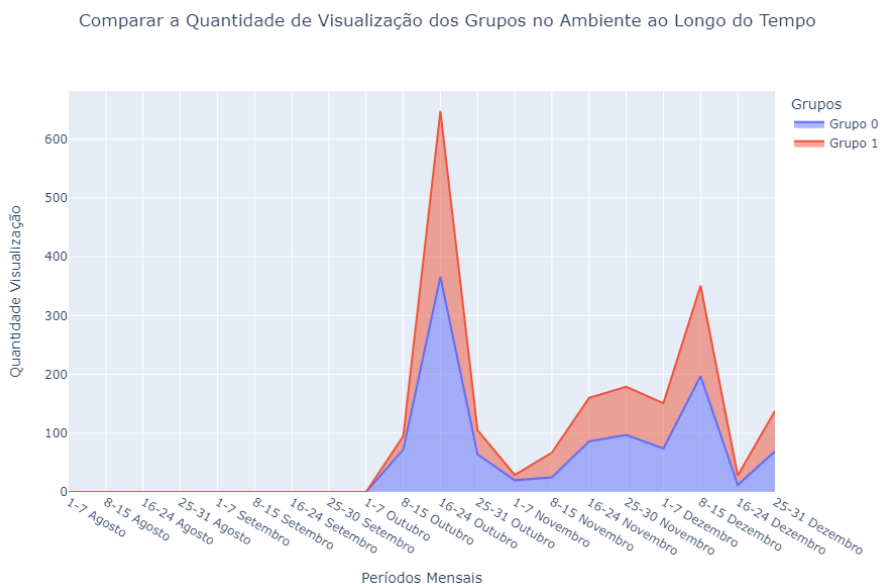


Figure 1. Compare the number of views to the environment by groups

Figure 2 presents the choice of visualization metaphor for the variable "the number of responses in forum posts," which has the quantitative and objective attribute of identifying the trend. Therefore, when analyzing the information of this variable with the

information in Table 1, it was noticed that the line chart visualization metaphor is recommended for this variable, forming the following structure: what does the user see? - a quantitative variable (Number of responses) on the vertical axis and an ordinal variable (weekly periods) on the horizontal axis, and why does the user intend to use it? - to track the trend of responses in forums over time.

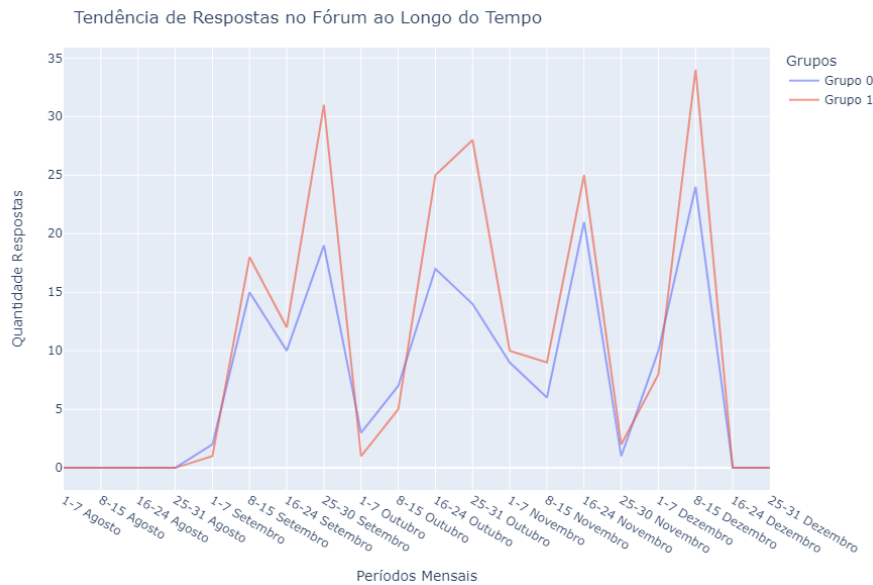


Figure 2. trend of responses in forums

Figure 3, presents the choice of visualization metaphor for the variable "Number of visualizations to activities", which has the quantitative and objective attribute of summarizing the accesses. Therefore, when analyzing the characteristics of this variable with the information in Table 1, it was found that the metaphor of visualization of the heat map is recommended for this variable, forming the following structure: what does the user see? - a quantitative variable (Number of accesses) represented by the color scale, a categorical variable (groups) on the horizontal axis, and an ordinal categorical variable (weekly periods) on the vertical axis; and why does the user intend to use it? - to summarize the amount of access to activities by groups over time.

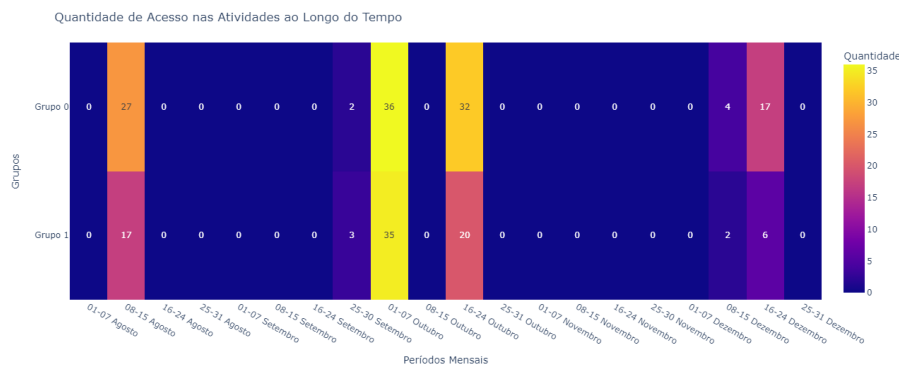


Figure 3. Summary of access to activities

Figure 4, presents the choice of visualization metaphor for the variable "Notes"

composed of three types of assessments: "Regular Call," "Final Assessment," and "Second Call," which have a quantitative attribute and objective of identifying the distribution of grades. Therefore, when analyzing these characteristics with the information in Table 1, it was verified that the visualization metaphor composed of a scatter matrix is adequate for this variable, forming the following structure: what does the user see? - a quantitative variable (Number of students) on the vertical axis and another quantitative variable (grades) on the horizontal axis; and why does the user intend to use it? - to follow the distribution of the number of students by grades in the assessments.

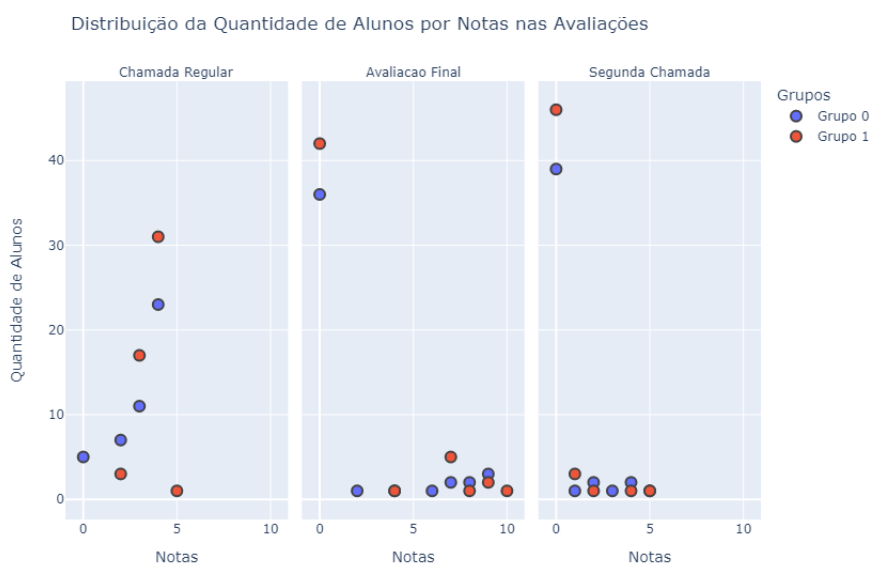


Figure 4. Distribution of the number of students by grades in the assessments

In Figure 5, the choice of visualization metaphor for the variable "Notes" is presented, referring to the evaluation of the checklist type. The choice of another visualization metaphor different from the one chosen for the other three types of evaluations is due to the values of this evaluation ranging from 0 to 100, fitting better into grade intervals, which makes this data categorical. Therefore, when analyzing these characteristics with the information in Table 1, it was verified that the visualization metaphor composed of a bar graph is adequate for the characteristics of this variable, which allows composing the following structure: what does the user see? - a quantitative variable (number of students) on the vertical axis and another categorical variable (range of grades) on the horizontal axis; and why does the user intend to use it? - to monitor the composition of students by grades in the checklist assessment.

In the figures presented, it is possible to notice the difference in the values obtained by the two groups in the forum, activities, and environment, which formed two groups with different engagement profiles. Where "Group 1" has higher values than "Group 0", it can be concluded that "Group 1" is composed of students with a more engaged profile, while "Group 0" comprises less engaged students. Regarding the metaphors of visualizations referring to the evaluations, in Figure 4, it is noticed that "Group 1" is present in the highest score values and is the one with the highest number of students with a score of 0 in the "Final Evaluation" and "Second Call" indicating that the majority did not need to do this type of evaluation and those who did manage to obtain a high

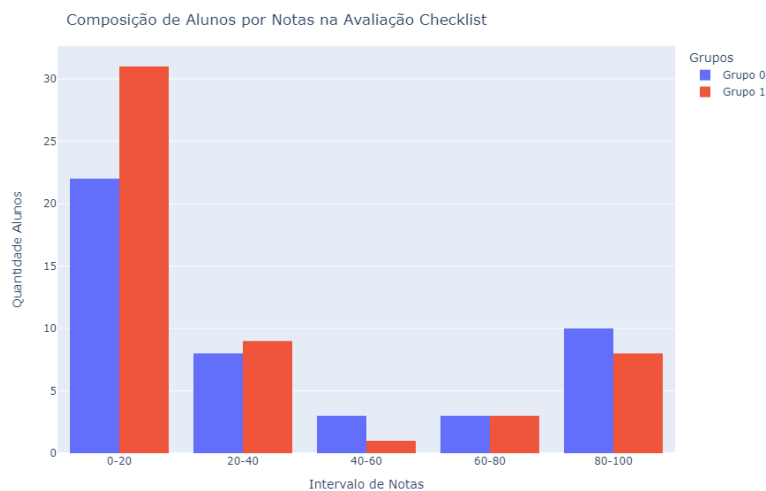


Figure 5. Composition of students by grades in the checklist assessment

score. However, in Figure 5, "Group 0" has a slightly greater difference than "Group 1" in the highest grade range "80-100", which would be a point to be analyzed to understand whether students had some difficulty in this type of evaluation.

4.2. Structured Interview with Experts

In this subsection, the responses obtained in the structured interview conducted with two male and female teachers, who answered six open questions, will be briefly explained:

(1) How do you identify students with the possibility of dropping out?

Three interviewees mentioned that it is due to the student's participation in the activities and the forum. In addition, they also mentioned that they take into account whether the answers are not a "copy and paste from the internet", in addition, one interviewee said that it was due to the feedback from the tutors. Interviewee 1- "Due to the student's way of interacting and participating, meeting deadlines for activities, copies, and pastes are made from the internet, that is, the way the student interacts with the environment and with the class". Interviewee 1 - "When they don't participate, mainly in the first forums. Furthermore, some criteria for participation in forum responses include copying and pasting from the internet and authorship". According to the responses, the importance of analyzing the interaction in the forums was clear, quantitatively and qualitatively.

(2) What tool/method do you use to track student engagement?.

Different means are used to follow up. One interviewee mentioned that she makes a follow-up worksheet; the others mentioned that she follows up through the environment, email, and tutor feedback. Interviewee 1 - "I make a more pedagogical worksheet with my tutor passing parameters such as participation, observation, and students' names". Interviewee 1- "Basically, I use three ways: the environment, looking to see if there are students online if they are doing activities; the email, where I receive communications from the environment; and the feedback from the tutors, where I receive information about face-to-face meetings and the specific activities that tutors pass". With these responses, it was possible to perceive the absence of a specific tool for monitoring students that can be

used both for teachers and tutors.

(3) How often do you usually accompany students?.

We had very varied answers to this question. One respondent mentioned "daily", another "weekly" and two "monthly". Interviewee 1 - "Daily I check if the students are online to study my subject" Interviewee 2 - "I follow up with the tutors every month, also because the activities take longer to carry out." It is noticed that the monitoring of students happens regardless of the frequency in which it is carried out.

(4)How can you identify students with different engagement profiles?.

In this question, three interviewees mentioned that it is due to the interaction in the forum, and only one mentioned that it is due to the tutors' feedback. Interviewee 1- "For the answer and the interaction with other students in the forum". Interviewee 2- "In the forum, I see the same student or groups that are generally the first to participate and that participate more in the first weeks but then disappear". These responses highlight the relationship between forums and student engagement through participation with other students and the frequency of access over time.

(5) How do you manage to follow up with students who are disengaged?.

Three interviewees said they post messages on the environment, and two mentioned that they contact the tutors. Interviewee 1- "I ask my virtual tutor to get in touch with the face-to-face tutor so she can see what is happening, with that the face-to-face tutor acts by going after the student and reports to the virtual one, then they contact me to try to give another chance for the student".

(6)What do you think about tracking groups of students instead of them individually?.

When questioned, the interviewees found it essential and mentioned factors such as the size of the classes and pedagogical proposals for the groups. Interviewee 2 - "Very important because it is not feasible to do individual follow-up due to the class size". Interviewee 1- "It is much better to look at groups to be able to think of proposals and pedagogical practices of those groups".

Furthermore, when the metaphors of visualizations above were presented, some observations were made to the interviewees: Interviewee 2 - "It would be important to put the number of students in each group"; Interviewee 2 - "Analyze the answers on the forums, as there may be short answers like "I agree", so it would be good to look at the quality of the answers as well".

Given these responses, it is clear that there is no tool for monitoring student data, especially regarding forum data, which is of great importance in identifying engagement. In addition, the importance of following through groups of students with similar profiles should also be highlighted, given the size of the classes.

5. Conclusion

When transformed into interactive visualizations, the data of students collected in EaD environments help the teaching-learning process. Given that, teachers can identify the profile of students and make decisions based on this information, which will imply the

intervention of possible failures and dropouts.

Given this importance, it is necessary to analyze the choice of visualization metaphors taking into account the type of data and the objective and adapt the proposal to the target audience's needs to develop "good visualizations" that make monitoring easy and intuitive.

This research used the analysis framework: What-Why-Who-How to develop metaphors of visualizations with educational data. The visualization metaphors were the five most used in studies that approach grouping: heat map, scatter, line, bar, and area.

In addition, with these visualization metaphors, it was possible to identify groups with different engagement profiles and monitor them, which is one of the essential factors for teachers. Given that, in the interview with the specialists, the absence of a specific tool that allows the monitoring of students was identified. In addition, the importance of this visualization through groups was also identified due to the size of the interviewees' classes.

However, it is possible to perceive the importance of a good visualization of educational data for teaching-learning. For future works, it is intended to make some adjustments to the visualizations mentioned by the interviewees, conduct interviews with distance education tutors and apply grouping and visualizations to other disciplines in the database.

6. Thanks

The authors would like to thank the professors interviewed; the Núcleo de Educação a Distância NEAD - UPE; and the FACEPE for granting the scholarship to the first author.

References

- Carrillo, R., Renaud, C., Prié, Y., and Lavoué, (2017). Dashboard for monitoring student engagement in mind mapping activities. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pages 433–437.
- de Oliveira, P., Rodrigues, R., and Maciel, A. (2022). Revisão sistemática da literatura sobre modelos de visualizações utilizados para analisar fenômenos educacionais em plataformas de lms. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 1006–1015, Porto Alegre, RS, Brasil. SBC.
- Dourado, R., Rodrigues, R., Ferreira, N., and Gomes, A. (2018). Mapeamento sistemático sobre o uso de visualização de dados para análise processual da aprendizagem em ambientes virtuais. *VII Congresso Brasileiro de Informática na Educação (CBIE 2018)*.
- H. R. Macêdo, P., B. Santos, W., and M. A. Maciel, A. (2021). Análise de perfis de engajamento de estudantes de ensino a distância. *Revista Novas Tecnologias na Educação*, 18(2):326–335.
- Haron, H., Natrah Aziz, N. H., and Harun, A. (2017). A conceptual model participatory engagement within e-learning community. *Procedia Computer Science*, 116:242–250. Discovery and innovation of computer science technology in artificial intelligence era: The 2nd International Conference on Computer Science and Computational Intelligence (ICCSICI 2017).

- Inep (2021). Senso da educação superior 2021. <https://download.inep.gov.br/educacaosuperior/censosuperior/documentos/2021/apresentacaocensodaeducacaosuperior2021.pdf>.
- Lacerda, D., Dresch, A., Proença, A., and Antunes Júnior, J. A. V. (2012). Design science research: A research method to production engineering. *Gestão Produção*, 20:741–761.
- Matcha, W., Ahmad Uzir, N., Gasevic, D., and Pardo, A. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, PP:1–1.
- S. Stoltzman (2018). What type of data visualization do you choose (if any)? Disponível em: <https://opendatascience.com/data-visualization-part-3/>. Acesso em: 01 de setembro 2022.
- Sedrakyan, G., Mannens, E., and Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, 50:19–38.
- Silva, N. A. d. L., Dourado, R. A., Rodrigues, R. L., and Cavalcanti, A. A. (2018). Investigating the suitability of multi-dimensional data visualization as an instrument for assisting distance learning instructor. *VII CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (CBIE 2018)*, pages 1–1.
- T. Munzner (2014). *Visualization Analysis and Design*. AK Peters, 1th edition.