Show Me the GIFference! Using data-GIFs as Educational Tools

Lorenzo Amabili
TU Wien
lorenzo.amabili@gmail.com

M. Eduard Gröller
TU Wien and VRVis
groeller@cg.tuwien.ac.at

Renata G. Raidou
TU Wien
rraidou@cg.tuwien.ac.at

ABSTRACT

We investigate the use of data-GIFs, i.e., graphics interchange format files containing short animations, to engage visualization viewers in learning about data visualization design pitfalls. A large number of data visualizations—among which, also several with bad data designs—are generated every day to convey information to lay audiences. To support non-expert viewers in recognizing common visualization design mistakes, we propose *data-GIFs*. Data-GIFs are short educational animations played in automatic repetition with a single core message on how the design flaws of a given visualization can be identified. After defining what bad data visualization design entails, we inform the design requirements for the data-GIFs. We, subsequently, design four variants: two data-GIFs, which use respectively interchangeability and smooth transitions, a static variant with juxtaposition, and a data-video approach with audio. In a controlled user study with 48 participants, we compare the four variants. We demonstrate that interchangeability and smooth transitions effectively support viewers in assessing *why* elements characterizing bad data visualizations are indeed bad. Yet, smooth transitions are more engaging, and data-videos are more efficient for the identification of differences between bad and good data visualization designs.

Keywords

Data-GIF, visualization education, visualization design pitfalls, design study, information visualization.

1 INTRODUCTION

Among the many data visualizations produced every day, we often encounter cases with unsuitable encodings and visuals [20, 21]. These visualizations may confuse non-expert viewers and make data interpretation difficult, or may even (un)intentionally communicate misleading information. Understanding what makes data visualizations bad, i.e., confusing or less effective in communicating data, and raising awareness about the existence of bad visualization designs is a core research topic of our field [8, 17, 30]. In this work, we investigate *how to engage visualization viewers in effectively identifying design components that distinguish bad (i.e., poorly designed) from good (i.e., well-designed) visualizations*.

We focus on designing and evaluating an engaging approach that communicates to large audiences how to effectively identify pitfalls when interpreting visualizations. We start by setting the formal definition of a *bad data visualization*, in conjunction with established taxonomies [20, 21]. After defining what bad visualization design entails, we determine a set of *learning goals* to drive further our approach design.

We identify data-GIFs as a suitable medium to show comparatively the differences between two visualization designs of interest (bad vs. good). The concept behind this solution is to expose viewers to bad and good visualizations of the same data and to make them aware of *design differences*. *Data-GIFs* are data-driven graphics interchange format (GIF) files containing short animations played in automatic repetition. Being concise in size and duration, they are versatile in conveying a single core message about the pitfalls of the visualization design, and in facilitating the comparison of different designs of the same data. Moreover, the use of data-GIFs is anticipated to engage viewers in learning visualization design concepts [1].

The *contribution* of this work is the development and assessment of data-GIFs as an effective and engaging approach for communicating pitfalls in data visualization designs to lay audiences, and for conveying how to recognize and interpret such pitfalls.

2 RELATED WORK

Recent works shed light on common design errors affecting visualizations by analyzing several misleading real-world cases and by developing a taxonomy to categorize them. Lo et al. [20] categorize 74 types of issues and form a taxonomy of misleading elements in visualizations to propose guidelines for the community. They identify five main categories of issues; namely *input*, *visualization design*, *plotting*, *perception*, and *interpretation*. McNutt et al. [21] propose a conceptual model to show how visualization mirages appear at every stage of the visual analytics process, distinguishing between *data-driven mirages* (or *data representation* issues) and *design-driven mirages* (or *data presentation* issues). Also, there is a large number of early contributions in the visualization community, providing tools for improving visualization literacy concerning the employed design and visual encodings [9, 30]. In this work, we focus on how all this knowledge can be harvested to educate lay audiences.

2.1 Tools for Visualization Literacy

Prior work has stressed the importance of the concept of *deconstruction and construction* in visualization education, and how deconstructing and constructing data visualizations can support their interpretation and design. Bishop et al. [4] developed *Construct-A-Vis*, a tablet-based tool that can guide visualization activities with children based on the learning paradigm of *constructionism*, scaffolding mechanisms, and shared interactions. This work shows the potential of a free-form constructive approach, which can lead to engaging children with data and their related mapping processes. Börner et al. [6] proposed visualization exercises based on the construction–deconstruction concept to teach visualization. The authors describe how to assess learners' insights by defining a visualization literacy framework.

Current approaches are built upon (more-or-less) complex models and implementations with particular target groups in mind, e.g., children [9]. Given their level of complexity or specific audiences, these solutions are not suitable for mass consumption, such as within a social media setup, or for more general audiences. To our knowledge, there is no previous investigation that targets lay audiences through an easy-to-implement—yet, effective and engaging—strategy for improving visualization literacy.

2.2 Data-GIFs to Convey Information

Graphics interchange format (GIF) files were released in 1987 to convey automatic, looped animations of individual, short messages without sound. Despite the increasing interest in employing GIFs to communicate information [3], there is not much research on the direction of data-driven GIFs yet. Shu et al. [28] introduced a review of data-GIFs used in the wild and investigated what makes a data-GIF understandable, by conducting a qualitative analysis. The work demonstrates the impact of the design factors of a data-GIF on how the users understand the core message presented by it. The work further proposes guidelines for designing understandable data-GIFs, without focusing on their user engagement level or the educational power. Other sources of inspiration include the large collection of data-GIFs by Jeremy Singer-Vine, the work of Lena Groeger, and Dark Horse Analytics—all showcasing different data-GIFs or short video formats categorized based on their main visualization goal.

3 CONCEPTUALIZING DATA-GIFS

Our goal is to educate *lay audiences* in recognizing a *poor visualization design* and to inform them about *a good visualization design alternative*—effectively and engagingly. The choice of focusing on lay audiences is based on the assumption that this group of viewers is anticipated to suffer the most from poorly designed visualizations. Based on the definition of our goal, we make two conceptual choices.

The first conceptual choice links to the **design space of a visualization**. The available design space is multi-dimensional and very complex for laypeople without prior knowledge in data visualization [20]. We, therefore, reduce this complex space to two "simple" categories: *bad* and *good* visualizations. We define *bad data visualization* as a representation that fails to effectively convey information, misleads the viewer, or obscures the underlying meaning of the data due to design flaws, inaccuracies, or poor choices in visual encoding. Oppositely, a *good data visualization* is a representation that effectively conveys information, enhances understanding, and facilitates insights by using appropriate visual design principles and techniques. An example of a bad data visualization is depicted in Fig. 1, together with its transformations into a good one.

The second conceptual choice is associated with the use of the **construction–deconstruction concept** [6, 14]. Deconstructing a visualization is simpler than starting from scratch. By comparing bad and redesigned good visualizations, learners can easily grasp the differences. This learning-by-contrast approach engages viewers and facilitates understanding.

4 DESIGNING DATA-GIFS

In this section, we identify the learning goals and requirements that drive the design of our approach.

4.1 Learning Goals

Bloom et al.'s [5] taxonomy of learning objectives classifies educational learning objectives into six complexity levels: *knowledge, comprehension, application, analysis, synthesis,* and *evaluation*. Considering our target audience, this work targets only the first three levels of the taxonomy. Based on Bloom et al.'s taxonomy and the learning-by-contrast approach, we define our learning goals:

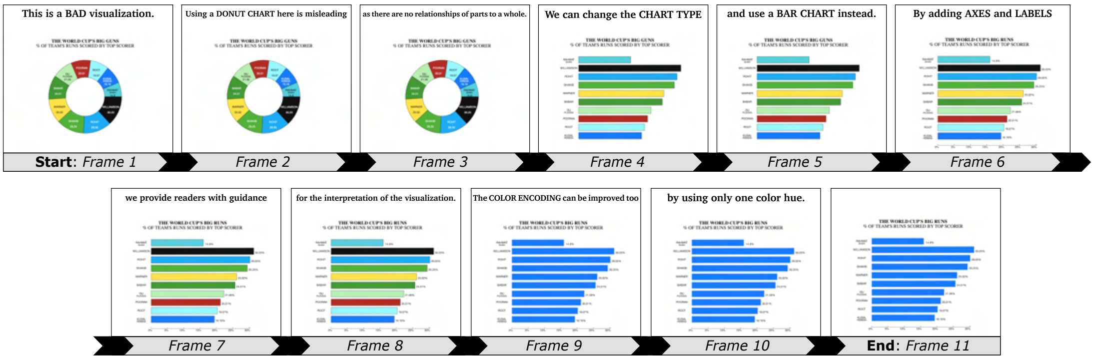

Figure 1: A data-GIF fractionated into eleven frames with a total duration of 38 s, where transitions (smooth or non-smooth) occur at frames 3→4, 5→6, and 8→9. Text is employed as guidance in each frame (top).

L1 ***Identify the attribute(s) of a bad data visualization, as opposed to those of a good visualization.*** For instance, learners should recognize that 3D visuals are unnecessary (bad) if the data can be sufficiently displayed in 2D (good).

L2 ***Recognize why the attribute(s) of a bad data visualization is (are) actually bad.*** For instance, learners should understand that a 3D chart is redundant if the data have only two dimensions.

L3 ***Comprehend how the attribute(s) of a bad data visualization could be improved.*** For instance, learners should recognize that an unnecessary 3D chart can be improved by simplifying it to 2D.

4.2 Design Requirements

The main requirements for our approach design are:

R1 ***Support the learning goals.*** The design of our approach should support the specific learning goals set in Sec. 4.1 to ensure that the target audience can identify *what* a bad visualization is (**L1**), *why* it is characterized as bad (**L2**), and *how* it can be transformed into a good one (**L3**).

R2 ***Be simple, understandable, and self-contained.*** Our approach must have a clear closed linear structure with a beginning (bad visualization) and an end (good visualization).

R3 ***Avoid information overload.*** The amount of information conveyed to the audience must not be overwhelming. Therefore, we should convey one message at a time.

R4 ***Maintain object constancy.*** During visual comparison, our approach must maintain object constancy, i.e., every graphical object can be visually tracked. So, secondary details should be de-emphasized to keep viewers oriented [27].

4.3 Exploring the Data-GIF Design Space

Based on what we defined in Sec. 4.1 and 4.2, we envision data-GIFs as short animations that incrementally progress from depicting a bad visualization state to a good one. For example, if a visualization has two flaws, then it comprises three *states*: the initial bad state, the state after fixing the first flaw, and the state after fixing the second flaw, which is also the final good state. Each state consists of *frames*. The way each frame transitions to the next one and how its content is communicated to the audience affects how viewers perceive, consume, and compare information through different states.

Therefore, our data-GIFs heavily rely on *comparison strategies* and a *messaging medium*, as means of communication (**R1,4**) with the audience. The former affects how the data-GIFs present their content, i.e., how they convey the comparison of a bad vs. a good data visualization state to the viewers. The latter affects which additional mechanisms the data-GIFs employ to support the viewers in understanding the conveyed message. In addition to these two aspects, we also consider in our design the guidelines proposed previously by Shu et al. [28] to improve the understandability of data-GIFs. Their work gives insights into designs that support **R2–4**, such as recommendations for the use of animation, the use of pauses to denote the end, and the structure of the visual content (i.e., text and animation).

4.3.1 Content Presentation by Comparison

We consider three different cases for the presentation of the content, inspired by previous work from the domains of comparative visualization [15] and narrative visualization [12, 16, 27]:

Smooth transitions—By interpolating the differences between frames within a specified time interval, one can smoothly transition from one frame to the next. Visualizations integrated with smooth transitions are generally preferred, being more engaging and effective in facilitating understanding [12]—given that they follow the *congruence* and *apprehension* principles [31]. We consider smooth transitions as potentially the most engaging and effective means for directing the viewer's attention to the most relevant information in the visualization, i.e., the differences between visualization states.

Interchangeability—Transitions by interchangeability consist of switching from one frame to another one without interpolation. Previous work found that interchangeable transitions are as effective as smooth transitions in some cases [16] and this solution may be preferred if the data-GIF frames are not largely different from each other [28]. Otherwise, object constancy may be affected.

Juxtaposition—In juxtaposition, all states are put next to each other [15] and there are no transitions. We consider this a special case that requires only one frame per state (i.e., it is a static visualization). Previous studies [25, 31] claim static visualizations to be equally or more effective than animation in some cases. Hence, the animation is not indispensable to effectively compare two or more visualization states—yet, the number of states should be limited.

4.3.2 Messaging Medium

During the viewing process, we need messaging mechanisms to help the viewer understand and remain attentive to the conveyed message. GIFs often include text [28] and—technically—do not support audio, which is a strong messaging medium. Other visual cues and guidance mechanisms can also be supported [18, 19, 27], but we consider them out of scope for this initial investigation. We consider only:

Text—Integrating text into our proposed data-GIFs augments their communicative value, as a viewer can easily extract information from textual explanations [22]. Previous work found that effectively linking captions, headlines, introductory text, summaries, and text annotations improves user engagement in interpreting visualizations [32]. The *intra-frame text* aims to explain how to read the visualization within the data-GIF and provides additional information about it. The *inter-frame text* plays the role of a narrator (often across states). Still, text should be staged, i.e., designed with attention to wording, number of words, style, and arrangement, to avoid information overload.

Audio—GIFs do not technically support audio. Previous studies observed that the influence of audio narration is higher than the influence of visual cues on learning outcomes [19]. Audio narrations can substitute the inter-frame text, allowing viewers to focus faster on the intra-frame text (e.g., legends or labels of the visualization). We expect that vocal guidance may reduce the number of times that a data-GIF needs to be observed by using another sensory channel to reduce visual overload. We intend to investigate whether the inherent lack of audio in GIFs is a limiting factor.

5 IMPLEMENTING DATA-GIFS

5.1 Attributes of Bad Data Visualizations

To showcase (and later assess) the design of the data-GIFs, we first need a good overview of the potential attributes characterizing bad visualizations. We start our search from existing taxonomies [20, 21] that comprehensibly summarize severe or prominent types of issues encountered in visualization designs. These taxonomies guided us in building a corpus of bad data visualizations from real-life examples. These examples come from various sources in the wild, i.e., blogs, social media, or already-existing collections of bad data visualizations such as VisLies. During a meta-selection process, we excluded scenarios, where many different issues could be encountered and kept only samples with a few (i.e., up to three) simple pitfalls. This inclusion strategy aimed to exclude complex scenarios requiring an excessively long viewing time and, therefore, not compatible with the purpose of our approach (and the typical short duration of GIFs). Hence, the attributes included in our corpus is not an exhaustive list—rather, an *indicative registry* for a first investigation. The final corpus includes 92 examples and is available in our online repository. Samples are shown in Fig. 1 and 2.

Upon collection, the first author analyzed our corpus of bad data visualizations to identify and list their pitfall(s). The other authors went through this analysis and verified its appropriateness. The arising disagreements were discussed among the authors and solved collaboratively. Successively, the first author conducted a qualitative analysis (also available in our repository) to code the encountered bad data visualization attributes into meaningful categories. Several visualizations were affected by more than one attribute and, therefore, were assigned to multiple categories. This was done in an iterative process consisting of identifying and repeatedly refining the coding until the list was saturated. The same process as in the pitfall identification verification was followed for the verification of the coding results by all authors. The coding resulted in the identification of eight indicative attributes, which are shown in Fig. 2 (together with their corresponding improvements). We hereby report them, ordered by frequency of occurrence in our corpus of bad data visualizations. Some cases were assigned to more than one category, therefore the denoted percentages do not sum up to 100%.

A1 ***Excessive display of visual elements*** *(34.8%; 32/92)*, which leads to clutter and information overload. This is a presentation issue related to *plotting* (chaotic canvas) in Lo et al.'s taxonomy [20]. An example is shown in Fig. 2 (**A1**).

A2 ***Misleading visualization axes*** *(28.3%; 26/92)*, which is a data representation issue and relates to the *visualization design* (choice of axis) stage in Lo et al.'s taxonomy. Fig. 2 (**A2**) shows an example with misleading, truncated axes.

A3 ***Inappropriate choice of visualization typology*** *(22.8%; 21/92)*, which relates to choosing the appropriate representation for a given data set.

Figure 2: The eight scenarios used in the study, each covering a bad data visualization attribute (**A**): (**A1**) excessive display of visual elements; (**A2**) misleading visualization axes; (**A3**) inappropriate choice of the visualization typology; (**A4**) unnecessary use of 3D visuals; (**A5**) lack of guidance; (**A6**) miscalculated geometric areas; (**A7**) difficult-to-read text; (**A8**) wrong use of color encoding. For each scenario, we also denote its good state.

This is a data representation issue that occurs in the *visualization design* (choice of chart) stage in Lo et al.'s taxonomy. Fig. 2 **(A3)** exemplifies an inappropriate choice of visualization, where a part-to-whole visualization is used for depicting percentages that add up to more than 100%.

A4 ***Unnecessary use of 3D visuals*** *(20.7%; 19/92)*, which is also a common type of chartjunk. This is a data presentation issue that relates to the *perception* stage in Lo et al.'s taxonomy. Fig. 2 (**A4**) shows such an example.

A5 ***Lack of guidance*** *(19.6%; 18/92)*, where annotations, legends, captions, chart titles, axis labels, etc. are omitted. This is a data presentation issue and occurs in the *plotting* (incomplete chart) stage in Lo et al.'s taxonomy. An example is shown in Fig. 2 (**A5**), where annotations reveal additional information.

A6 ***Miscalculated geometric areas*** *(15.2%; 14/92)*, where data are not represented faithfully. This is a data representation issue that also happens in the *perception* stage in Lo et al.'s taxonomy. An example is shown in Fig. 2 (**A6**).

A7 ***Difficult-to-read text*** *(14.1%; 13/92)*, which is a data presentation issue and relates to the *plotting* (chaotic canvas) stage in Lo et al.'s taxonomy. Fig. 2 (**A7**) gives an example of a visualization with difficult-to-read text.

A8 ***Wrong use of color encoding*** *(13%; 12/92)*, which is a data representation issue in the *visualization design* (color mess) stage in Lo et al.'s taxonomy. Fig. 2 (**A8**) shows an example of how color can be misused and distract the audience.

5.2 Visual Narratives with Data-GIFs

We implement our data-GIFs in a linear structure that depicts an initially bad data visualization, which we then correct and transform into a good visualization. We include a visualization *state* for each design correction (i.e., the stages between the bad and the good visualization) to gradually reveal improvements. All bad visualization cases that we implemented are presented in Fig. 2 together with their final transformations into good visualizations. In this figure, we show only the bad vs. good state, but Fig. 1 shows an example with all intermediate steps for **A7**.

Fig. 1 showcases the storyboard of a data-GIF with four states shown in frames 1, 4, 6, and 9. The data-GIF has been fractionated into eleven frames (1–11). The first frame announces that "*This is a bad visualization*" and we present the depicted donut chart as a bad design choice. Then, we show *what* is wrong with the depicted visualization and explain *why*: the chart choice is not appropriate (frame 2) because there is no parts-to-a-whole relationship (frame 3). We also indicate *how* to improve the visualization by exchanging the chart type for a bar chart (frames 3→4). Similarly, we incrementally introduce and correct two more flaws: we add axes and labels for readability (frames 5→6), and we change the color encoding (frames 8→9). At the end (frame 11), we show the good state that concludes the GIF.

The frames can be set one after the other with a smooth or non-smooth *transition*. To determine the optimal transition duration for the smooth transitions in data-GIFs, we follow the suggestions by Heer and Robertson [12]. We achieve object constancy between each frame by only changing one design pitfall at a time and by keeping constant all other visual elements. If this is not possible, we opt for a keyframing animation involving tweening [29]. The duration of the animation is set based on the study by Heer and Robertson [12]. For example, the data-GIF depicted in Fig. 1 requires 19 *s*. The last frame with the good visualization includes also a pause to denote the end [28]. Alternatively, the bad vs. good data visualization states can be put side-by-side [27], similarly to the configuration in Fig. 1.

For the *messaging medium*, we use inter-frame text as in Fig. 1 (at the top of each frame), or audio. In the first case, the number of words included in the inter-frame text of data-GIFs takes into account the results of a previous study, claiming that people can read 175-300 words per minute [7]. Considering that the text may include terms unknown to the audience, either due to specific visualization and data-related terminology, or due to non-native command of English, we use the lower boundary, i.e., 175 words per minute. We, then, double the time to ensure that viewers can process the message in the data-GIF [26]. For audio, we use the same guidelines and resulting text as for the textual messaging medium case. We, then, employ an automatically generated female voice-over at a normal speed (150 words per minute) and pitch (200 Hz).

6 USER STUDY

We conducted a user study to assess four *variants* of our approach, resulting from the design space of Sec. 4.3:

V1 ***Data-GIFs with smooth transitions***, featuring smooth transitions and text as messaging medium.

V2 ***Data-GIFs with interchangeability***, featuring interchangeability and text as messaging medium.

V3 ***Static visualizations***, featuring juxtaposition and text as messaging medium.

V4 ***Data-videos***, featuring smooth transitions and audio as messaging medium.

The main goal of our study is to assess whether data-GIFs can communicate to a general audience how to distinguish bad visualizations from good ones and how to recognize common mistakes in visualization designs—*effectively*, *engagingly*, and *efficiently*. We are primarily interested in understanding which of the four variants is a more engaging and more effective tool. However, efficiency is also an interesting factor in learning as it correlates negatively with cognitive load and negative emotions, such as frustration [24]. We formulate three hypotheses for our user study:

H1 *Data-GIFs* **(V1,2)** *are* ***more effective*** *for the identification of differences between bad and good data visualization designs than static visualizations* **(V3)** *and data-videos* **(V4)**.

H2 *Data-GIFs* **(V1,2)** *support a* ***more efficient*** *identification of differences between bad and good data visualization designs than static visualizations* **(V3)** *and data-videos* **(V4)**.

H3 *Data-GIFs* **(V1,2)** *are* ***more engaging*** *for the identification of differences between bad and good data visualization designs than static visualizations* **(V3)** *and data-videos* **(V4)**.

6.1 Participants

We recruited 48 participants between 24 and 37 years old, located in 13 different countries, with different educational backgrounds and a good command of English. The recruitment was done by snowball sampling. Among the 48 participants, 16 are laypeople, 16 are professionals working with data visualizations (e.g., data scientists or journalists), and the remaining 16 are visualization experts (e.g., researchers). For conciseness, we will hereby refer to these three groups as *low*, *medium*, and *high visualization literacy* groups respectively—although the categorization reflects rather their prior experience in working with data.

6.2 Study Design

We designed eight scenarios for the user study—one for each bad data visualization attribute (**A1–8**) defined in Sec. 5.1. These are also depicted in Fig. 2. We implemented all variants (**V1–4**) for all scenarios in `d3.js` with the strategy described in Sec. 5.2. All cases are English-based, and their mean duration is 26 *s* (*SD* = 9.86 *s*), with the exception of variant **V3**, which is static. All cases are included in our repository.

We separated the 48 participants into four groups (one for each variant). Each group consists of 12 participants and is balanced with four participants from each visualization literacy group. During the study, we exposed each of the four groups to one variant using the eight designed scenarios, in randomized order. We employ a 12×8×4 mixed design with two between-subjects independent variables (i.e., variant and visualization literacy level) and one within-subjects independent variable (i.e., attribute scenarios **A1–8**). In total, the study included 384 runs (i.e., 48 participants × 8 attributes). To deal with possible effects of confounding factors, we randomized our sample by arbitrarily assigning each participant to a group, as well as the order of attribute scenarios **A1–8**.

6.3 Tasks

During the study, each participant consumed the eight attribute scenarios (**A1–8**) through one of the four variants (**V1–4**). Subsequently, they were asked to view and understand the content of each scenario, before conducting **two tasks**. First, they watched the eight scenarios and, for each scenario, performed an **XYZ test** [11] following the think-aloud method. This task targeted **H1**–**H2**. Second, they completed an **engagement questionnaire** inspired by previous works [2, 13]. This questionnaire was completed at the end of the entire session to address **H3**.

6.4 Study Setting

We conducted all studies individually and online through video calls. All meetings were recorded (screen sharing and audio) and each meeting lasted around 25–45 minutes. First, we shared the materials with the participants and instructed them verbally on what to do. For each scenario, the participants were allowed to view the given variant for as long as needed to understand its story. Yet, they were informed that the consumed time was being measured and they had to verbally communicate when "processing" was complete. Then, we started posing questions to the participants, as part of the first task (XYZ test, Sec. 6.5). The second task was provided as a link to an online engagement questionnaire (Sec. 6.6). After the study, we used the recordings to analyze the answers.

6.5 The XYZ Test

We employed the XYZ test, introduced in psychology by Haim Ginott [11], to assess whether our learning goals are met. The XYZ test consists of three open questions, which directly map back to our learning goals, expressed in a less technical language:

X ***What*** *are the differences between the initial and the final data visualization design?* (links to **L1**)

Y ***Why*** *was the initial data visualization design changed into the final one?* (links to **L2**)

Z ***How*** *would you change the initial data visualization design to obtain the final one?* (links to **L3**)

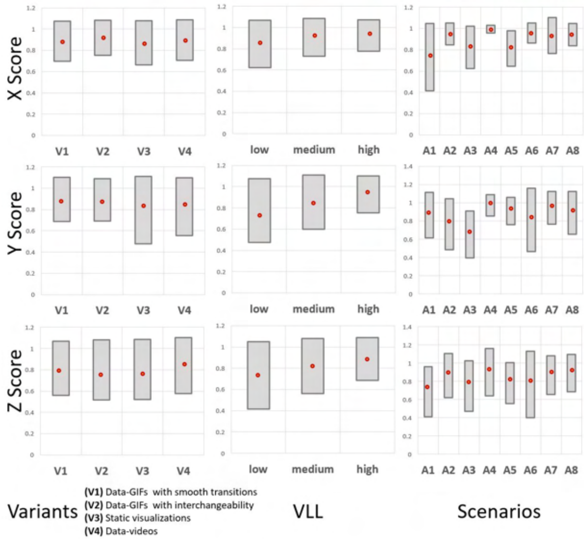

Figure 3: Crossbar plots of the X score (first row), Y score (second row), and Z score (third row) w.r.t. variant (first column), visualization literacy level (second column, VLL), and scenario (third column). The red dots indicate the mean values for each measurement.

To assess the accuracy of the responses to the XYZ questions, the first author used a weighted scoring method to assign a score (i.e., 0 = wrong, 0.25 = mostly wrong, 0.5 = semi-correct, 0.75 = mostly correct, 1 = correct) and reference answers, generated before the study. The XYZ test responses help assess **H1**, while consumption time was measured before the XYZ test to assess **H2**. A visual summary of these results is shown in Fig. 3. Our repository includes the details of the statistical analysis and we hereby summarize only the most significant outcomes.

Statistical Analysis for H1—We performed a statistical analysis on the scores related to the XYZ test, i.e., the score related to the X, Y, and Z components to investigate the effectiveness of the tested variants. A visual summary of these results is shown in Fig. 3. Given that the scores are based on the weighted scoring method mentioned above, they are categorical and ordered. For this reason, we performed multiple ordinal logistic regressions and Pearson's χ-squared tests to analyze the relationship between the XYZ scores and the factors of our study, namely the variants, the scenarios, and the visualization literacy level. We found that *participants watching data-GIFs with smooth transitions* (**V1**) are 2.16 and 2 times more likely (p=0.011 and p=0.045) to obtain *a higher Y score* (i.e., to better recognize *why* the attributes of bad data visualization are actually bad) than those watching static visualizations (**V3**) or data-videos (**V4**), respectively. This observation is shown in Fig. 3 (see first column, second row).

Additionally, we found that differences in *the XYZ scores also depend on the visualization literacy level*

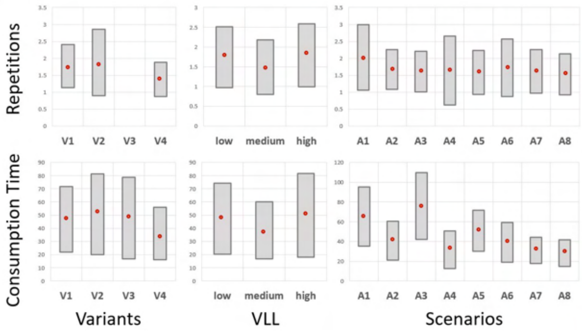

Figure 4: Crossbar plots of the watching repetitions (i.e., the ratio of consumption time over data-GIF duration, in the first row) and consumption time (second row), w.r.t variant (first column), visualization literacy level (second column, VLL), and scenario (third column). Repetitions of **V3** (static visualization) cannot be calculated. The red dots indicate the mean values for each measurement.

of the participants, as shown in the second column of Fig. 3. Our analysis results indicate participants with a high visualization literacy level are 2.19 times more likely (*p*=0.012) to obtain a higher Y score than participants with a medium visualization literacy level. While an average visualization literacy level is enough to correctly identify the attributes of bad data visualizations as opposed to those of good visualizations, recognizing why the attributes of bad data visualizations are actually bad requires a high visualization literacy level. As expected, *the higher the visualization literacy level of the participants the better they reflect on the acquired knowledge on how the bad data visualization can be improved* (Z score, *p*=0.026). Our findings also support that the *scenarios influence the score* of the participants, as depicted in the third column of Fig. 3. There are no other underlying patterns with regard to attributes or insights in this observation.

Statistical Analysis for H2—We also statistically analyzed the consumption time and the ratio of consumption time over data-GIF duration (i.e., how many times a data-GIF is watched) to assess efficiency. For these time-related continuous variables, we used ANOVA and for all the analyses, we checked all test assumptions. A summary of these results is shown in Fig. 4.

The number of repetitions needed by the participants to understand the content of the given variant was on average 1.865 for the data-GIFs with smooth transitions (**V1**), 1.739 for the data-GIFs with interchangeability (**V2**), and 1.375 for the data-videos (**V4**). For the static variant (**V3**), time cannot be measured without additional methods (e.g., eye-tracking). In essence, *data-videos need to be watched fewer times than the other two variants* (*p*=0.0009 and *p*=0.0012, respectively). This is also shown in Fig. 4 (see first row, first column). On the other side, *the consumption time is influenced by the variant, the participants' visualization literacy level, and the scenario* (*p*=0.004) (see the second row of Fig. 4). According to our results, scenarios **A1** and **A3** (and in some cases also **A5**) require more time than other scenarios with most of the variants (see second row, last column of Fig. 4). Participants with low and medium levels of visualization literacy needed more time for other attributes, such as **A2** (see second row, second column of Fig. 4). This was expected, as scenarios **A1**, **A3**, and **A5** contain more states between bad and good visualization, i.e., they might be more complex, or not so easy to process and remember.

6.6 The User Engagement Questionnaire

Our questionnaire for user engagement (**H3**) is inspired by previous work [2, 13] and consists of 28 questions on a seven-point Likert scale. With this questionnaire, we assess user engagement including cognitive involvement, affective involvement, enjoyment, presence, experience, and aesthetics. Our supplementary material includes the questionnaire and the statistical analysis.

Statistical Analysis for H3—We performed multiple ordinal logistic regressions to learn more about the differences between the variants. First of all, we observed no differences among the study participants concerning the perceived *aesthetics* of the variants. However, we found that the *cognitive involvement* of the participants can change based on the variant watched. According to our results, the participants watching data-GIFs with smooth transitions (**V1**) are 6.57 times more likely to *reflect on the content* during viewing than participants watching the static visualizations (**V3**) (*p*=0.024). Also, *learning something* new by watching the data-GIFs with smooth transitions (**V1**) is 14.56 times more likely than by static visualizations (**V3**) (*p*=0.001).

The visualization literacy level of participants plays also a role in their perceived *cognitive involvement*. Participants with a low visualization literacy level are 4.69 times more likely to reflect on the content and 7.32 times more likely to remember some parts of the data-GIFs than participants with a medium visualization literacy level (*p*=0.038 and *p*=0.007, respectively). Similarly, participants with a high visualization literacy level are 5.77 times more likely to remember some parts of the GIFs and 4.01 times more likely to learn something new than participants with a medium visualization literacy level (*p*=0.020 and *p*=0.047, respectively).

Finally, for data-GIFs with smooth transitions (**V1**) participants are 8.06 times more likely of being involved than for data-GIFs with interchangeability (**V2**) (*p*=0.007). They are also 5.24 times more likely to concentrate on the GIFs while watching data-GIFs as compared to static visualizations (**V3**) (*p*=0.040). Regarding the *experience*, data-GIFs with smooth transitions (**V1**) are 4.99 times more likely to be watched than static visualizations (**V3**) (p=0.036).

6.7 Summary of Findings and Discussion

Regarding our Hypotheses—The effectiveness of data-GIFs (**V1,2**) surpasses static visualizations (**V3**) in discerning bad from good data visualization designs—**affirming H1**. Data-GIFs (**V1,2**) help viewers understand why certain elements are flawed and this observation is particularly influenced by the viewers' visualization literacy and depicted scenarios. Conversely, the efficiency of data-GIFs (**H2**) is challenged, leading to **rejecting H2**. Data-videos (**V4**) consume less time and require fewer repetitions due to their audio aids, while other variants need multiple viewings. Finally, data-GIFs with smooth transitions (**V1**) and data-videos (**V4**) excel in engagement, fostering cognitive involvement and motivation for learning. Thus, **supporting H3**.

Link to Learning Goals—Identifying what distinguishes a bad visualization from a good one (**L1**) relies on the individuals' visualization literacy and the specific attributes of the flawed visualization scenario. Higher literacy correlates with better X score performance, but not necessarily with reduced performance time. Participants struggled more with complex scenarios, such as excessive visual elements or inappropriate visualization types. The suitability of tested variants for **L1 remains inconclusive**. Variants significantly impact identifying why changes were made (**L2**), with **data-GIFs yielding higher performance and smooth transitions enhancing engagement.** Finally, literacy level and scenarios affect participants' ability to reflect on transforming bad visualizations, with higher literacy aiding elaboration (**L3**). Variant suitability for **L3 remains inconclusive**.

Lessons Learnt and Future Recommendations— Overall, we recommend choosing data-GIFs with smooth transitions (**V1**) to prioritize audience engagement. Smooth transitions guide attention between states, aiding visualization interpretation. Techniques like "do-it-yourself" and "pair analytics" could be further employed to refine frame duration settings. Also, we suggest audio to expedite interpretation, especially with non-expert or complex scenarios. Juxtaposed static visualizations (**V3**), oppositely, can reduce cognitive burden when multiple tasks are involved.

The potential **generalization** of data-GIFs in educational contexts beyond mere design comparisons is significant; yet, it is accompanied by notable **limitations**. Data-GIFs offer a passive viewing experience with fixed sequences, restricting viewer interaction and dynamic engagement. Their brevity and limited information capacity pose challenges outside of contrast-based learning contexts and may impede accessibility for viewers with disabilities. While GIFs offer simplicity in communication, interactive formats provide greater versatility and engagement, ensuring broader usability and personalized experiences.

Regarding the **design of effective data-GIFs**, beyond content presentation and messaging medium, other features such as interactive capabilities, visual cues like highlighting and annotations, and aesthetic considerations warrant exploration to enhance engagement and effectiveness. Individual differences, including visualization literacy levels and background, influence learning performance and user engagement, necessitating further investigation [10, 23]. **Further studies** incorporating diverse demographics, and cultural or learning backgrounds, as well as other evaluation methods, are encouraged to ascertain the effectiveness of data-GIFs in visualization education—particularly in terms of memorability and perceived workload.

7 CONCLUSION

We proposed data-GIFs as educational tools for engaging viewers in identifying effectively differences between bad and good data visualization designs. Our study with 48 participants indicates that smooth transitions are engaging and effective, especially for recognizing *why* visualization designs are misleading. However, our results show that data-videos are generally more efficient than the other variants in terms of consumption time, suggesting that the inherent lack of audio messaging in GIFs is a limiting factor. In our future work, we intend to investigate further the design space of data-GIFs and explore its full potential for visualization education.

8 ACKNOWLEDGMENTS

The paper was partially written in collaboration with the VRVis Competence Center (funded by BMVIT, BMWFW, Styria, SFG, and Vienna Business Agency in the scope of COMET under Grant 854174 managed by FFG).

9 REFERENCES

[1] E. Altintas et al. Evaluation of Use of Graphics Interchange Format (GIF) Animations in Mathematics Education. *Educational Research and Reviews*, 12(23):1112–1119, 2017.

[2] F. Amini et al. Hooked on Data Videos: Assessing the Effect of Animation and Pictographs on Viewer Engagement. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, pages 1–9, 2018.

[3] S. Bakhshi et al. Fast, Cheap, and Good: Why Animated GIFs Engage Us. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 575–586, 2016.

[4] F. Bishop et al. Construct-A-Vis: Exploring the Free-Form Visualization Processes of Children. *IEEE Transactions on Visualization and Computer Graphics*, 26:451–460, 2020.

[5] B. S. Bloom et al. Taxonomy of Educational Objectives: The Classification of Educational Goals. *Handbook I: Cognitive domain*, 1956.

[6] K. Börner et al. Data Visualization Literacy: Definitions, Conceptual Frameworks, Exercises, and Assessments. *Proceedings of the National Academy of Sciences*, 116(6):1857–1864, 2019.

[7] M. Brysbaert. How Many Words do we Read per Minute? A Review and Meta-analysis of Reading Rate. *Journal of Memory and Language*, 109:104047, 2019.

[8] A. Cairo. *How Charts Lie: Getting Smarter about Visual Information*. WW Norton & Company, 2019.

[9] F. Chevalier et al. Observations and reflections on visualization literacy in elementary school. *IEEE Computer Graphics and Applications*, 38(3):21–29, 2018.

[10] Y. Cui et al. Adaptive assessment of visualization literacy. *arXiv preprint arXiv:2308.14147*, 2023.

[11] H. G. Ginott. *Teacher and Child: A Book for Parents and Teachers*. Scribner Paper Fiction, 1972.

[12] J. Heer and G. Robertson. Animated Transitions in Statistical Data Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1240–1247, 2007.

[13] Y.-H. Hung and P. Parsons. Assessing User Engagement in Information Visualization. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1708–1717, 2017.

[14] S. Huron et al. Constructive visualization. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS 14, pages 433–442, 2014.

[15] K. Kim et al. Comparison Techniques Utilized in Spatial 3D and 4D Data Visualizations: A Survey and Future Directions. *Computers & Graphics*, 67:138–147, 2017.

[16] Y. Kim et al. Designing Animated Transitions to Convey Aggregate Operations. *Computer Graphics Forum*, 38(3):541–551, 2019.

[17] C. N. Knaflic. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons, 2015.

[18] H.-K. Kong et al. Internal and External Visual Cue Preferences for Visualizations in Presentations. *Computer Graphics Forum*, 2017.

[19] H.-K. Kong et al. Understanding Visual Cues in Visualizations Accompanied by Audio Narrations. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–13, 2019.

[20] L. Y.-H. Lo et al. Misinformed by visualization: What do we learn from misinformative visualizations? In *Computer Graphics Forum*, volume 41, pages 515–525, 2022.

[21] A. McNutt et al. Surfacing visualization mirages. In *Proceedings of the 2020 CHI Conference on human factors in computing systems*, pages 1–16, 2020.

[22] A. Ottley et al. The Curious Case of Combining Text and Visualization. In *EuroVis (Short Papers)*, pages 121–125, 2019.

[23] S. Pandey and A. Ottley. Mini-vlat: A short and effective measure of visualization literacy. *arXiv preprint arXiv:2304.07905*, 2023.

[24] R. Pekrun et al. Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1):36–48, 2011.

[25] G. Robertson et al. Effectiveness of Animation in Trend Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1325–1332, 2008.

[26] F. Schmidt-Weigand et al. A Closer Look at Split Visual Attention in System- and Self-paced Instruction in Multimedia Learning. *Learning and Instruction*, 20(2):100–110, 2010.

[27] E. Segel and J. Heer. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16:1139–1148, 2010.

[28] X. Shu et al. What Makes a Data-GIF Understandable? *IEEE Transactions on Visualization and Computer Graphics*, 27:1492–1502, 2021.

[29] J. Thompson et al. Understanding the Design Space and Authoring Paradigms for Animated Data Graphics. In *Computer Graphics Forum*, volume 39, pages 207–218. Wiley Online Library, 2020.

[30] E. R. Tufte et al. Visual Explanations: Images and Quantities, Evidence and Narrative, 1998.

[31] B. Tversky et al. Animation: Can it Facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.

[32] Q. Zhi et al. Linking and Layout: Exploring the Integration of Text and Visualization in Storytelling. In *Computer Graphics Forum*, volume 38, pages 675–685. Wiley Online Library, 2019.