Towards Data Storytelling to Support Teaching and Learning

Vanessa Echeverria
University of Technology Sydney,
NSW, 2007, Australia;
Escuela Superior Politécnica del Litoral, ESPOL
Guayaquil, Ecuador
Vanessa.i.EcheverriaBarzola@student.uts.edu.au

Roberto Martinez-Maldonado
University of Technology Sydney,
NSW, 2007, Australia
Roberto.Martinez-Maldonado@uts.edu.au

Simon Buckingham Shum
University of Technology Sydney,
NSW, 2007, Australia
Simon.BuckinghamShum@uts.edu.au

ABSTRACT

Data science is now impacting the educational sector, with a growing number of commercial products and research prototypes providing *learning dashboards* as feedback for both educators and students. From a human-centred computing perspective, the end-user's interpretation of these visualisations is a critical challenge to design for, with empirical evidence already showing that 'usable' visualisations are not necessarily effective from a teaching and learning perspective. Since an educator's interpretation of visualised data is essentially the construction of a *narrative* about that student's progress, we draw on the growing body of work on 'Data Storytelling' (DS) as the inspiration for a set of enhancements that could be applied to data visualisations to improve their communicative power. We present a pilot study that explores the effectiveness of these DS elements based on educators' responses to paper prototypes. The dual purpose is understanding the contribution of each visual element for data storytelling, and the effectiveness of the enhancements when combined. The results suggest that DS elements could add clarity, especially when there are multiple possible stories in a complex visualisation.

CCS CONCEPTS

• **Human-centered computing** → Empirical studies in visualization; • **Applied computing** → Interactive learning environments

KEYWORDS

Data storytelling, information visualisation, learning data.

ACM Reference format:

V. Echeverria, R. Martinez-Maldonado, and S. Buckingham Shum. 2017. Towards Automated Storytelling with Learning Data. In *Proceedings of the 29th Australian Conference on Human-Computer Interaction, Brisbane, QLD, Australia, November 2017 (OzCHI 2017)*, 5 pages. https://doi.org/10.1145/3152771.3156134

1 INTRODUCTION

Data science is now impacting the educational sector, with a growing number of commercial products providing "learning dashboards" as feedback for both educators and students. The opportunities and challenges that are raised by the mainstreaming of student tracking technology, algorithms and artificial intelligence (AI) in education, are the focus of fields such as *Learning Analytics, Educational Data Mining*, and *AI in Education* [2, 18]. Since the interpretation of such visualisations is central to the effectiveness of such tools, there is a corresponding interest in the human factors of learning analytics visualisations [3, 12, 16], especially as evidence emerges that teachers can have difficulty interpreting them [7], which may be due to a disparity between the designer and teacher's perspectives. In some cases (e.g. [5, 19]), learning dashboards have been even detrimental in motivating students. Human-centred design requires that we design student-facing software that is not only usable, but motivating, and informing appropriate action on the user's part. These challenges are the motivation for examining *Data Storytelling* (DS), which refers to techniques employing narrative components (e.g. structures, elements, concepts, plots) to tell a story grounded in data [10].

Journalists have been using data storytelling to inform news in an effective way by *"combining narratives with interactive graphics"* [9, 17]. Ryan [14] and Knaflic [8] connect DS to visualisation principles [20] (detailed in Sec. 2). Some research has explored how DS principles can be used for information visualisation in the communication of scientific data [11], explaining dynamic networks through comic strips for an educational audience [1], presenting "data stories" to a broader audience [17], or supporting presenters to tell a story through data visualisations effectively [8]. However, although these works use various DS techniques, they have mostly been targeted at helping people present data more effectively to engage an audience. Some work has started to be done towards bringing these DS elements into the design space of automatically generated visualisations. For instance, some work has been done to adapt visual elements (i.e. emphasise text or charts) depending on user's cognitive ability to interpret a graph [4]. Authoring tools have recently been deployed with the aim of semi-automatically crafting stories from data with little human input by adding automated textual narratives from the data [13, 15]. Nonetheless, these works have not explored which specific visual elements can support 'the storytelling'.

In this paper, we perform the first step towards defining the DS elements that can be added to the visualisations to support teacher's sensemaking. The proposed DS elements are distilled from data storytelling principles [8, 14, 17]. We present a pilot

OzCHI '17, November 28-December 1 2017, Brisbane, QLD, Australia

ACM ISBN 978-1-4503-5379-3/17/11...$15.00
https://doi.org/10.1145/3152771.3156134

study that explores the effectiveness of these DS elements from visualisations in the context of a teamwork learning activity. We report on a mixed-methods analysis of a number of teachers' reactions to prototype visualisations, with the dual purpose of understanding the contribution of each visual element for data storytelling, and the effectiveness of the enhancements when combined.

2 THE DATA STORYTELLING ELEMENTS

Ryan [14] and Knaflic [8] proposed some practical *golden rules* based on empirical knowledge and visualisation principles (e.g. [20]), to help professionals communicate more effectively with data. We have distilled five storytelling rules, that have been reported by both as follows:

R1. Keep a clear goal: defining which is the purpose of the visualisation and who is the audience, could help the users to find insights more effectively.
R2. Eliminate extra-ink: remove elements that do not add informative value.
R3. Using narratives wisely: labels should be used wisely and should be descriptive to explain important points.
R4. Driving attention: push everything to the background and then use pre-attentive attributes to only highlight important aspects of the visualisation.
R5. Call for action: if the visualisation is clear and concise, it should explain the "story" that users should take from it.

Whilst these rules can be applicable in different ways almost using any type of visualisation, we have distilled a set of visual Data Storytelling design elements (DS elements) that can be related to each rule. Table 1 shows each rule with the corresponding actions that can be performed directly on the visualisations. We have omitted R1 because this rule cannot be translated as a specific DS element but rather to be pervasively considered in the way the visualisation is deployed as a whole.

Table 1: Data storytelling (DS) golden rules with their corresponding DS elements.

Golden Rules	Data Storytelling elements
R2. Eliminate extra-ink	**(A) Decluttering** (*removing*): Data labels, Legends, Data markers, Tick marks, Grids, Axis labels
R3. Using narratives wisely	*Adding*: **(B) Narratives** to critical data points; **(C) Shaded areas** to cluster information
R4. Driving attention	**(D)** Highlighting with **colours**; **(E)** Emphasising key **data points**; **(F)** Making **lines** thicker
R5. Call for action	*Adding*: **(G)** Prescriptive **title** delivering a straightforward insight from the data

For the other four rules (R2-5) we defined their corresponding DS elements as follows. Eliminating extra-ink (R2), can be achieved by (A) *decluttering* the visualisation. This means *removing* or *sending to the background* unnecessary data labels (text in each point), data markers, the grid in the background, legends and tick marks. It also involves rotating labels (horizontally) to improve readability. This could be considered a pre-step, before *adding* other DS elements. We can use narratives (R3) to further explain certain aspects of the visualisation. This can be achieved by adding (B) textual *narratives,* to critical parts of the visualisation; or (C) *shaded areas,* to cluster information. Driving attention (R4) can be supported by first, sending everything to the background (making all visual elements into grayscale as part of the decluttering process) and then, highlighting only the essential parts of the visualisation that help to tell the story. Driving attention could be achieved by (D) *highlighting data with colours*, (E) *emphasising key points* or (F) *making lines thicker.* Finally, to communicate a call for action (R5), a (G) prescriptive *title can deliver a straightforward insight from the data.*

3 PILOT STUDY

3.1 Context of the Study

The dataset used in the pilot study was collected from a groupware tool that supports face-to-face database design [6]. The tool captures team members' actions performed on an interactive tabletop and the partial solutions of the database designs they create. With these data, the tool automatically generates two types of visualisations about participation and performance. Examples of these are shown in Fig. 1. Fig. 1-#1 shows the individual actions (participation) each student within group performed during the activity using the collaborative tool. Fig 1-#2 depicts the team performance and individual participation over time. When any student added an element to the proposed solution, the collaborative tool assigned a score by comparing the partial solution with the teacher's solution. Then, each data point is labelled with the score and the number of different objects (i.e. green and orange dots for attributes/entities and relationships of the database design) of the students' solution at that time. These visualisations were displayed to the students and their teacher just after each session. Results of their reactions have been reported in [6], indicating that, similarly to other dashboards, students and teachers found it hard to focus on specific aspects of participation and performance for post-hoc sensemaking. These results inspired us for adding enhancements that could be applied to data visualisations.

3.2 Study Design

3.2.1 Participants. Five participants (4 male, 1 female, avg. age = 31, sd.=5.36) were asked to participate in the pilot study. All participants had previous experience in teaching and database design. Two had experience on teaching databases. Four are PhD students and one holds a PhD.

3.2.2 Data. The dataset used for this study was collected from a groupware tool that was used support five teams in face-to-face database design (see section 3.1). For this pilot study, we selected two teams (A and B) because they featured quite distinctive stories in terms of participation and performance as following. Participation in Team A was balanced with one participant leading the activity. Participation in Team B was unbalanced, with each participant dominating the participation at different times. The Performance of Team A was high, the team was able to determine all key aspects of the task. The Performance of Team B was low, the team was only able to determine few key aspects of the task.

3.2.3 Design. Having in mind these clear 'stories', we crafted a set of visualisations with DS elements following the rules expressed in section 2. Regarding the participation visualisation, we added A, B, C, D and G DS elements (e.g. see example result in Fig. 1, #3); whereas for the performance visualisation, we added A, B, C, E, F and G elements. (e.g. Fig. 1, #4).

3.2.4 Task. The pilot study consisted of the following steps. *First*, we presented the set of participation and performance visualisations, with and without the DS elements, for Teams A and B to each participant (8 visualisations in total: 4 without DS elements, and 4 with DS elements) gradually. For example, Fig. 1 depicts all the visualisations corresponding to Team A. We asked participants to externalise their reactions as a think-aloud exercise and try to "tell the story" from each visualisation. Although the number of participants in the pilot study is small, we varied the sequential order to which visualisations were presented to minimise the learning effect. *Second*, we asked participants to explain how each DS element was helpful to support the 'story'. Additionally, participants rated each element (from 1 to 5 - being 1 the most helpful and 5 the least helpful) illustrated in the visualisations with DS elements.

We analysed the audio-recordings and notes taken from the comments made by each participant in order to understand how DS elements were used to support the interpretation of the story. Next section presents the results in terms of the effectiveness of the DS elements to support sensemaking and the perceived contribution of each DS element to support the story.

4 RESULTS AND DISCUSSION

4.1 Sensemaking Support

Considering that each visualisation had its own story, here we summarise the results from participants' reactions when interpreting the story from each type of visualisation.

4.1.1 Visualisations about participation. All participants identified the stories from visualisations with and without DS elements. From the former, two participants recognised the *title* as a critical element to understand the 'story' effectively. Nevertheless, a criticism was that the orange *colour* displayed at the first plane dominated the visualisation. They expressed that the colours in the visualisation without DS elements, were better and the opacity helped to observe overlapping participations at a glance. However, one participant indicated that, "*while the information could be easily explored in the visualisation without DS, the [visualisation with DS] helped him to focus only on the specific and critical moments of the team participation*".

4.1.2 Visualisations about performance. It was difficult for participants to infer the story from the visualisations without DS elements. Only one participant commented that the visualisation was about performance. From the visualisation with DS elements, four participants correctly recognised the stories. For instance, participants stated that the information presented in the version with DS elements *"was clearer and consistent"*.

Reflecting upon these results, we can see that, when a lot of

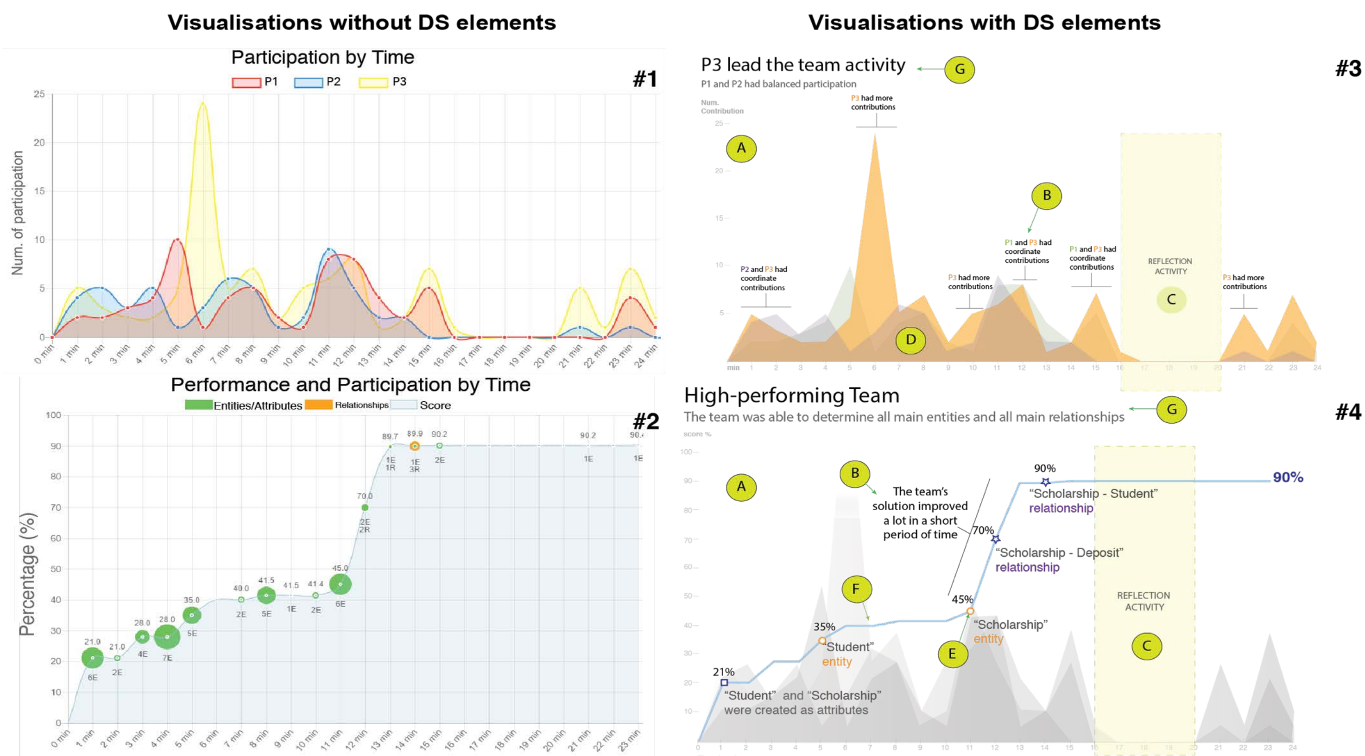

Figure 1: Participation (1st row) and Performance (2nd row) visualisations without (left) and with (right) DS elements.

information is displayed in the visualisation, users cannot determine effectively the story behind it due to the extreme cognitive load process for interpreting it. Another cause could be given by the multiple stories the visualisation tells. By adding the DS elements to this visualisation, we attempted to guide the user to focus on one story at time. This proved effective for the visualisations about performance, most likely because there was one main story to tell. However, in regard to visualising participation, although the DS elements helped participants to focus on only one story at a time, they also wanted to keep other possible 'data stories' available for exploration.

4.2 Helpfulness of Data Storytelling Elements

After exploring how the visualisations with DS elements could support the sensemaking process, we wanted to understand how helpful each particular DS element was. Fig. 2 shows the ratings from the five participants according to the helpfulness of each DS. In the following subsections, we present the results from the most to the least helpful element for each visualisation.

4.2.1 Visualisations about participation. The DS element *(B) narratives* was the most helpful element to support the story (see Fig, 2, left). Participants agreed that this element helped them understand the data points from the visualisation by offering brief information of the things that happened. One participant expressed this as following: "*narratives are descriptive. I could understand what happened here [pointing to the narrative and the data]*". The second most helpful element was the *(G) title delivering a straightforward insight from the data.* Participants mentioned that *"this [title] explains the main idea"*, and *"describes the things that are shown in the visualisation."* Furthermore, *(D) highlighting with colour* was rated as the third most helpful element. Participants said that this element makes it easier to identify what student participated the most. A participant said: *"this is important if your intention is to emphasise this student".* However, some participants agreed that as the colour is brighter, it is difficult to see the overlapping participation by other students. For example, one participant expressed that "*it is harder to observe overlapped participations*". Finally, *(A) Decluttering* and adding a *(C) shaded area to cluster information* were rated as the least helpful elements. In brief, participants agreed that the DS elements helped to make the story behind each visualisation clearer to some extent. However, the colours and opacity were found as detrimental to understand the participation of the three students. For instance, one participant said that: *"I would rather prefer to use the colours of the original visualisation".*

4.2.2 Visualisations about performance. As depicted in Fig. 2

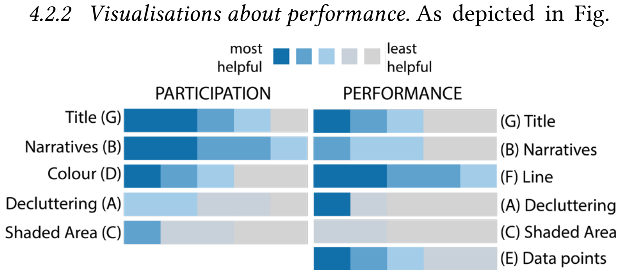

Figure 2: Ranking from each DS element.

(right), *(F) making lines thicker* was catalogued as the most helpful element. Participants expressed that, because of the thicker line, they could understand the main idea of the visualisation. One participant expressed that: *"The line is clear. This [pointing to the line] is more readable".* Next, *(E) emphasising key data points* was categorised as the second most helpful element. Participants stated that showing this element helped them focus on important aspects of the task that occurred in the activity. One participant expressed this as follows: *"these [key points] summarise the most important things of the task which allows me to focus only on this".* The third most important element was the *(G) title delivering a straightforward insight from the data.* Some participants expressed that the title contains the main idea of the visualisation. Conversely, other participants expressed that the title *"was not crucial for supporting the message of the visualisation, because the other elements made the visualisation easy to understand [already]".* Furthermore, *(B) narratives* was rated as somewhat helpful. Some participants said that these annotations *"support the main idea"*, *"describe briefly what happened"* and *"show what the teacher wants to know".* By contrast, participants who rated this element as less important indicated that annotations *"cause noise to the visualisation"* and *"did not add any extra information to the story".* Again, *(A) Decluttering* and *(C) shaded area* were rated as the least helpful to supporting the story.

From these results, we can notice that, as each visualisation has its own story, each DS element can play different roles in the sensemaking process. Participants recognised the pre-attentive attributes (highlight with colour, key points, lines) as the most helpful elements to interpret the message. By contrast, decluttering and the shaded area did not seem to add relevant information to the story but may have helped to make other elements outstand.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we presented a pilot study that explores the potential of a set of data storytelling elements to leverage the teacher's sensemaking process. The prototypes of visualisations with DS elements were generated and presented to participants to provoke reflection and record their reactions. Preliminary results indicate that DS elements that permit exploration of each 'Data Story' in turn could add clarity, especially when there are multiple possible stories in a complex visualisation. We are not claiming that by adding DS elements, these visualisations are better than others. Instead, our aim was to explore if DS elements help users to leverage the sensemaking process and to focus on different aspects of the 'data story'. Further work should include a more detailed analysis on how these DS elements correlates with user's attention, cognitive processes and cognitive load to generalise these results. Also, a deeper analysis on the time taken to correctly interpret the visualisation should be consider in order to determine if these DS elements really support a better sensemaking process. Finally, trying to incorporate these elements automatically into learning visualisations are the next steps in our research agenda.

REFERENCES

[1] Benjamin Bach, Natalie Kerracher, Kyle Wm Hall, Sheelagh Carpendale, Jessie Kennedy and Nathalie Henry Riche. 2016. Telling Stories About Dynamic Networks with Graph Comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, Santa Clara, California, USA.

[2] Ryan S. J. D. Baker and Kalina Yacef. 2009. The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM-Journal of Educational Data Mining* 1, 1, 3-17.

[3] Sven Charleer, Joris Klerkx, Erik Duval, Tinne De Laet and Katrien Verbert. 2016. Creating Effective Learning Analytics Dashboards: Lessons Learnt. In *Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning, Ec-Tel 2016, Lyon, France, September 13-16, 2016, Proceedings*, Katrien Verbert et al. Eds. Springer International Publishing, Cham, 42-56.

[4] Cristina Conati, Giuseppe Carenini, Dereck Toker and Sébastien Lallé. 2015. Towards User-Adaptive Information Visualization. In *Proceedings of Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI, 4100-4106.

[5] Linda Corrin and Paula de Barba. 2014. Exploring Students' Interpretation of Feedback Delivered through Learning Analytics Dashboards. In *Proceedings of 31st International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education*. 629-633.

[6] Vanessa Echeverria, Roberto Martinez-Maldonado, Katherine Chiluiza and Simon Buckingham Shum. 2017. Dbcollab: Automated Feedback for Face-to-Face Group Database Design. In *Proceedings of 25th International Conference on Computers in Education (ICCE 2017)*.

[7] Rebecca Ferguson. 2012. Learning Analytics: Drivers, Developments and Challenges. *International Journal on Technology Enhanced Learning* 4, 5/6, 304-317.

[8] Cole Nussbaumer Knaflic (Ed.). 2015. *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons,

[9] Robert Kosara and Jock Mackinlay. 2013. Storytelling: The Next Step for Visualization. *Computer* 46, 5, 44-50.

[10] Bongshin Lee, Nathalie. H. Riche, Petra Isenberg and Sheelagh Carpendale. 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications* 35, 5, 84-90.

[11] Kwan-Liu. Ma, Isaac Liao, Jennifer Frazier, Helwig Hauser and Helen-Nicole Kostis. 2012. Scientific Storytelling Using Visualization. *IEEE Computer Graphics and Applications* 32, 1, 12-19.

[12] Abelardo. Pardo, Oleksandra Poquet, Roberto Martínez-Maldonado, and Shane Dawson. 2017. Provision of Data-Driven Student Feedback in LA & EDM. *The Handbook of Learning Analytics*, Charles Lang, George Siemens, Alyssa Wise, Dragan Gašević (Eds.). Alberta, Canada: Society for Learning Analytics Research, pp. 163-174.

[13] Puripant Ruchikachorn. 2016. *Integrating Visualizations into the Authoring of Narratives*. Ph.D. Dissertation. State University of New York at Stony Brook, USA. Number: 10158099.

[14] Lindy Ryan (Ed.). 2016. *The Visual Imperative : Creating a Visual Culture of Data Discovery*. Elsevier Science.

[15] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring Narrative Visualizations with Ellipsis. *Computer Graphics Forum* 33, 3, 361-370.

[16] B. A. Schwendimann, M. J. Rodríguez-Triana, A. Vozniuk, L. P. Prieto, M. S. Boroujeni, A. Holzer, D. Gillet and P. Dillenbourg. 2017. Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies* 10, 1, 30-41.

[17] E. Segel and J. Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6, 1139-1148.

[18] George Siemens and Ryan S. J. d Baker. 2012. Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. ACM, Vancouver, British Columbia, Canada, 252-254.

[19] Jennifer Pei-Ling Tan, Elizabeth Koh, Christin Rekha Jonathan and Simon Yang. 2017. Learner Dashboards a Double-Edged Sword? Students' Sense-Making of a Collaborative Critical Reading and Learning Analytics Environment for Fostering 21st Century Literacies. *Journal of Learning Analytics* 4, 1, 117-140.

[20] Edward R. Tufte and Glenn M. Schmieg (Eds.). 1985. The Visual Display of Quantitative Information. American Association of Physics Teachers.