Jun 25th, 9:00 AM

Data-painting: Expressive free-form visualisation

Miriam Sturdee
Lancaster University, United Kingdom

Soren Knudsen
IT University of Copenhagen

Sheelagh Carpendale
Simon Fraser University

Citation

Sturdee, M., Knudsen, S., and Carpendale, S. (2022) Data-painting: Expressive free-form visualisation, in Lockton, D., Lloyd, P., Lenzi, S. (eds.), *DRS2022: Bilbao*, 25 June - 3 July, Bilbao, Spain. https://doi.org/10.21606/drs.2022.257

DRS 2022
Bilbao 25th June — 3rd July

Data-Painting: Expressive free-form visualisation

Miriam Sturdee[a*], Søren Knudsen[b], Sheelagh Carpendale[c]

[a]Lancaster University, United Kingdom
[b]IT University of Copenhagen, Denmark
[c]Simon Fraser University, Burnaby, Canada

*Corresponding author e-mail: m.sturdee@lancaster.ac.uk
https://doi.org/10.21606/drs.2022.257

Abstract: Data visualization can be powerful in enabling us to make sense of complex data. Expressive data representation – where individuals have control over the nature of the output – is hard to incorporate into existing frameworks and techniques for visualization. The power of informal, rough, expressive sketches in working out ideas is well documented. This points to an opportunity to better understand how expressivity can exist in data visualization creation. We explore the expressive potential of Data Painting through a study aimed at improving our understanding of what people need and make use of in creating novel examples of data expression. Participants use exact measures of paint for data-mapping and then explore the expressive possibilities of free-form data representation. Our intentions are to improve our understanding of expressivity in data visualization; to raise questions as to the creation and use of non-traditional data visualizations; and to suggest directions for expressivity in visualization.

Keywords: painting; information visualisation; data visualisation; expressivity

1. Introduction

Data visualisations always *express* data but are not always *expressive*. We use them to make sense of data, in a variety of visual arrangements, valuing clarity and appropriateness of the method for the resulting image (Tufte, 2006). However, in some instances, the use of visual embellishment can have positive effects (Bateman et al. 2010), such as to engage new audiences (Akbaba et al. 2021). The rise in popularity of Infographics also suggests that figurative imagery in data visualisations can add value in terms of readability and accessibility (McCandless, 2012), and Ware (2012) states that even cave painting can be seen as a type of visualisation. Yet, on the topic of art-science visualisation collaborations Campbell & Samsel (2015) postulate that *"…the first question to ask is what is the primary goal of the collaboration? Is it a visualization? A piece of art?"*. We ask, could it not be both?

There are calls that the 'next generation' of data visualisation tools should be based upon an 'expressive and effective' model (Ribarsky & Foley, 1994). The problem with designing such tools is that with high levels of expressivity comes the possibility of reduced comprehension, and with highly effective tools comes the problem of rigidity.

Just a decade ago, even using the word 'expression' in conjunction with traditional visualization may have seemed odd. After all, traditionally visualization has been defined as a mapping (Marr, 1982) from the data to the visuals – where the visuals should be definable in terms of the data. Drawing from Hadamard's correspondence with 100 famous mathematicians and physicists (Hadamard, 1945), it is apparent that for most, their process of discovery of new scientific understanding came initially from drawing images and diagrams where the meaning was not at first clear, sometimes even to themselves. Considering this, we suggest there may be more we can learn by thinking more broadly about expression in visualization. There may be a similar benefit in working with data, not only as clear and precise visualizations, but also as 'cloudy' data representations, which may be a reflection of the mindset and questions of the creator, as well as of the data itself.

There is a recent upsurge of discussion in the visualization literature about the importance of expression in visualization. Novel applications and tools such as DataInk (Xia et al. 2018), Data Illustrator (Liu et al. 2018), and Charticulator (Ren et al. 2019) all offer some form of expressive interaction. However, compared with the wealth of possible expressive visualizations, these tools still remain largely within the context of existing visualization conventions.

By moving towards a better understanding of the meaning of expression in data visualization, we hope to take a step towards discovering new ways to work with, think with, and communicate with data (Card et al. 1999). In addition to the value that expressive forms of creation may have for the creator, expressive visualization tools may also provide avenues for engaging new and broader audiences, for communicating data, in new and exciting ways. To further expand our understanding of the potential of expressivity in data visualization, we conducted a study with 20 participants of varying levels of skill in data visualisation, who each created an expressive data visualisation utilising *data painting.* We reflect here upon insights and processes relating to our approach and consider how to further explore the possibilities of data painting.

2. Background

2.1 The Art of Visualisation

Painting has always been about conveying information, whether an emotion, a scene, or hidden meanings embedded in the innocent or bizarre. For example, at first look, Cy Twombly's *Nine Discourses on Commodus* (1963) look like abstract paintings with rough scales and tallies, but behind the art is the story of a cruel and unpredictable Roman Emperor (Neely, 2010). More directly linked to data are Jill Pelto's paintings depicting arctic climate change (2016), Nathalie Miebach's marine environments (Campbell, 2015), or the iconic *Dear Data* postcards (Lupi & Posavec, 2016).

In the professional sense, we can gain insights from how people sketch data (Lee at al. 2013; Walny et al. 2015) and use it to develop tools (Lee et al. 2015). In the public sense we can *democratise* data, by providing physical toolkits to create physical visualisations of highly personal data (Thudt et al. 2018), or the tools to make them (Wun et al. 2018). These latter examples value making data accessible and personal, but there are also ways of making more complex data expressive, such as through the medium of *Data Comics* (Bach et al. 2017) which has also been developed into an accessible tool to aid their digital development (Kim et al. 2019). Similarly, the storyline tool developed by Tang et al. (2019) takes a well-known paradigm of storytelling and enables it digitally, and, of specific interest is Tang's observation that *"people usually disobey a well-established design principle ... to create storylines while considering narrative details".* The disobedience the researchers found in their participants' hand-drawn sketches directly informed the tool design.

2.2 Data Literacy & Data Personalization

To help novices author their visualizations, different types of software such as *Tableau Public*, and *Microsoft Power BI*, offer people templates into which they can place their own data. While these applications can be limited stylistically, they usually follow established conventions for data visualization to aid readability and data investigation (Cleveland, 1993; Murray, 2013). However, even with expert help, novices find it difficult to develop visualizations (Grammel et al. 2010). More accessible approaches might help bridge the gap between data and visualisation, such as tangible tiles (Huron, 2014) or simple print-making (Wun et al, 2018), whilst also offering opportunities for breaking convention. In this manner, Li et al. (2020) sought to combine the direct artistry of creative tools with programming languages to offer direct feedback to the creators whilst designing visualisations. Creative ways of enabling data artistry can only serve to further expressive potential in visualisation, especially as our underlying data becomes more complex, and increasing numbers strive to better understand and manipulate data to express their intent and personality.

2.3 Expressivity in Existing Applications

The full spectrum of such applications that people can use to make data visualisations is vast, and each offer different approaches and have varying interaction and output possibilities (Mei, 2017), hence we discuss some well-known examples.

In their analysis of existing visualisation authoring tools, Xia et al. (2018) provide detailed reasoning as to the potential of expressivity in data visualisation. *DataInk* is part of a new breed of data-bound tools for drawing visualisations, alongside examples such as i*Volver* (Mendez et al. 2016), *Charticulator* (Ren et al. 2019), *Data Illustrator* (Liu et al. 2018), and *StructGraphics* (Tsandilas et al. 2020). These tools show promise of increased customisability and expression in designing and interacting with data visualisations and stand in contrast to widely used template-based tools for non-experts, such as *Microsoft Excel, SPSS,* and *Tableau*. While the examples of data-bound drawing tools offer a glimpse of expressive possibilities, established tools such as *Adobe Illustrator* today provide great expressive freedom, but lack any notion of data binding. At the far end of the 'expert' spectrum, we might also consider programming libraries used to generate visualisations, such as *d3.js* (Bostock et al., 2011), which support novel representations but require considerable programming skills. At face-value, more complicated applications have increased support for expressive features and customisation, but they remain less accessible for non-experts.

Sketch-based applications also offer opportunities for expression and personalization. For example, *SketchViz* (Brade et al. 2012) focuses on the basic forms of sketching and offers an immediate, hands-on approach which resonates with traditional mark-making. In contrast, *SketchInsight* (Lee et al. 2015) and *SketchStory* (Lee et al. 2013) allow people to draw and interact with simple charts directly on a whiteboard. Xia et al.'s *Data Ink* (2018) focuses on the directly drawn image in relation to glyphs to support expressive outputs, and similarly, *Data Driven Guides* supports image import and vector-based editing of data to create highly sophisticated graphical outputs (Kim et al. 2017). These expression-focused applications all rely heavily on direct sketching and vector-based drawing to support visualisation. Given the acknowledged value of sketching in expressive data visualisation tool design, we propose to look toward other forms of artistry to inform expressivity in the design and development of novel data visualisation tools, specifically, to examine the practice and process of painting in conjunction with data visualisation.

Some applications built with expressivity in mind are developed using a bottom-up process, whereas adding expressivity to existing applications employs a top-down process. However, as neither approach has thus far generated highly expressive tools, we propose something different: *Studying how expressivity can inform the design and development of data visualisation applications.*

3. Concept & Exploration

We designed a study in which we use a novel approach to data-to-colour mapping. Specifically, we use volume of paint as a direct mapping to data value and asked participants to imagine that the paint contained data that they mapped onto a blank canvas. We curated this by using medical syringes to extract acrylic paint, and silicone brushes and receptacles to limit paint-loss: for example, a data point of 1.5 would translate directly to

1.5ml of paint. Our approach links numeric data to a free-form creative activity – painting – to allow people the full freedom of expression without the constraints of existing concepts of data visualisation or programming.

3.1 Participants

Twenty participants aged between 18-56 took part (8M/12F). Participants, with normal or corrected-to-normal vision, were recruited via posters on public notice boards and via snowball sampling. Individuals had to be sighted, as the work was with visual media. Participants filled out a demographic questionnaire assessing familiarity with data visualization software and artistic practices. Less than half of the participants (8) had high levels of familiarity with visualisation in practice (e.g. computing/design students and staff). The remainder (12) came from the general university population (staff/students), and visitors to the site who had seen the research call and were interested.

3.2 Materials

The study took place in a private room (library/meeting room) within the visualisation department, which had good natural light and large tables. We provided participants with nine colours of identical brand acrylic paint (red, light blue, dark blue, brown, black, green, yellow, purple, and white), and nine syringes ranging from 5-10ml capacity. To support accuracy, we also provided silicone pots and a range of silicone ‘brushes’ and spatulas which do not retain paint as bristles do. We used non-porous, smooth A1 cardstock, with additional sheets if needed. We also provided A4 sheets of paper for notes and sketches, and pens, fine-liners and coloured pencils for annotation (Figure 1). Cleaning materials were provided. Participants were asked to think aloud as they worked and live transcription (touch typing) was taken in real time. They were also filmed from above and the side so that the process could be revisited during analysis.

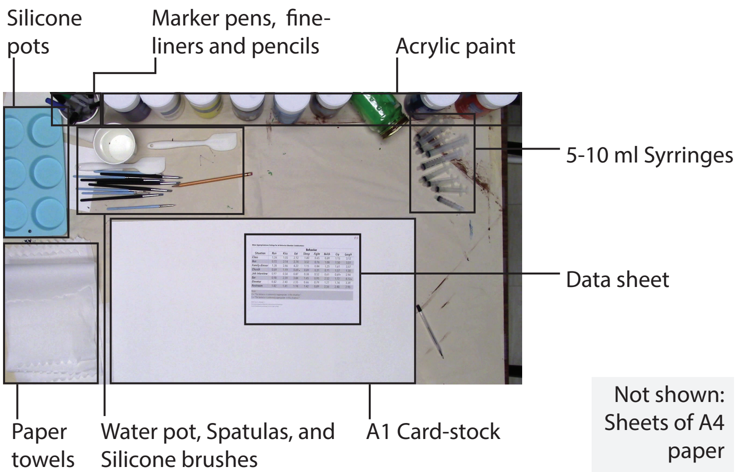

Figure 1 Study materials

3.3 Data

We provided a sample dataset adapted from a survey of behavioural appropriateness and situational constraint, containing an 8 by 8 matrix shown in Figure 2 (Price & Bouffard, 1974). The numbers were adapted to range from 1–5 due to limitations of paint volume versus paper area. For example, if participants wished to use paint in its thinnest possible state, the area covered by 5ml would encompass an entire A1 sheet. We asked participants to measure the paint as close to the second decimal point as possible for accuracy. Participants were given no constraints as to which data points they wished to link to paint, or how they wished to use the paint once measured.

Mean Appropriateness Ratings for 64 Behavior-Situation Combinations

Situation	Behavior							
	Run	*Kiss*	*Eat*	*Sleep*	*Fight*	*Belch*	*Cry*	*Laugh*
Class	1.26	1.05	2.12	1.80	0.65	0.89	1.15	3.12
Bus	0.72	2.14	2.74	3.52	0.76	1.08	1.50	3.51
Family dinner	1.28	2.46	4.22	1.15	0.84	1.25	1.61	3.57
Church	0.69	1.19	0.69	0.89	0.31	0.71	1.57	1.30
Job interview	0.97	0.54	0.87	0.38	0.52	0.61	0.69	2.94
Bar	0.98	2.59	3.84	1.45	0.95	2.52	1.72	4.12
Elevator	0.82	2.40	2.55	0.66	0.79	1.27	1.74	3.39
Restroom	1.42	1.41	1.18	1.42	0.89	2.56	2.40	2.95

Note:
0 = "The behavior is *extremely inappropriate* in this situation."
5 = "The behavior is *extremely appropriate* in this situation."

SOURCE: Price, R.H. and Bouffard, D.L.
Behavioral Appropriateness and Situational Constraint as Dimensions of Social Behavior
Journal of Personality and Social Psychology, 1974, Vol. 30, No.4, 579-586

Figure 2 The data set presented to the participants, adjusted by halving original numerical data (Price & Bouffard, 1974)

3.4 Protocol

Participants filled out a consent form, and a pre-study questionnaire. The study facilitator then explained: 1) That they would be working with paint which related by volume to data points on a supplied sample set; 2) That they were free to choose any number and type of data points to represent; 3) That they were free to use as many sheets of card as desired; 4) That there was no wrong method or style of expression; 5) There was no time limit; 6) That they would like the participants to think aloud during the data-painting process. A post-study questionnaire querying the experience, and memorability of the resulting data-painting was given at the end of the study.

3.5 Analysis

We used an open-coding and annotation process to tease out themes and categories. All paintings were fixed to the walls of a large meeting room. We spent two hours to immerse ourselves in the data, talking about, describing, and noting down observations. We produced 90 post-it notes containing our initial findings. By going over these, and linking them to the observations and paintings, we condensed our themes and started to draft descriptions of these.

4. Process & Results

Participants produced 20 visualisations of varying process, format, and presentation. They embraced *data painting* with varying degrees of expressivity. The open coding process – utilising the final data paintings and observation notes (including direct transcriptions) – produced 90 codes ranging from top level dimensions (e.g., artistry and temporality) to specific codes (e.g., symbolism and shapes).

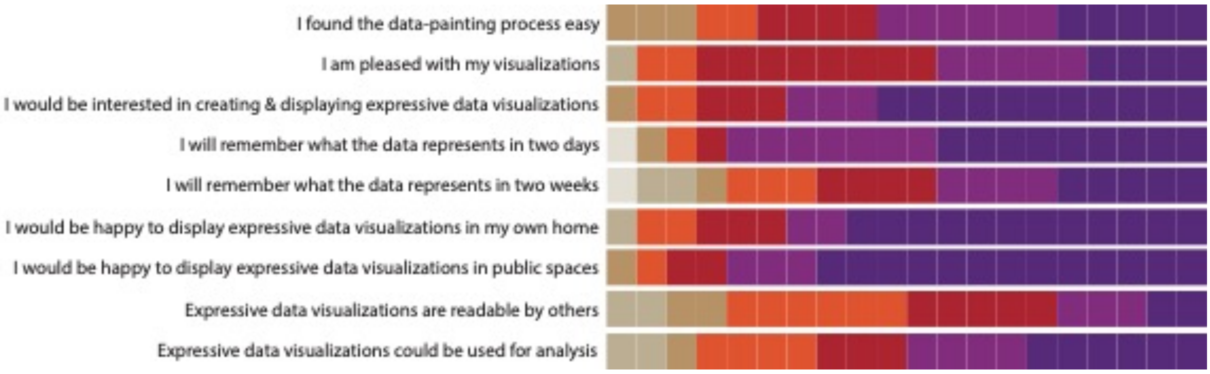

Figure 3 Likert responses. From left to right: Shown as a heat map with low agreement (rated 1, shown as beige) – High agreement (rated 5, shown as purple)

The pre-study questionnaire collected basic demographic data and asked participants about their level of experience with data visualisation, as well as their level of engagement with artistic practice and if they regularly collected personal data (e.g. forms of self-tracking). All but one participant stated that they were familiar with data visualisations at various levels (from basic bar graphs and pie charts to complex engineering data), and all but two participants stated that they collect personal digital data – these participants also classed themselves as 'not artistic', a feeling shared by only one other.

The post study questionnaire used 7-point Likert scales to evaluate the data-painting experience (Figure 3). These questions were based on our own insights of the process and also to currently unpublished work examining memorability of data. Whilst most found the concept and process easy to grasp, there was a mixed reaction to data-painting, although higher agreement was seen when participants were asked whether they would be interested in creating and displaying expressive data visualisations in everyday life. Interestingly, most participants felt that they would be able to remember the underlying data in two weeks, with memorability likely to wane as time went on. The display potential for expressive visualisations was thought to be high (both public and private contexts), but feelings about readability and analysis were mixed: several participants commented that it depends on the level of expressivity and process of creation.

The resulting images were varied and often unexpected – ranging from carefully executed bar charts and Nightingale Roses (the latter made serendipitously by non-experts), to highly detailed and embellished figurative works, similar to the continuum in data sketching found by Walny et al. (2015) (Figure 4). All participants were able to understand and execute the process, regardless of background and familiarity in either data visualisation or artistic practice, except for P19 who measured the paint but did not seem to grasp the concept of data explicitly linking to the paint. Through observation notes and open-coding of the imagery, we identified high level themes within the process, and subsequently used the observation notes and affinity diagramming of the codes to generate insights for expressive data visualisation tools.

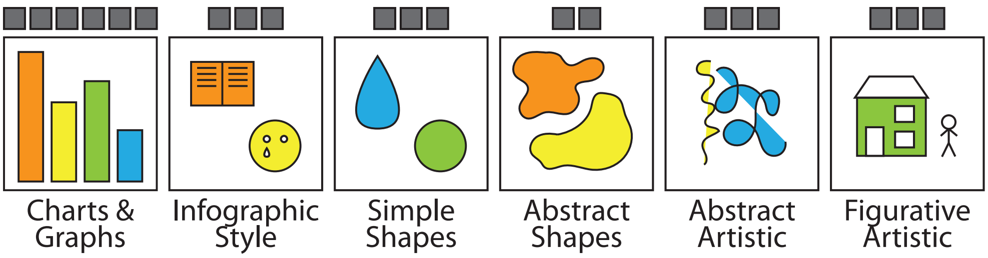

Figure 4 Continuum of responses from data painting with counts.

All participants made use of the range of tools available (spatulas, different brush heads, syringes), with some requesting extra tools (ruler, P20), or improvising with other items (paper-towels, P3). P2 and P3 also used their hands to directly manipulate the 'data'.

In the *infographic-style* and *figurative-artistic* categories, participants made use of recognisable imagery. Although in some cases, the 'data' itself was worked to fit the image, rather than to accurately show the spread (e.g., P9). Where simpler shapes were chosen, the accuracy of the data was clearer, but the expression became more limited, embedded instead in the choice of colour and placement (P8 & P11). P12 attempted to blend both approaches, but although the relation between the data points was clear, the paint was carefully built-up and 'physicalized' (3D) rather than smoothed across the page. The figurative-artistic approach is best shown in P10 and P18, where the full amount of paint was used, without compromising the vision, although the paint went further than anticipated.

All participants started the process by examining the dataset and choosing data-points to represent. After this, half of the participants began to make notes and sketch ideas. Due to the unpredictable nature of the paint, several participants had to abandon their original idea, and embraced alternate modes of expression. Participants either measured out all of the data-paint at once or extracted measures one by one as they added a data point. Several participants used the syringes to squirt the paint directly onto the paper (8/20) whereas just over half (11/20) chose to place the paint in the palettes to work with afterwards – in this case, some chose to dilute the paint with measured amounts of water, or mix the paint in a systematic manner relating to their data set.

Participants 1, 7, 8, 11, 14 and 15 all chose to use the spatula to rigorously spread the paint to its full extent, embracing the concept that the image should portray the data as accurately as possible. The spatula was also used by P4 to create contrasting areas, P2 for overlaying and spreading, and by P17 to apply partially mixed paint straight from the palette to the cardstock. Those that chose to apply the paint directly from the syringes (as above) either smeared the paint with a spatula or brush afterward, or, in the case of P16 (and to some extent P4) left the paint in relief to dry as a physical representation. To support expression, P3 utilised the paper towels provided to smear the paint into a 'Rothko' style piece of artwork, whereas P4 used their fingers to make marks.

Participants who chose to fully utilise the paint over a large area found the unexpected spread led to rethinking their output: P1 abandoned the idea of a figurative image, and then a simple bar chart, to create a blended bar chart which fully utilised the paint; P2 and P8 extended the cardstock to maintain their vision. P15 spent time layering and waiting for each to dry before spreading out the next data point, but it took too long, and halved their data points. More 'artistic' visualisations tended to maintain original plans, but possibly at the expense of accuracy. Most of those that chose an 'exacting' method finished their painting when all the data had been 'used up', then chose to annotate, or leave the image unembellished. Several added keys, scales, or titles (e.g., P12/13/16). Those that worked with the paint wet, and/or in relief, had more control over the end point (e.g., P6/16), these participants 'felt' that the data-painting was completed when it supported their vision. This suggests that expressivity can be approached by following a rigorous plan, or by embracing serendipitous results – and learning via mistakes and challenges. The complete data set can be seen in Figure 5.

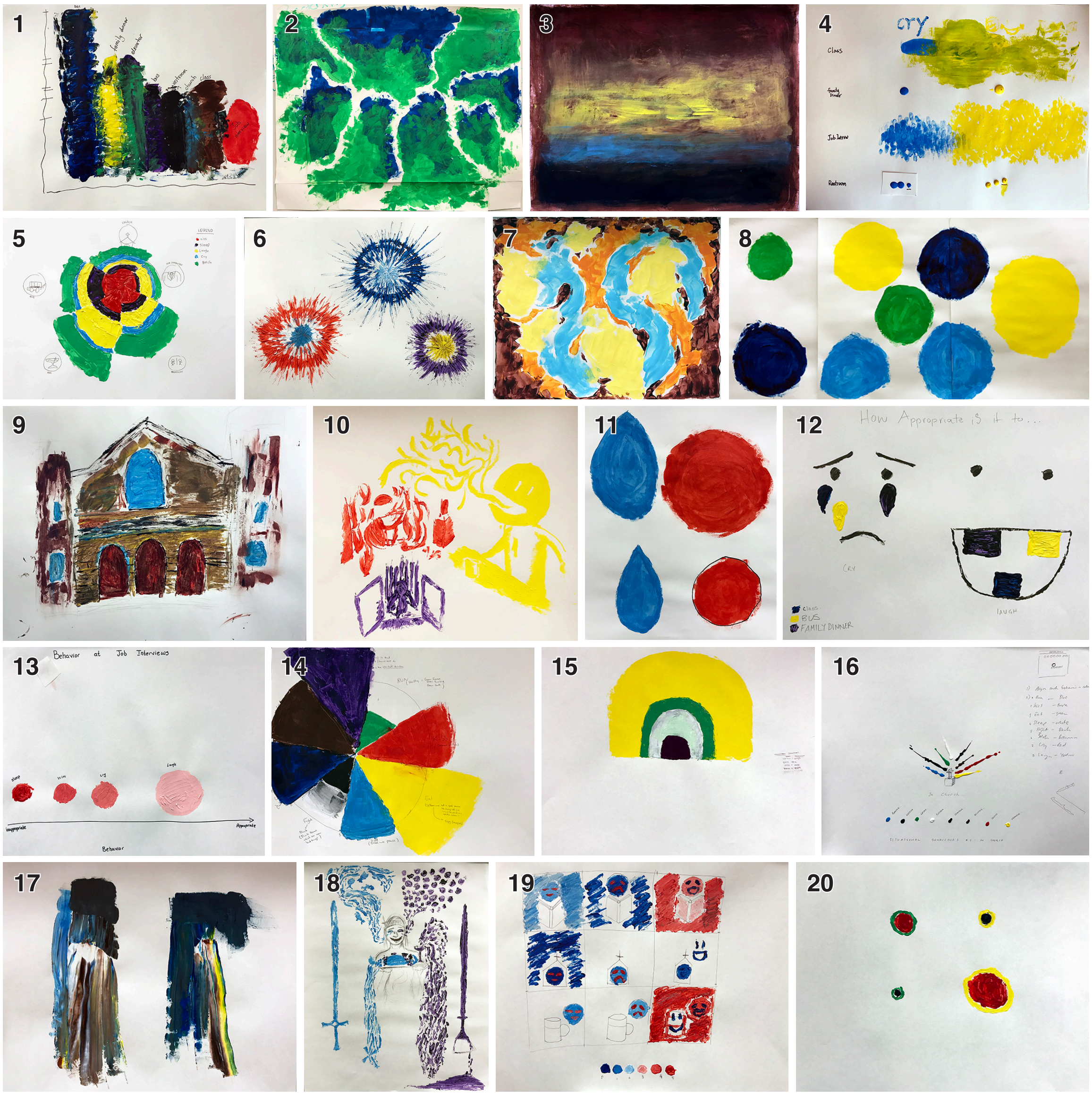

Figure 5. Participant Data-Paintings

5. Reflection & Implications

This section describes and visually demonstrates implications for designing data visualisation tools based on our insights. Whilst we discovered many parallels to existing research and tools e.g. layers (Gleicher et al. 2011; Javed & Elmqvist, 2021), and blending (Figure 6), they take on new meaning when examined within the context of data paint.

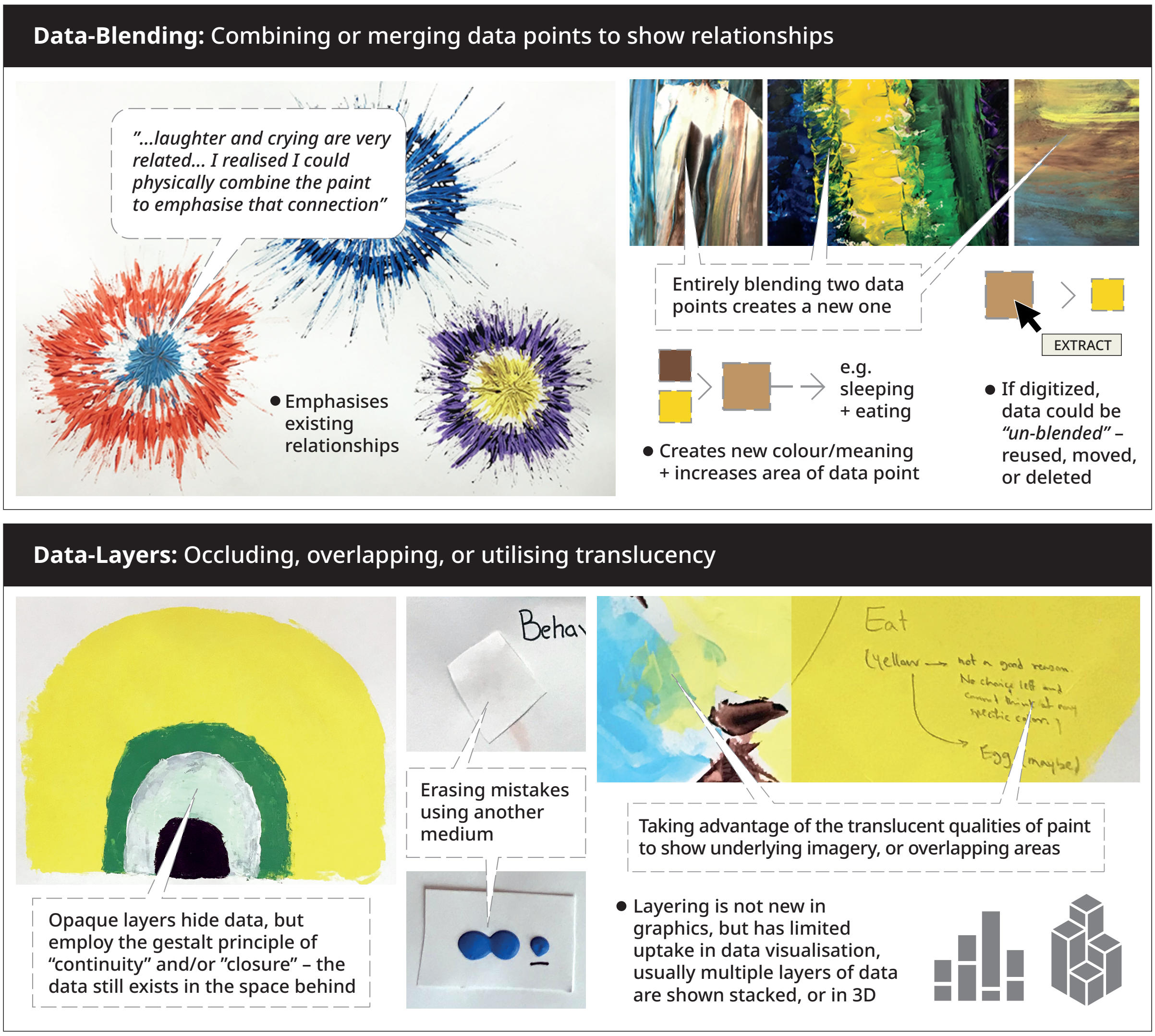

Figure 6 Reflections on blending and layering: Top – blending data points using data-as-medium, and potential digital applications; Bottom – Layering using data-as-medium, which has parallels to (Gleicher et al. 2011, Javed & Elmqvist, 2021), but could be re-thought in this context.

5.1 Data-as-Medium

Data-painting created a semblance of a hybrid material. Several participants described paint as 'data': *"I'm just measuring out my data"* (P1), blurring boundaries. P10 said *"... by turning it into a resource ... I was limited by it – not just looking at a couple of figures on paper."* Data-as-medium also links to other concepts: P2 found that air-bubbles were an absence of data, which could be realised as 'zero-points' (Figure 7), and also found that introducing another medium affected the appearance and texture of the paint (e.g., using tape), which could an allegory for combining datasets, or texturization.

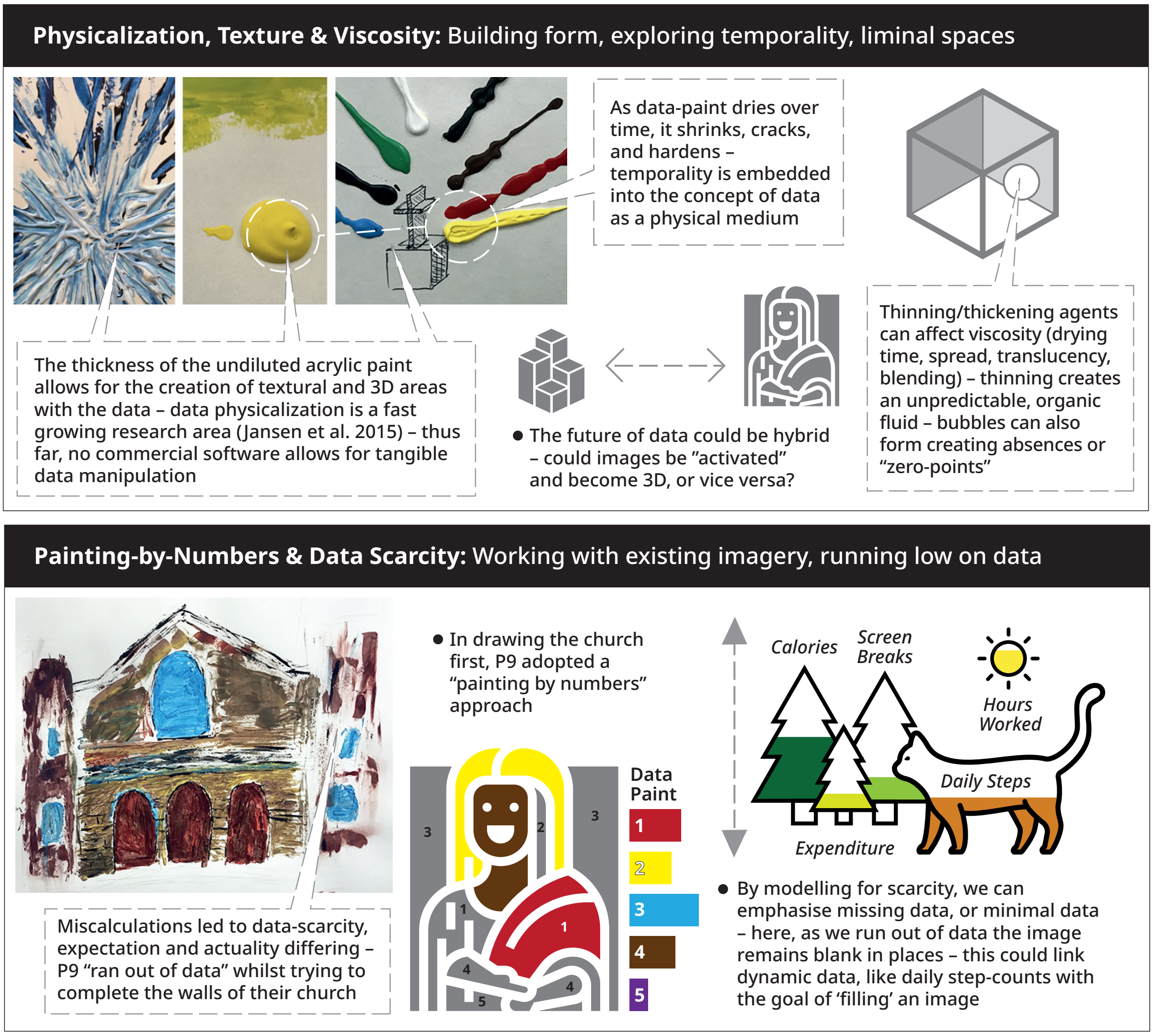

Figure 7 Reflections on physicalization and data-as-medium: Top – Data physicalization is becoming more interactive via shape-changing interfaces (Hogan et al. 2021), but variable physicalization and temporality of medium is relatively unexplored; Bottom: Possible uses of painting with data directly could be in self-tracking.

5.2 Manipulation & Disingenuity

Participants had contrasting approaches to working with data, some 'manipulating' data to fit their vision (e.g., building layers, blobs of paint): P12 built up 'tears' and 'teeth' to fill the desired space, due to an inaccurate estimation of spatial relationships. P3 used multiples of the data points (whilst maintaining the relationship) as the contrast between colours was *"too muddy"*. Some had more direct approaches, (P8, P11) carefully spreading the paint in a consistent manner to show the data as accurately as possible, or P9 who found they "ran out of data" and simply left their image unfinished (Figure 7).

5.3 Serendipity

Serendipity relates to how participants transformed their approach via unexpected discoveries. In current tools, this would mean re-starting when new insights occurred, but by encouraging a 'forward-only' linear process, new forms of data-expression emerged, and participants embraced what came before. Edges were a source of frustration, as paint spread further than expected, or participants realised they should have started centrally. In most cases, participants chose not to extend the canvas, instead moving along the edges, or adapting their process. The engagement of the participant throughout the process, rather than choosing an output and running a program allows for these insights: we value speed and efficiency in digital processes, but *slow processes* can offer additional insights, and mistakes can enable realisation – the linear nature created opportunities.

5.4 Temporality

Temporality relates to data points changing organically over time, rather than directly controlled animations – although this is linked to existing time-lapse data (e.g., climate change modelling). Here, this could be seen in the drying of the acrylic paint, which caused a fade in the glossy surface, shrinkage, and the transition between 'soft' and 'hard' data – the point at which the data point becomes consolidated, final, and in the case of the built-up paint: a physicalization (Figure 7). The temporality of data visualisation could also be seen in recording the process of painting, as a form of provenance and understanding for those interested in interpreting the data.

5.5 Using 'White' on White

Participants found that white paint worked well when used in relief (P16), or when blended loosely with other data points (P17). P13 used exact measures of white to show values in both size and colour simultaneously, whilst P15 used white as a contrast. Some stated they would not use white as it was hard to see, although mentioned the possibility of *"hiding data in plain sight"*.

5.6 Familiarity, Personalization & Emotion

Several participants made observations as they worked. The variable spread of paint made differences acute, and the opportunity to 'mix' data allowed participants to show relationships where there was overlap or interest. P2 stated *"...manipulating the data helped me to spend more time thinking about it"* outlining the potential of extended process in visualisation and working concretely with data mapping. P9 suggested *"…it got me thinking about the data"* – essentially, painting produced *data-familiarity*. Likewise, P7 emphasised: *"while you are doing it you are learning about the data, but afterwards people won't understand... they just see a painting, but for yourself there is benefit to that."* If we examine this in terms of Ware's stages of data visualisation (2012), we suggest that not only are the participants playing the part of *Human Information Analyst*, but also *graphics engine*, by manually manipulating the data and creating the visualisation from basic information.

Some participants had a bias toward 'emotionally-charged' data: those with an ascribable emotion such as 'fighting' or 'kissing'. This also influenced colour choice – dark or rich colours (purple/red) were used for emotionally linked concepts as seen in colour association studies (Ou et al. 2004). This suggests that 'emotional' or 'human' data was easier to make expressive due to existing associations with concepts and artistic expression.

Several participants mentioned they would have liked to have used their own data rather than the supplied dataset, as they felt this linked to the concept of creating artworks with data (Thudt et al. 2018). P8 said *"it feels like a very personal procedure, like if this were my data, it would be more meaningful … and I would think of adding touches like finger painting, things to make it more personal... Like putting in a piece of yourself."* Others said they tried to imagine the data belonged to them as it helped with the creative process. If heightened expressivity is intrinsically linked to personal aspects of data, we might argue for 'expressive' and 'standard' visualisation applications falling under different remits, with corresponding audiences: P3 and P6 felt that whilst their visualisations were very personal, they would still know the meaning, and enjoy the imagery privately. The physicality of painting appeared to evoke a connection or familiarity that is not found when working digitally with data.

6. Discussion

The concept of *data-painting* allows us to bridge a gap between expression, artistic practice, and data visualisation, with the aim to provide a process for exploring data understanding. Alongside the insights shown here, there are open questions inspired by the concept, such as how to address differences in personal abilities (democratisation), visualisation production and purpose, audience, and analytical possibilities. We focused on exploring ways in which people might better tap into their own creativity and expressiveness when engaging with data, but are not suggesting data-painting as an end point. For example sketching is used in many ways as an important part of the visualization and ideation process, and a way in which people approach develop a better understanding of data through visualization. We consider data painting in this way too. This mirrors Hadamard's (1945) survey – the data-painting process allows for thoughts to take shape and become embodied on the page. The *process* is as important as the *output*, it gives us a freedom that other approaches have not. In exploring these freedoms we can discover new insights. This exploration of data-painting is not exhaustive however, there are limitations in the approach which leave the work open to further interpretation – for example, how might a structured study with experts and non-experts play out? Are there cultural differences? Is it possible to reach saturation or is the process unbounded? We also acknowledge that unfamiliarity with paint or artistry may skew results, and more participants are needed to tease out the intricacies of the approach.

6.1 Readability vs. Artistry

When expressivity is increased, we come across the problem of *readability*. Whereas *chart-junk* ranges from 'clutter' (Tufte, 2006) to a useful memory assist (Bateman, 2010), when we create a visualisation as art, meaning can be lost, or available only to its creator. There is a trade-off between true expression and comprehension – such as choosing between design execution and rigorous execution (Bigelow et al. 2017), or the laboratory versus the 'cacophony' of an artists' studio (Samsel et al. 2013). However, by automating the ideas and processes we have identified, could we change that: by developing a 'reader' program that tells us the saturation of colour, produces readings of coverage or height, or unpicks the process in a meaningful way?

Most of the mathematicians and physicists in Hadamard's (1945) survey spoke of the importance of creating images and diagrams that were only personally understandable as an important part of their scientific process. If we consider visualisation as a personal representation with hidden meaning, then direct analysis is not always necessary (Thudt et al. 2018). Therefore, it makes sense that any application being built to support this would have different options for public and private viewing. We might also design for *uncertainty*, where data provenance is unknown, or the data is in flux. Ware's (2012) statement is applicable here: *"All are meaningful to those who understand them and agree to their meanings"* – expressive visualisations can be read and understood when the process behind their creation can be shared and explained.

6.2 On Belonging

By working with data in a hands-on, artistic manner, we found that participants were more personally involved in the resulting visualisation, and that they considered the data more deeply, representing more than if they had been making a chart using current software. This mirrors comparisons between *iVolver* and *Tableau* (Mendez et al. 2017) – *iVolver* requiring more planning and thought, compared to the quick, multiple choices of *Tableau*. Mendez et al. stress the importance of allowing for breadth in user-exploration when creating visualisations, and "thoughtful exploration". An important part of the *data-painting* study was the emphasis on *process* over *convenience*.

If we consider process and audience, expressive data seems to immediately find a home in personal data analysis. By making analysis part of the process, and outputs objects of expression rather than function, we produce visualisations. That relate to more than one issue. By making working with data a creative process, we allowed the user to feel *ownership* of that data as it is *physically put into their hands*, and the act of creation means they also feel a sense of belonging to the final visualisation. Many participants were explicit about wanting more time, wanting to start again, and feeling deeply engaged with the process. Enhancing potential for expressivity in data visualization may offer opportunities for increasing data engagement.

6.3 Future Work, Limitations & Conclusions

We suggest that ideas drawn from this concept may prove useful as an ideation process, to avoid design fixation, and help us explore novel data representation design. This process provides a link to the data and considerable expressive freedoms, and may provide a free-form personal ideation platform to personally, and privately, explore ideas as they are forming (Hadamard, 1945). This also has links to embodied cognition and the idea of the 'thinking hand' (for an example, see Bredekamp's essays on Galileo, 2019, or the 'Marvel Method' in comics creation – credited to Stan Lee), where action via artistic medium provides novel output, something that is important to explore further within the context of visualisation. If data painting helps people think about data, it may in turn help designers develop a deeper understanding that leads to better data representations.

If we also change how we think about 'audience', we may be able to reconfigure what data means to us and what it can be used for. Given the clear interest in the community about allowing expression, we designed and conducted this study to expand our understanding. While we expected existing visualisation tropes, we also saw a variety of directly expressive work, which suggests variations into what construes a data representation, and exactly what the possibilities and limitations of readability are.

Could similar freedoms and expressivity can be emulated in digital form? Are there features of working with data in an analogue manner that offer expressive experiences and opportunities for expressivity that we cannot provide digitally, or that will not work if translated into digital contexts? We hope it is possible that new ideas about interactivity can be inspired by how participants painted with data. By further exploring expressivity, we might discover new ways of supporting visual thinking, and working with data – potentially benefiting the field of data visualisation.

Acknowledgements: This research was supported in part by: the Natural Sciences and Engineering Research Council of Canada (NSERC); SMART Technologies; and Alberta Innovates - Technology Futures (AITF), and also in part by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 753816.

7. References

Akbaba, D., Wilburn, J., Nance, M. T., & Meyer, M. (2021). Manifesto for Putting 'Chartjunk' in the Trash 2021!. *arXiv preprint arXiv:2109.10132*.

Bach, B., Riche, N. H., Carpendale, S., & Pfister, H. (2017). The emerging genre of data comics. *IEEE computer graphics and applications*, *37*(3), 6-13.

Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010, April). Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2573-2582).

Bigelow, A., Drucker, S., Fisher, D., & Meyer, M. (2016). Iterating between tools to create and edit visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 481-490.

Brade, M., Brändel, C., Salmen, A., & Groh, R. (2012, September). SketchViz: a sketching interface for domain comprehension tasks illustrated by an industrial network use case. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies* (pp. 1-4).

Bredekamp, H., 2019. *Galileo's Thinking Hand*. de Gruyter.

Campbell, B. D., & Samsel, F. (2015). Pursuing value in art-science collaborations. *IEEE computer graphics and applications*, *35*(1), 6-11.

Cleveland, W. S. (1993). *Visualizing data*. Hobart Press.

Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, *10*(4), 289-309.

Grammel, L., Tory, M., & Storey, M. A. (2010). How information visualization novices construct visualizations. *IEEE transactions on visualization and computer graphics*, *16*(6), 943-952.

Hadamard, J. (1945). *The Psychology of Invention in the Mathematical Field*. Princeton University Press, Princeton, NJ.

Hogan, T., Hinrichs, U., Huron, S., Alexander, J., & Jansen, Y. (2021). Data Physicalization—Part II. *IEEE Computer Graphics and Applications*, *41*(01), 63-64.

Huron, S., Jansen, Y., & Carpendale, S. (2014). Constructing visual representations: Investigating the use of tangible tokens. *IEEE transactions on visualization and computer graphics*, *20*(12), 2102-2111.

Jansen, Y., Dragicevic, P., Isenberg, P., Alexander, J., Karnik, A., Kildal, J., ... & Hornbæk, K. (2015, April). Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3227-3236).

Javed, W., & Elmqvist, N. (2012, February). Exploring the design space of composite visualization. In *2012 ieee pacific visualization symposium* (pp. 1-8). IEEE.

Kim, N. W., Schweickart, E., Liu, Z., Dontcheva, M., Li, W., Popovic, J., & Pfister, H. (2016). Data-driven guides: Supporting expressive design for information graphics. *IEEE transactions on visualization and computer graphics*, *23*(1), 491-500.

Kim, N. W., Henry Riche, N., Bach, B., Xu, G., Brehmer, M., Hinckley, K., ... & Pfister, H. (2019, May). Datatoon: Drawing dynamic network comics with pen+ touch interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Lee, B., Kazi, R. H., & Smith, G. (2013). SketchStory: Telling more engaging stories with data through freeform sketching. *IEEE transactions on visualization and computer graphics*, *19*(12), 2416-2425.

Lee, B., Smith, G., Riche, N. H., Karlson, A., & Carpendale, S. (2015, April). SketchInsight: Natural data exploration on interactive whiteboards leveraging pen and touch interaction. In *2015 IEEE Pacific Visualization Symposium (PacificVis)* (pp. 199-206). IEEE.

Li, J., Brandt, J., Mech, R., Agrawala, M., & Jacobs, J. (2020, April). Supporting Visual Artists in Programming through Direct Inspection and Control of Program Execution. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Liu, Z., Thompson, J., Wilson, A., Dontcheva, M., Delorey, J., Grigg, S., ... & Stasko, J. (2018, April). Data Illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Lupi, G., & Posavec, S. (2016). *Dear data*. Chronicle books.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information, henry holt and co. *Inc., New York, NY*, *2*(4.2)

McCandless, D. (2012). *Information is beautiful* (pp. 978-0007294664). London: Collins.

Mei, H., Ma, Y., Wei, Y., & Chen, W. (2018). The design space of construction tools for information visualization: A survey. *Journal of Visual Languages & Computing*, *44*, 120-132.

Méndez, G. G., Nacenta, M. A., & Vandenheste, S. (2016, May). iVoLVER: Interactive visual language for visualization extraction and reconstruction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4073-4085).

Murray, D. G. (2013). *Tableau your data!: fast and easy visual analysis with tableau software*. John Wiley & Sons.

Neely, E. (2010). *Cy Twombly and the ethics of painting*. Columbia University.

Ou, L. C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, *29*(3), 232-240.

Pelto, M. (2016). *Recent climate change impacts on mountain glaciers*. John Wiley & Sons.

Price, R. H., & Bouffard, D. L. (1974). Behavioral appropriateness and situational constraint as dimensions of social behavior. *Journal of Personality and Social Psychology*, *30*(4), 579.

Ren, D., Lee, B., & Brehmer, M. (2018). Charticulator: Interactive construction of bespoke chart layouts. *IEEE transactions on visualization and computer graphics*, *25*(1), 789-799.

Ribarsky, W., & Foley, J. D. (1994). *Next-generation data visualization tools*. Georgia Institute of Technology.

Tang, T., Rubab, S., Lai, J., Cui, W., Yu, L., & Wu, Y. (2018). iStoryline: Effective convergence to hand-drawn storylines. *IEEE transactions on visualization and computer graphics*, *25*(1), 769-778.

Thudt, A., Hinrichs, U., Huron, S., & Carpendale, S. (2018, April). Self-reflection and personal physicalization construction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Tsandilas, T. (2020). StructGraphics: Flexible Visualization Design through Data-Agnostic and Reusable Graphical Structures. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), 315-325.

Tufte, E. R. (2006). *Beautiful evidence* (Vol. 1). Cheshire, CT: Graphics Press.

Walny, J., Huron, S., & Carpendale, S. (2015, June). An exploratory study of data sketching for visual representation. In *Computer Graphics Forum* (Vol. 34, No. 3, pp. 231-240).

Ware, C. (2012). Information Visualization: Perception for Design, Waltham, MA.

Wun, T., Oehlberg, L., Sturdee, M. and Carpendale, S., 2019, March. You say potato, i say po-data: Physical template tools for authoring visualizations. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 297-306).

Xia, H., Henry Riche, N., Chevalier, F., De Araujo, B., & Wigdor, D. (2018, April). DataInk: Direct and creative data-oriented drawing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

About the Authors:

Miriam Sturdee is a lecturer in the School of Computing & Communications at Lancaster University. She specialises in the application of creative practices to computer science, and is also the co-founder of the Sketching in HCI working group.

Søren Knudsen is an Assistant Professor at the Digital Design Department, IT University of Copenhagen. His research focuses on information visualization and HCI, and he is interested in supporting

people in understanding and discussing data in face-to-face and online contexts.

Sheelagh Carpendale is a Professor in the School of Computing Science at Simon Fraser University, where she holds the NSERC/AITF/SMART Technologies Industrial Research Chair in Interactive Technologies. Her research has been recognized through multiple awards including an IEEE Visualization Career Award.