DataSelfie: Empowering People to Design Personalized Visuals to Represent Their Data

Nam Wook Kim
Harvard University
namwkim@seas.harvard.edu

Hyejin Im
Tufts University
hyejin.im@tufts.edu

Nathalie Henry Riche
Microsoft Research
Nathalie.Henry@microsoft.com

Alicia Wang
Troy High School
aliwang2001@gmail.com

Krzysztof Gajos
Harvard University
kgajos@seas.harvard.edu

Hanspeter Pfister
Harvard University
pfister@seas.harvard.edu

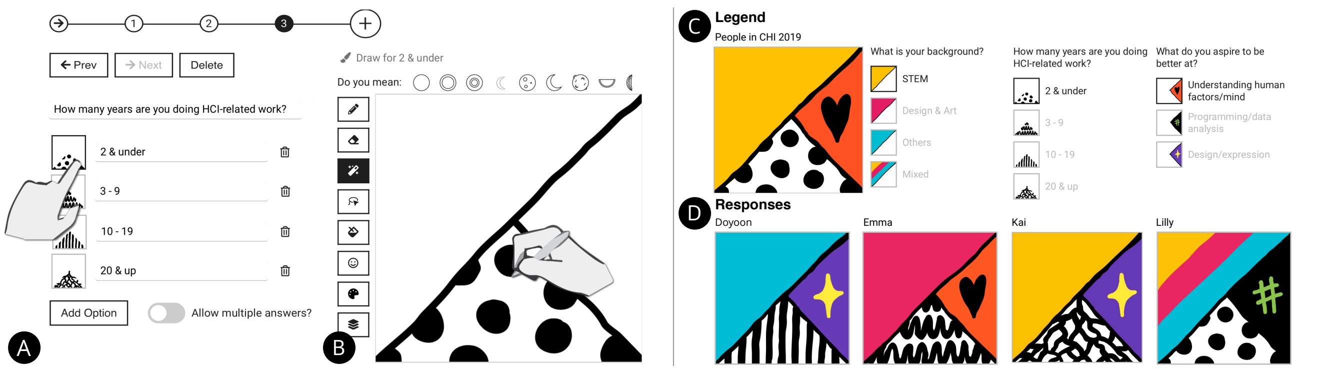

Figure 1: DataSelfie: A) The questionnaire editor for data collection. B) A visual mapping canvas in which a user can draw a unique personalized visual for a selected option. C) An interactive legend aids the interpretation of the visual mappings. D) Each questionnaire response generates a distinctive visual.

ABSTRACT

Many personal informatics systems allow people to collect and manage personal data and reflect more deeply about themselves. However, these tools rarely offer ways to customize how the data is visualized. In this work, we investigate the question of how to enable people to determine the representation of their data. We analyzed the Dear Data project to gain insights into the design elements of personal visualizations. We developed DataSelfie, a novel system that allows individuals to gather personal data and design custom visuals to represent the collected data. We conducted a user study to evaluate the usability of the system as well as its potential for individual and collaborative sensemaking of the data.

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300309

CCS CONCEPTS

• **Human-centered computing** → *Visualization systems and tools.*

KEYWORDS

Personal visualization, personal informatics, self-tracking, data selfies, data portraits, visualization, visual vocabulary.

ACM Reference Format:
Nam Wook Kim, Hyejin Im, Nathalie Henry Riche, Alicia Wang, Krzysztof Gajos, and Hanspeter Pfister. 2019. DataSelfie: Empowering People to Design Personalized Visuals to Represent Their Data. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300309

1 INTRODUCTION

Data can capture a snapshot of the world and allow us to understand ourselves and our communities better. Giorgia Lupi, a renowned visualization artist, recently advocated for *data humanism*, a more personal approach to collecting, analyzing and visualizing data [21]. In contrast to traditional data systems that focus on rapid processing, automated analysis, and summary presentation of a large quantity of data, data humanism puts more emphasis on perhaps slow but deeper engagements with data, as well as unique expressive visuals that embrace the context and subjectivity of the data they represent.

Lupi's manifesto echoes in a lot of existing research in the areas of personal informatics, self-tracking, and casual visualization [10], including an abundance of tools for collecting and visualizing personal data [7]. However, most of these tools focus on data collection, specifically automated tracking, and use predefined presentations of the data. Users remain mostly passive and less engaged with the data as such. On the other hand, there has been a surge of construction tools allowing people to create custom visualizations of data [13, 36]. Notwithstanding, these tools investigate authoring processes and tool expressiveness and have not been applied to a personal data context.

In this work, we are investigating the question of how to equip people with the ability to design their own visual vocabulary to represent qualitative and nuanced aspects of personal data. As an initial step toward understanding the empowerment through personal visualizations, we analyzed the Dear Data project [23] in which Giorgia Lupi and Stefanie Posavec collected, visualized, and shared the visual postcards of their data on a weekly basis for about a year. We performed open coding on the visual postcards as well as their reflections on the process of making them. We derived several design implications for personal informatics, including values of qualitative data for deeper reflection of self and common design elements used to personalize the visuals of the data.

Informed by the analysis of the Dear Data project, we developed DataSelfie, a web-based interactive system designed to enable any individuals to collect their data and decide how to visualize the data. DataSelfie combines a familiar survey authoring interface with a drawing capability so that users can create any questionnaire to ask questions about themselves and design a personalized visual vocabulary to represent the collected data. Unlike existing personal informatics tools, DataSelfie is offering users full autonomy over the visual presentation of their data. It also makes it possible to design the visual vocabulary at the time of creating the questionnaire; the final visual output is determined later on based on the response to the questionnaire. In this way, the construction of visualizations is not an afterthought, but a primary activity that users actively engage in while thinking about the goals of their data collection.

Users can use DataSelfie in both individual and collaborative scenarios. They can collect data on a recurring basis for self-tracking purposes or share a questionnaire with others to capture individual identities in a group setting. In the latter case, the small multiple of individual responses can create a single collective visualization representing the group. The users can also share their visual responses with each other, facilitating communication among individuals through data alone. Through a user study with 14 participants, we found that DataSelfie provides an easy-to-use and enjoyable interface to gather and visualize data. The variety of examples they created also suggests that DataSelfie enables to create expressive visualizations of data.

2 RELATED WORK

Personal Informatics & Self-Monitoring

Over the past decade, research has been abundant in personal informatics [17, 18], also known as various similar terms such as lifelogging and quantified self. A wide range of tools has been proposed to assist with collecting and managing various kinds of personal information, such as habits, activities, and moods, to encourage self-reflection and to promote self-knowledge and behavior change. Cho et al. characterize the design space of personal informatics based on whether data collection is fully automatic, semi-automatic, or fully manual [5]. Manual approaches include analog methods using pen and paper such as bullet journals [4] or digital methods using spreadsheets or note-taking apps. The manual approaches have high data capture burdens, often hindering people from sustaining long-term practices [5].

Automated tracking technology attempts to address this issue by leveraging personal devices embedded with various sensors to remove the need for manual data inputs [5]. However, complete automation of data collection often eliminates additional opportunities for engagement and reflection with personal data [5]. Also, sensors have limitations in the types of data they can collect, mainly focusing on collecting quantitative information, and fails to support people's practical goals and emotional needs [1, 5]. Most semi-automated approaches seek to strike a balance between both ends of the spectrum [5, 14]. For instance, SleepTight aims to make manual tracking easier rather than automate it by leveraging lock screen and home screen widgets [6], while OmniTrack provides manual trackers combined with triggers and external services that enable automated logging [13].

Existing self-tracking tools focus on data collection and management, not necessarily on how to display the data. As a result, the design of the tools dictates the presentation of the data, showing data summaries in the form of standard charts or tables. In this work, we set out to explore how to allow users to decide the representation of their data by augmenting a familiar survey tool with a drawing capability.

Casual and Personal Visualization

Most conventional visualization systems are designed to support domain experts to perform analytic tasks. As visualization has become widespread among a general audience, we are beginning to see more diverse uses of visualization. Pousman et al. describe casual information visualization that often involves ambient and artistic representations of data [26]. This type of visualization targets a broader population rather than experts alone and also serves personal or even aesthetic purposes [34]. Like other traditional visualizations, it generates analytic, often implicit, insights but often focuses more on awareness and reflective insights [26].

In the field of personal informatics, visualization has gained growing popularity as it provides a means to quickly make sense of complex data without requiring advanced statistical literacy. Huang et al. provide an overview of various design dimensions of personal visualizations and personal visual analytics [10], including who the data is about (e.g., self, family, and community) and the degree of control over data collection.

Visualizations in personal informatics tools are often personalized by using unconventional encodings such as visual metaphors, pictograms, and abstract drawings. For instance, UbiFit Garden uses the metaphor of a garden that blooms based on the performance of a user's physical activities [7]. Paper bullet journalists employ personally meaningful representations to meet their practical and emotional needs in tracking different types of data [1]. Such subjective visualizations can convey a unique perspective in personal visualizations [33], encourage further exploration of data [35], and establish a better sense of identity [12], although perceptions and preferences regarding designs of personal visualizations may depend on individual personalities [30].

While the past research suggests potential benefits of having personal visualizations of the data, users still do not have full control over them in existing personal visualizations. In our work, we provide a flexible framework that assists people with designing and sharing expressive visual representations of their data.

Visualization Authoring Tools

Significant efforts have been recently made to enable non-experts to create data visualizations both in industry and academia. Tableau [31] provides a simple drag & drop interface for constructing a chart while automating visual encoding under the hood based on perceptual effectiveness. Recent tools such as Lyra [29], Charticulator [27], and Data Illustrator [19] provide more customization options, including fine-grained mappings from data to geometric properties (e.g., fill, stroke, opacity), although visual marks mostly remain standard shapes. Other tools such as Data-Driven

Figure 2: An example postcard with the theme *distractions*: Giorgia Lupi (left) and Stefanie Posavec (right).

Guides [13] and DataInk [36] allow for more expressive controls over the design of visual marks by offering the ability to draw freeform sketches constrained by data.

While individuals can use these tools to create visualizations for their personal data, they have not been considered in a personal context. That is, visualization construction is considered as an afterthought not being part of data collection. DataSelfie can allow for prescriptive, rather than descriptive, visualization design tightly coupled with a data collection plan, generating a final visual in real time based on data input from a user.

3 ANALYZING DEAR DATA

To understand what types of data can capture personal lives and what kinds of visuals can be used to represent the data, we analyzed the Dear Data project [23] in which Lupi and Posavec collected data about themselves and drew custom visualizations with the data every week for a year.

Data

The project has a total of 104 postcards that spanned 52 weeks; each designer creates each half of them. The postcards contain hand-drawn visualizations of data about their lives along with a visual legend explaining the visual encoding and context of the data (Figure 2). The project also accompanies each designer's weekly retrospective, reflecting on the experience of creating and sharing the postcards, which is available on the website [1].

Method

We used qualitative methods of open and axial coding to analyze both postcards and retrospective texts. While visual postcards revealed common design elements to represent personal data, retrospective texts allowed us to gain insights into the underlying rationale for the design choices.

Results on Visual Postcards

Figure 3 shows the overview of extracted categories and observation counts out of 104 visualizations. Among three major themes identified, *things around me* (52) and *what i*

[1]http://www.dear-data.com/by-week/

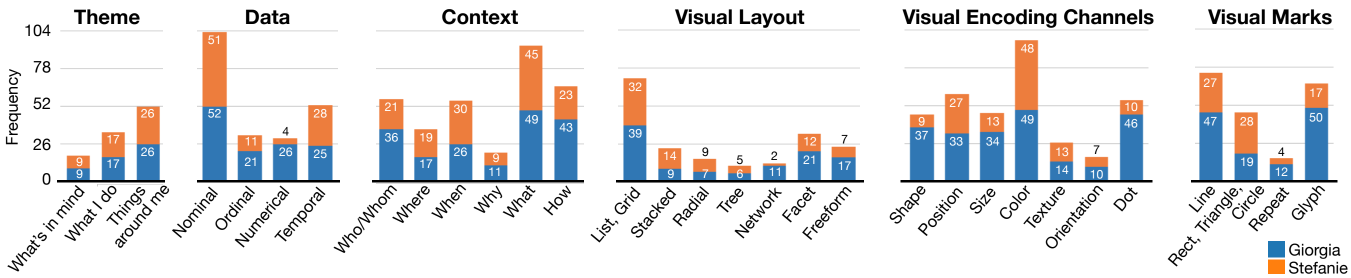

Figure 3: A histogram of categories identified through the open coding of 104 visual postcards.

do (34)—external states— were more common than *what's in mind"* (18)—internal states. Regarding data scales, *nominal* (103) and *ordinal* (32) were used most often in comparison to *numeric* (30). In fact, numerical data was also often binned into ordinal categories (8 out of 30).

Regarding the types of contextual questions addressed, we found *what* (94, e.g., what I did? what for?) was the most frequently used. *how* (67, e.g., how long or how intense was it?), *who/whom* (56, e.g., who were involved?), *when* (56), and *where* (36) were next in the order of frequency, while *why* (20) was the least frequently observed. About half of the time (54), they used *visual annotations* (e.g., circling or giving a unique visual) to highlight specific items (e.g., husband or boyfriend) and to communicate further context (e.g., missing or uncertain data).

They regularly used standard layouts such as *grid* (50) and *list* (22). For example, when showing data over time, it was common to arrange items in chronological order or distribute them in a 2-dimensional space in which both axes represent time and day respectively. We also observed moderate uses of other layouts such as *facets* (33), *stacked* (23), *networks* (13), and *trees* (11). They also often used *freeform* layouts (24), such as randomly distributing items in space. In most cases, they used more than one layout in one postcard.

Most frequently used visual variables include *color* (97), *position* (60), *size* (47), and *shape* (46). Size encoding was not necessarily used for numerical data but also for categorical data as well. They also often used *texture* (27), *fill/no fill* (12), and *orientation* (17), which are not commonly observed in digital visualizations. Most often used visual marks include *lines* (74), *glyph* (67), and *dots* (56), while *circles* (32) and *rectangles* (14) were also common. We also observed *repeat* marks (16, multiple same elements are stacked to encode a single value) and *real objects* (3, e.g., cosmetics).

There was also a notable difference between Lupi and Posavec (Figure 3). For instance, Lupi used more numeric scales compared to Posavec (26 vs. 4). As a result, Lupi used more size encoding as well (34 vs. 13). Lupi's visuals were fine-grained while Posavec's visuals were vibrant but straightforward. For instance, Lupi also frequently used small visual attributes (e.g., *dots*: 46 vs. 10, and *glyphs*: 50 vs. 17) attached to a primary mark (see Figure 2).

Results on Retrospective Texts

We identified five different stages of the process. We summarize what each stage entails below. We use 'P' for Posavec, 'L' for Lupi, and 'W' for both along with a week number to indicate the source of evidence.

Preparation. Lupi and Posavec chose a theme in which they are both interested. They often decided the theme for a performative purpose as a way to promote certain behavior (e.g., *Being nicer* - W23). They interpreted the theme in their personal contexts, such as thinking about how they perceive it or reflecting on the past experience they had. The choice of the theme and the interpretation had an impact on what kinds of data to collect and tools to use. They often set up specific questions and categories for data tracking prior to data collection, while in other cases they jot down logs about the topic and sort through the logs afterwards.

Data collection. They mostly took qualitative methods for gathering data, employing both analog and digital tools (e.g., notebooks, Reporter, Moves). The methods were largely divided into two kinds: 1) surveying data at once and 2) tracking data over time. A side effect of this qualitative approach was that they were aware of gathering data, which often had an impact on their moods and behaviors throughout the week (e.g., recording positive feelings creates an optimistic mood - L31). This often made data collection *performative*, intentionally manipulating certain data points.

Struggles in the manual process involved capturing transient things and noting down every instance of a data point (e.g., having to pause and note a laugh in a circle of people - P42). In particular, the latter resulted in what they call *data-gathering fatigue*. They embraced this imperfect nature of the manual process; for instance, *data voids* indicating a special moment such as a wedding or being drunken. Similarly, they marked with a special mark to denote playful data manipulation by others (e.g., *Physical contacts* - W6).

While gathering data, they reflected on their characters, habits, and preferences. For instance, they reflected on how they organize things (P16), where they have come from (P22), and what makes them happy (L31).

Data processing. Once data is collected, they would organize the data to find interesting facts and stories that they found meaningful and want to share. They often added additional data to provide context (e.g., adding demographic information to friends - L25) or simplified the complexity of the data to ease the drawing or understanding of it (e.g., highlighting the top 5 emotions-P11). This stage also served as a firewall to ensure privacy by hiding certain information in the data (e.g., crossing out the husband's name - P14), which they enjoyed indicating it in the postcards. While aggregating and categorizing data, they discovered new aspects of themselves and had opportunities for self-reflection. The final organization had an influence on mostly visual layouts and groupings in postcards.

Visual encoding. They sketched and iterated on multiple ideas before reaching the final drawing of the data. They mostly used colored pencils, as well as pens & markers, but often incorporated new materials and drawing techniques (e.g., cut and collaged papers - W20 and lipsticks - L19). Most frustrations in this stage came from the hand-drawing of the data, including difficulties of redrawing mistakes, maintaining accuracy, and mapping too many data points.

A variety of factors inspired the visual representation of the data. These include: 1) visual metaphors for data or theme (e.g., scribbles alluding the textured sonic waveforms - P32), 2) personal visual metaphors for data or theme (e.g., musical scores for complaints - L07), 3) personal styles and preferences in general (e.g., plant-based shapes - P14 or abstract arts - L14), and 4) random and spontaneous drawings rather than dictated by data.

There was a tension between readability and aesthetics; e.g., they were not pleased with the legibility of data insights when they liked drawings, and vice versa. Often, aesthetic focus was intentional, such as to detract attention from data (P27) or to compensate for lack of patterns in the data (L14). In general, they appreciated having control over aesthetics by noting that both beauty and functionality are important and which one to emphasize depends on data and context.

Sharing & reflection. By sharing visual postcards, they learned about each other and improved self-knowledge via comparisons. They appreciated different ways of interpreting a theme (e.g., the definition of a nice act - W23), collecting data (e.g., close vs broad range of friends - W25), and drawing postcards (e.g., circular vs linear layouts - W06).

The postcards brought back vivid memories of each week, providing ample rooms for reflection. They also reflected on the impact of the Dear Data project on their lives, including changes in future behavior (e.g., "After this week, I really figured I could be nicer and credit people more often..." - L15) and the value of the postcards as personal records.

Implications for Personal Informatics

Below, we summarize key insights we learned from the analysis results. Please note that the project involved two professional designers and thus the results may not generalize to a general audience.

I1. Capture qualitative aspects of self. Our findings suggest that qualitative data may enable more nuanced and richer reflection that is not possible with current automated tracking. A challenge is that the data is often ephemeral or intrusive to gather.

I2. Reveal missing and uncertain data. We observed that inaccurate and missing data is embraced in a personal context. The mistakes and failures are part of the data and can provide additional insights into their lives.

I3. Provide different modes of data collection. Data gathering can be either tracking data over time or surveying it once. In the former case, it may start with a specific collection plan or involve iteratively refining data or discovering it during post-processing.

I4. Support data exploration for story harvesting. Visualization for communication and sharing would require more than listing all data points. Filtering & simplifying data and classifying it into categories are ways to experiment with different personal stories before the final visualization.

I5. Use visual annotations to encode moments. Annotation and highlighting are useful tools for indicating specific data and memorable moments in a personal visualization. They can be also used to add humor or a personal touch to the visualization.

I6. Enable designing personalized visuals. Designing a unique visual of personal data may increase enjoyment and attachment toward the data. It is also a means for self-expression, communicating a perspective on the data and revealing one's personality through the visual.

I7. Choose visual variables for personalization. We found that *color* and *shape* variables may be more versatile for customization, while quantitative variables such as *position* and *size* are mostly governed by data. Traditionally underused variables (e.g., *texture* and *orientation*) due to lack of perceptual effectiveness might be acceptable in a personal context.

I8. Leverage visuals as personal documentaries. Our findings also point to the forgotten value of visualization; i.e., records. While communication and analysis are the focus of current visualization research, visual postcards demonstrate the use of visualizations to record personal memories.

I9. Address challenges with qualitative processes. We observed that main struggles lie in data collection and visual encoding stages. Our findings mostly confirmed previous research [2, 17], including forgetting to gather data and recovering from mistakes. A notable difference is the difficulty of tracking transient things like emotions or smiles.

I10. Support conversations through data. A significant value of sharing is that it opens up opportunities to see one's data in the context of the other's data. Comparing and contrasting differences in data as well as in visuals can provide additional channel for self-reflection. Also, others might help to identify different data insights that are not intended by the author.

4 DESIGN DECISIONS

Informed by the analysis of the Dear Data project, we wanted to build a system for individuals to engage with their personal data via creating expressive visual representations.

Our design goal is provide structured support for novices to crisply articulate the data they want to collect and visualize, as opposed to the Dear Data project that required substantial expertise in data collection and visualization.

We also draw inspirations from the work by Lupi and the Accurat team on automatically generating data portraits of individuals based on their responses to a survey [20, 22]. We aim to generalize this idea to a general audience along with the insights learned from the Dear Data project.

D1. Using a questionnaire to collect qualitative data. (I1, I3, I9) To enable flexible collection of personal data, we seek to leverage the familiar experience of creating a questionnaire in a survey form. The questionnaire can serve as a data collection plan as observed in the preparation stage. It should be also editable to allow for revisions. The questionnaire is versatile to support contextual questions (e.g., who, what, why etc) and can capture diverse aspects of daily lives.

D2. Designing a personalized visual vocabulary. (I6, I7, I9) To allow people to craft unique and expressive visual representation of their data, we propose to augment the questionnaire editor with a drawing canvas in which a user can draw visual mappings for collected data. The visual mappings create a personalized visual vocabulary and should be also editable to allow for iterative design. The mappings can be constructed after data collection or at the time of creating the questionnaire; the latter has a benefit of visualizing the goals of data collection in advance.

D3. Support sharing visuals of data: (I8, I10) To foster collaborative reflection, we intend to support sharing data visuals with others. To assist in interpreting visual mappings, a legend should be generated and accompanied with the visuals. Also, sharing with more than two people should be possible.

5 THE DATASELFIE INTERFACE DESIGN

DataSelfie consists of three main components: a questionnaire editor (Figure 1A) and a canvas for drawing visual mappings (Figure 1B), an interactive legend (Figure 1C), and a response viewer (Figure 1D).

Creating a Questionnaire

The questionnaire editor has the same interface as other survey tools like Google Forms. It currently supports a multiple choice question (radio buttons) with an option to allow more than one answer (checkboxes). This type of question enables collection of mostly qualitative data (I1), i.e., categorical data, and also renders a rapid response by quickly selecting one. Although we do not support numerical data, users can still capture quantitative data by binning it into intervals (e.g., less than 10, 10 to 100, more than 100) as observed in the Dear Data project. While out of scope for this paper, the questionnaire editor can be extended to incorporate automated tracking similar to OmniTrack [14].

Collecting Data. To collect a data point, a user simply needs to respond to the questionnaire, which is similar to filling out a survey. The user can respond multiple times, generating more than one data point. To assist data collection on a recurring basis (e.g., daily, weekly, or monthly), DataSelfie supports setting a reminder email to prompt for a new response (I9). DataSelfie also allows the user to refine the questionnaire during data collection (I3).

Drawing Visual Mappings

The questionnaire editor embeds a drawing canvas that affords free-form sketching so that a user can draw personalized visuals to represent each answer option per each question (I6, I7-shape). To construct a mapping from a question to a visual, the user can simply taps an option thumbnail and draw a corresponding visual on the canvas, and keep doing this for other options (Figure 1B).

In this way, the user can create a *personalized visual vocabulary* integrated with the questionnaire. That is, the visual vocabulary is a set of drawings tied to contextual questions for data collection, generating a unique visual per each response to the questions. DataSelfie allows iterative refinements of the visual mappings at any time (I9).

Drawing Tools. The drawing canvas comes with a set of tools including a pencil for sketching, selection tool, color fill, and palette (color, stroke, opacity). The user can use the layer view (Figure 4C) to see how the drawings for the currently selected option would interact with those for other options and questions. It is necessary since the same canvas area is shared across the questionnaire, stacking all drawing layers into a single layer to generate a final image when a user submits a response.

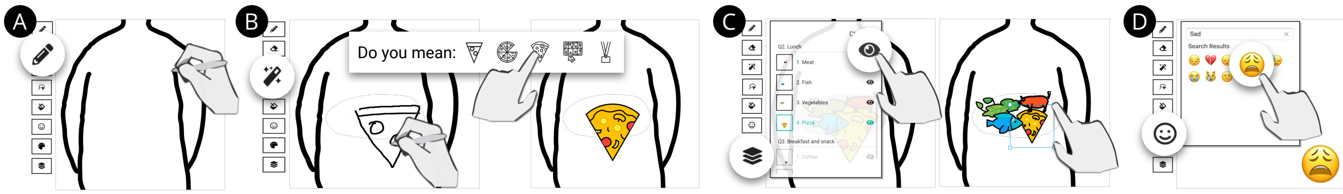

Figure 4: The drawing interface of DataSelfie: A) Drawing canvas, B) Auto-drawing, C) Layer view and selection, D) Emoji tool.

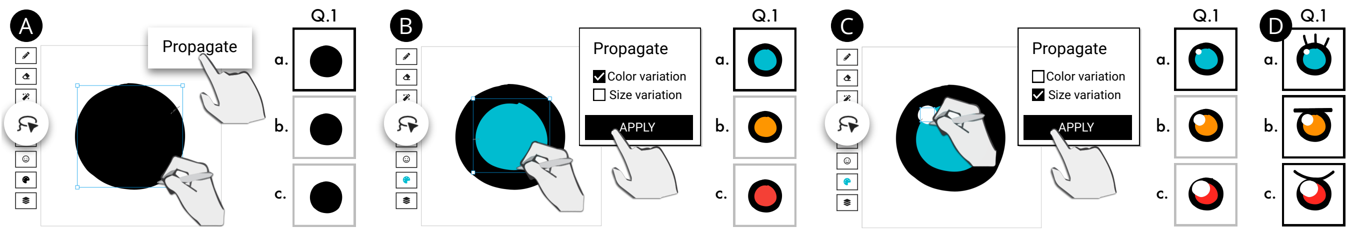

Figure 5: Automatic encoding support: A) Reusing a shape, B) Color encoding, C) Size encoding, D) Customization.

Assistive Drawing. To assist users with less experience in drawing, especially with a mouse, we integrated an auto-drawing feature [15] that suggests a predefined icon based on a series of strokes, dragging & dropping images, and a search box for adding emoji icons (Figure 4D). To enable reusing of the same shape across all options, we also support duplicating a shape across all options in a question, along with automatic size and color encodings (I7-color, Figure 5). The duplicate shapes compose a group constrained by the applied encodings; i.e., if a user changes the size of a shared shape, it also updates other shapes in the same group (I9).

Generating Visuals. Figure 6 shows a schematic diagram of how DataSelfie generates each visual response. Once a user submits a response to the questionnaire, it maps each answer option to a corresponding drawing while other drawings for other options are hidden in the final outcome. In this way, the visual vocabulary can generate a combinatorial number of visuals.

We currently take a layered approach to model the underlying visual vocabulary, which is similar to the layer model in a conventional drawing tool like Adobe Illustrator. Background and question layers are merged into a single layer, overlaid on top of each other in the order of questions. Each question layer can toggle the visibility of one or more option layers depending on the response.

Parametric Visualization Construction

Although simple, it is worth noting the way of generating a visualization is prescriptive than descriptive. The user create a visual vocabulary that can be used to generate a final visualization without actual data, rather than attempting to describe existing data. This is clearly different from typical visualization authoring tools as the users of those tools assume the existence of the data. In our case, the generation process is more close to parametric design. Each question can be considered as a parameter whose value is later specified by the user.

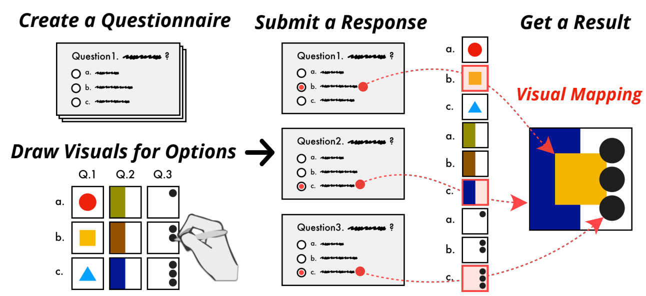

Figure 6: The production rule of a visualization in DataSelfie, generating a unique visual for each questionnaire response.

Sharing & collaboration with others

To facilitate a casual conversation through data, a user can share the questionnaire with others (I10). Each of them can then submit a response and receive a unique visual, contributing to a collage of visual responses (Figure 1D) [20]. DataSelfie also supports sharing each visual response with others using the hyperlink generated after submitting a response (I8). To assist in interpreting the data selfies, DataSelfie automatically generates an interactive visual legend from the visual vocabulary (Figure 1C).

6 USAGE SCENARIO

We introduce individual and collaborative scenarios that correspond to two kinds of data collection scenarios we observed in the Dear Data project: tracking data over time and surveying data once.

Scenario 1: Self-Tracking

This scenario introduces how an individual user might use DataSelfie for self-tracking or lifelogging.

Sam is recently concerned that he eats pizza a lot. He would like to have a healthy diet but struggles to do so due to a busy school life, and decided to track what kinds of food he eats every day. He starts by creating a questionnaire with three simple questions asking what types of food he had during the morning, afternoon, and evening (Figure 4).

Once finishing the questionnaire, he begins drawing to visualize his personal goal of the questionnaire. He wants the visuals to reflect his diet directly. He first draws a background representing his body (Figure 4A). For each option, he draws a visual representing the type of food he may or may not want to eat. He thinks he is not good at drawing, so uses the auto-drawing feature to draw the pizza (Figure 4B) and adds an emoji to represent his feelings after the diet.

He set a daily reminder to fill out the questionnaire. When he received a reminder through an email, he taps the link using his mobile phone to go to the questionnaire to submit a response. He often intentionally eats healthy food in order to get the visual he wants, making the data collection performative. At the end of the month, he reviews all the responses in a collective form. He gets a sense of how he maintained his goal by quickly gauging the prominence of all the greens.

Scenario 2: Community Gathering

The second scenario demonstrates a collaborative activity, similar to the one demonstrated by Giorgia Lupi in which she used a questionnaire to survey fun facts about conference attendees [22].

Emily is an organizer of a HCI workshop. She wants to engage the attendees in a way that they get to know each other as they belong to the same research community. She thinks about the topic of the workshop and comes up with a set of questions that may be well suited to represent members of the group (Figure 1C). She draws a background that serves as a template for all questions like a coloring book. The drawings for each question fills a different region in the background (See divide lines in Figure 1B).

Since this is not about tracking data over time, she turned off reminders. Instead, she shares the web link to the survey with the attendees. Each attendee gets a unique visual that captures their identity based on their response (Figure 1). Emily prints this visual and make a personalized badge for each attendee. She also creates a collage of all responses showing the community collectively (Figure 1D).

7 IMPLEMENTATION

DataSelfie is a web-based application written in Javascript. It uses React.js [8] for building user interface components and Redux.js [9] for application state management. We use a Flask [28] web server and MongoDB [25] to persist the user and questionnaire data and enable sharing through hyperlinks. We heavily use Paper.js [16] to implement the drawing canvas integrated into the questionnaire editor. For assistive drawing, we leverage Autodraw [15] and Emoji Mart [24]. DataSelfie will be open-sourced and available online upon acceptance.

8 USER STUDY

We conducted a qualitative user study to evaluate the usability of DataSelfie and gain insights into its usefulness with an emphasis on lifelogging.

Participants

We had a screening survey to recruit 14 participants (nine female and five male, six aged 18-24 and seven aged 26-35, one aged 36-45) who have prior experience on manual journaling and automatic self-tracking, as well as basic tool literacy on survey tools and presentation software. We used Safari browser on iPad 12.9 inch using Apple Pencil. We paid participants with a $25 gift card for an hour-long session.

Procedure & Tasks

Each hour-long study session started with a background survey at the beginning. We had a short tutorial and four tasks in total. For the tutorial, we used an example adapted from an existing template[2].

In the first task, we asked participants to reproduce the example from the tutorial in order to make them familiar with DataSelfie. In the second task, they were asked to replicate a new example as quickly as possible without any guidance from a researcher. We varied the examples in the two tasks so that users can see two use case scenarios: 1) surveying a single state or current identity and 2) tracking recurring states over time. In the third task, they needed to design a new questionnaire with a minimum of three questions and craft the corresponding visual vocabulary. We assigned each participant to a specific scenario (A. *tracking* or B. *surveying*) and topic (1. *what's in mind*, 2. *what I do*, or 3. *things around me*).

We concluded the session with a usability survey adapted from the System Usability Scale [3] using a 5-point Likert

[2]https://ideas.ted.com/how-to-draw-your-own-selfie-using-your-personal-data/

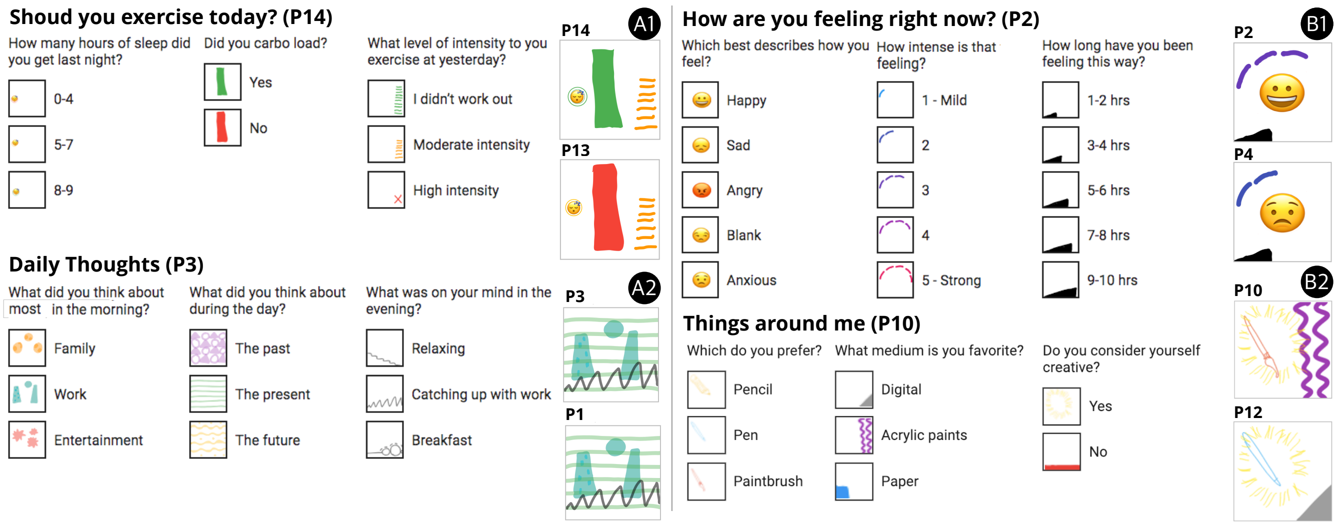

Figure 7: Four examples of questionnaires created by participants during the third task for two usage scenarios: A: tracking recurring states over time and B: capturing a current state.

scale (1—strongly disagree, 5—strongly agree), and a semi-structured interview discussing the overall experience and potential benefits in their current practice.

The fourth task happened later online, moderated by a researcher. We paired two participants who were assigned to the same scenario and topic in the third task. They submitted responses to each other's surveys and shared the final responses. We asked them to reflect on how they perceive the experience of sharing and exchanging the visual responses.

Results

All participants completed all the tasks. For the second task, it took 4 minutes on average (min=3, max=6), while the third task took 13 minutes on average (min=9, max=17). In the usability survey, participants rated higher on the ease of use (M=4.43, SD=0.76), the learnability (M=4.36, SD=0.74), and the usefulness of the tool (M=4.29, SD=0.83).

Diversity of questionnaires and visuals. They generated a variety of questionnaires (Figure 7, see the supplement for the full collection). The question topics include exercise, feelings, music, sleep, meeting, etc. A majority of questions were *how* and *what* questions, while some questions did not belong to 5W1H (e.g., Q: I like to take a shower. A: morning, afternoon, or before bed). In terms of visuals, all participants used *color* and *shape* variables while some also used the *size* variable using *repeat* (or *density*) marks (Figure 7B1). They also attempt to divide the canvas region to layout visual elements based on questions (Figure 7A1).

Ease of use & learnability. Participants had positive experiences with the tool with a few saying *"the interface is very intuitive"* - P1 and *"It is well-designed and fun to use"* - P5. One issue we found was foreseeing how layers will be merged, often ended up having next questions blocking previous ones. However, one participant deliberately manipulated this by having transparent elements to see through previous layers. Participants liked the assistive drawing support. One participant kept using the auto-drawing and said "I know it is not perfect, but I just like it" - P6, while another participant said *"I like having the emoji interface in a drawing tool"* - P8. One participant suggested to have drawing templates to further alleviate the fear of a blank canvas.

Enjoyment & engagement. Participants made positive comments on the engaging nature of the tool. For instance, they mentioned that *"I like how it integrates the answers into a succinct and amusing visual form"* - P5 and *"It reminds me of an art class. It would be useful for therapeutic purpose"* - P7. Other participants commented that *"I enjoyed the roughness of the hand-drawn sketches and I think the raw emotion feeling would be lost if the graphics were very clean cut and digitized."* - P4, and *"Freedom to draw anything to represent my data makes it fun and lively"* - P13.

Trade-offs in manual data collection. They liked the idea of collecting qualitative data, saying *"It's a good way to gain insights that can't be traced automatically"* - P11, and *"when it comes to qualitative data, it is more personal"* - P13. These comments are in line with the Dear Data project. However, a few participants said they would not use it for self-tracking due to lack of time for creating a questionnaire (e.g., *"I wouldn't use this for self-tracking. It takes too much time"* - P7).

Benefits of drawing personalized visuals. Participants' reaction to the drawing capability was mostly positive, saying *"it allows you to visualize goals and impacts in advance"*

- P9, *"I like visuals as they invoke more thoughts"* - P7, and *"it would be good for emotion and well-being that lack clear forms"* - P8. Several participants also commented that *"it's more interactive and personal than existing (automatic) tracking tools"* - P13 and *"I think I would use it to track my menstrual cycles"* - P10. Others had mixed opinions: *"it's good for qualitative data, but probably not for quantitative data in which precision is important"* - P8, *"I would still like to see aggregate summaries"* - P12.

Benefits of sharing. They also made positive comments on sharing (P3, P11, P12), although it is too short of time to judge its true benefit. One participant mentioned that *"if there is one reason for me to use this tool, it would be the sharing aspect"* - P11; this participant expressed reluctance to use the tool because of potential privacy issues. Other participants commented that: *"surprised since she had the same response as me! I think the visual result helps to immediately tell how similar or different our responses were"* - P3 (Figure 7A2) and *"It was also very relatable to see how another (stranger) is feeling about the same subject"* - P4.

Other use cases. Several participants unexpectedly came up with different use cases that we did not intend to support, saying *"I can design a visual to lead to make a decision. It can make a real change on who you are in each date."* - P14 (See Figure 7A1). Other participants also mentioned, *"I think I can use this tool for testing different configurations for a garden I am designing."*- P1 and *"My psychologist friend would love to use this for her research tracking moods from subjects."* - P2.

9 DISCUSSION

Limitations and Opportunities

The user study surfaced several opportunities for further improvements. We did not support question types for numerical data, as it can be binned to categories. However, we observed that participants were often annoyed as they need to write out all the intervals manually (e.g., Figure 7B1). It may be desirable to automatically generate these intervals while the user provides the min, max, and step size of the data. The main struggle, although not significant, was to come up with questions in the third task, particularly regarding the first topic: *what's in mind*. This topic was the least frequent in the Dear Data project as well. Providing questionnaire templates or allowing them to share with others would alleviate the struggle. Likewise, customizable drawing templates would also mitigate the perception of efforts necessary, as well as the concern about the lack of drawing skills.

Empowerment through Personalized Visuals

Our user study illuminated potential benefits of having control over the representation of personal data. One thing that recurrently stands out was the joyfulness of creating personalized visuals. The randomness of the outcome also seemed to contribute to it, as we saw some participants uttered their surprise when they see a visual response. This observation is also in line with previous research on the poetic use of unconventional encodings to create an element of surprise and stimulate reflection [33]. Another suggested benefit was being able to visualize the goals of data collection, as well as the impacts of the data on their lives. In particular, P14 articulated the use of our tool for visual decision making that can have an immediate impact on behavior change. Our study was qualitative and we believe it requires controlled experiments or a live deployment study to investigate the true potential of empowering people through personalized visuals more formally. The findings from our study provide initial hypotheses.

Beyond visuals and simple mappings

We focused on creating visuals in this work, but designing a multi-sensory experience using personal data is an exciting future direction. Similarly, one participant in the user study mentioned that sounds could often be much more emotional than visuals. It would be interesting to explore implications of using sounds [37] as well as tangible materials [11, 32] in a personal context.

In addition, we believe there is an interesting avenue for exploring the concept of the parametric, decision-oriented visualization generation. Our current framework for visualization construction was based on simple linear mappings from an option to a visual in a single canvas. But there are fruitful opportunities to incorporate more advanced composition rules to the framework. For example, it can incorporate branches like a decision tree and allow for laying out the layers in multiple canvases that are further parameterized by the position and size.

10 CONCLUSION & FUTURE WORK

In this work, we analyzed the Dear Data project to gain insights on what is like to design custom visuals to represent personal data. We developed DataSelfie as a step toward empowering individuals to create personalized visuals to depict their data. We conducted a user study to understand the potential benefit of the empowerment. For future work, we plan to conduct a long-term deployment study to study its advantages and disadvantages in more depth for a self-tracking scenario.

11 ACKNOWLEDGMENTS

We would like to thank Giorgia Lupi and anonymous reviewers for useful feedback. Nam Wook Kim would like to acknowledge the support from the Kwanjeong Educational Foundation and the Siebel Scholars Foundation.

REFERENCES

[1] Amid Ayobi, Tobias Sonne, Paul Marshall, and Anna L. Cox. 2018. Flexible and Mindful Self-Tracking: Design Implications from Paper Bullet Journals. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 28, 14 pages. https://doi.org/10.1145/3173574.3173602

[2] Alex Bigelow, Steven Drucker, Danyel Fisher, and Miriah Meyer. 2014. Reflections on How Designers Design with Data. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 17–24. https://doi.org/10.1145/2598153.2598175

[3] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[4] Ryder Carroll. 2018. Bullet Journal. https://bulletjournal.com/. Accessed: 2018-06-16.

[5] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A. Epstein, Felicia Cordeiro, Matthew Kay, Gregory D. Abowd, Tanzeem Choudhury, James Fogarty, Bongshin Lee, Mark Matthews, and Julie A. Kientz. 2017. Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. *IEEE Pervasive Computing* 16, 1 (Jan. 2017), 74–84. https://doi.org/10.1109/MPRV.2017.18

[6] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A. Kientz. 2015. SleepTight: Low-burden, Self-monitoring Technology for Capturing and Reflecting on Sleep Behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 121–132. https://doi.org/10.1145/2750858.2804266

[7] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1797–1806. https://doi.org/10.1145/1357054.1357335

[8] Facebook. 2018. React: A JavaScript Library for Building User Interfaces. https://reactjs.org/. [Online; accessed 16-Sep-2018].

[9] Facebook. 2018. Redux: a Predictable State Container for JavaScript Apps. https://redux.js.org/. [Online; accessed 16-Sep-2018].

[10] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2015. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3 (March 2015), 420–433. https://doi.org/10.1109/TVCG.2014.2359887

[11] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3227–3236. https://doi.org/10.1145/2702123.2702180

[12] Jin Kang, Jomara Binda, Pratik Agarwal, Bruno Saconi, and Eun Kyoung Choe. 2017. Fostering User Engagement: Improving Sense of Identity Through Cosmetic Customization in Wearable Trackers. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '17)*. ACM, New York, NY, USA, 11–20. https://doi.org/10.1145/3154862.3154878

[13] Nam Wook Kim, Eston Schweickart, Zhicheng Liu, Mira Dontcheva, Wilmot Li, Jovan Popovic, and Hanspeter Pfister. 2017. Data-Driven Guides: Supporting Expressive Design for Information Graphics. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 491–500. https://doi.org/10.1109/TVCG.2016.2598620

[14] Young-Ho Kim, Jae Ho Jeon, Bongshin Lee, Eun Kyoung Choe, and Jinwook Seo. 2017. OmniTrack: A Flexible Self-Tracking Approach Leveraging Semi-Automated Tracking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 67 (Sept. 2017), 28 pages. https://doi.org/10.1145/3130930

[15] Google Creative Lab. 2018. AutoDraw. https://www.autodraw.com/. [Online; accessed 16-Sep-2018].

[16] Jürg Lehni and Jonathan Puckey. 2018. Paper.js: The Swiss Army Knife of Vector Graphics Scripting. http://paperjs.org/. [Online; accessed 16-Sep-2018].

[17] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 557–566. https://doi.org/10.1145/1753326.1753409

[18] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 405–414. https://doi.org/10.1145/2030112.2030166

[19] Zhicheng Liu, John Thompson, Alan Wilson, Mira Dontcheva, James Delorey, Sam Grigg, Bernard Kerr, and John Stasko. 2018. 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 (CHI '18)*. ACM, New York, NY, USA, Article 123, 13 pages. https://doi.org/10.1145/3173574.3173697

[20] Giorgia Lupi. 2017. Collaborative Data Wallpaper for STORY. http://giorgialupi.com/collaborative-data-wallpaper-for-story/.

[21] Giorgia Lupi. 2017. Data Humanism, the Revolution will be Visualized. http://www.printmag.com/information-design/data-humanism-future-of-data-visualization/.

[22] Giorgia Lupi and Accurat. 2017. Data Portraits at TED 2017. https://www.accurat.it/works/ted/.

[23] Giorgia Lupi and Stefanie Posavec. 2016. *Dear Data.* Chronicle Books.

[24] Missive. 2018. Emoji Mart. https://missive.github.io/emoji-mart/. [Online; accessed 16-Sep-2018].

[25] Inc. MongoDB. 2018. MongoDB. https://www.mongodb.com/. [Online; accessed 16-Sep-2018].

[26] Zachary Pousman, John Stasko, and Michael Mateas. 2007. Casual Information Visualization: Depictions of Data in Everyday Life. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov. 2007), 1145–1152. https://doi.org/10.1109/TVCG.2007.70541

[27] Donghao Ren, Bongshin Lee, and Matthew Brehmer. 2018. Charticulator: Interactive construction of bespoke chart layouts. *IEEE Transactions on Visualization and Computer Graphics* (2018), 1–1. https://doi.org/10.1109/TVCG.2018.2865158

[28] Armin Ronacher. 2018. Flask: A Microframework for Python. http://flask.pocoo.org/. [Online; accessed 16-Sep-2018].

[29] Arvind Satyanarayan and Jeffrey Heer. 2014. Lyra: An Interactive Visualization Design Environment. *Computer Graphics Forum (Proceedings of EuroVis)* 33, 3 (2014), 351–360. http://doi.org/10.1111/cgf.12391.

[30] Hanna Schneider, Katrin Schauer, Clemens Stachl, and Andreas Butz. 2017. Your Data, Your Vis: Personalizing Personal Data Visualizations. In *IFIP Conference on Human-Computer Interaction*. Springer, 374–392.

[31] Tableau Software. 2018. Tableau. https://www.tableau.com. Accessed: 2018-09-15.

[32] Alice Thudt, Uta Hinrichs, Samuel Huron, and Sheelagh Carpendale. 2018. Self-Reflection and Personal Physicalization Construction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 154, 13 pages. https://doi.org/10.1145/3173574.3173728

[33] Alice Thudt, Charles Perin, Wesley Willett, and Sheelagh Carpendale. 2017. Subjectivity in personal storytelling with visualization. *Information Design Journal* 23, 1 (2017), 48–64. https://doi.org/10.175/idj.23.1.07thu

[34] Fernanda B. Viégas and Martin Wattenberg. 2007. Artistic Data Visualization: Beyond Visual Analytics. In *Proceedings of the 2Nd International Conference on Online Communities and Social Computing (OCSC'07)*. Springer-Verlag, Berlin, Heidelberg, 182–191. http://dl.acm.org/citation.cfm?id=1784297.1784319

[35] Shimin Wang, Yuzuru Tanahashi, Nick Leaf, and Kwan-Liu Ma. 2015. Design and Effects of Personal Visualizations. *IEEE Computer Graphics and Applications* 35, 4 (July 2015), 82–93. https://doi.org/10.1109/MCG.2015.74

[36] Haijun Xia, Nathalie Henry Riche, Fanny Chevalier, Bruno Araujo, and Daniel Wigdor. 2018. DataInk: Direct and creative data-oriented drawing. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. https://doi.org/10.1145/3173574.3173797.

[37] Haixia Zhao, Catherine Plaisant, and Ben Shneiderman. 2005. "I Hear the Pattern": Interactive Sonification of Geographical Data Patterns. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1905–1908. https://doi.org/10.1145/1056808.1057052