Data-Oriented Drawing

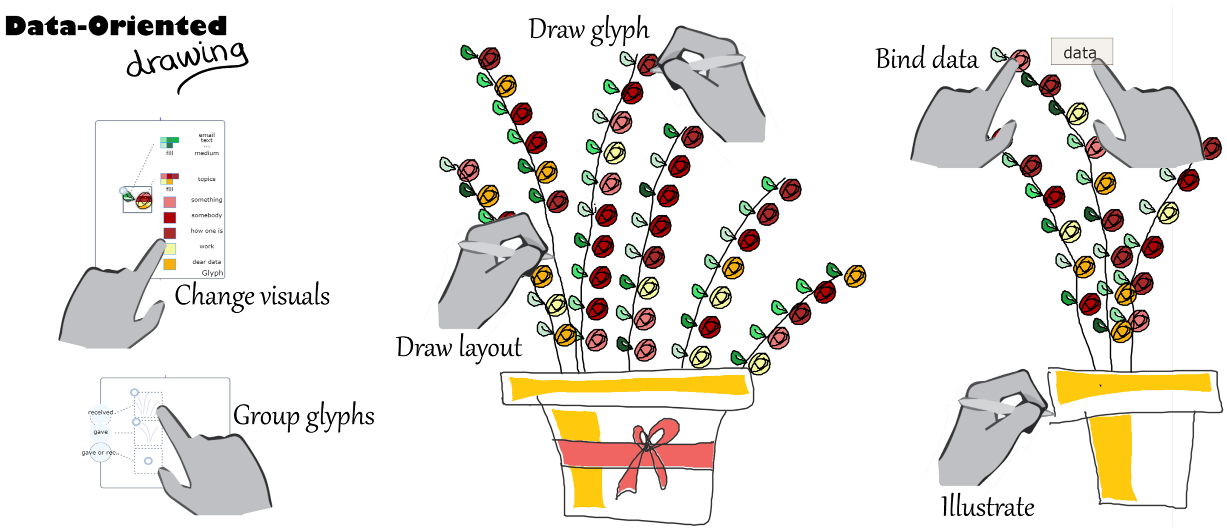

Figure 1. Data-Oriented Drawing (DOD) is an approach enabling for the easy creation of visually creative data visualizations. Visualization authors can interleave illustration (right, lower) with data binding (right, upper) to quickly create visualizations that can be iteratively specified through direct manipulation (left) and laid out using drawn paths (center, middle).

ABSTRACT

Creating whimsical, personal data visualizations remains a challenge due to a lack of tools that enable for creative visual expression while providing support to bind graphical content to data. Many data analysis and visualization creation tools target the quick generation of visual representations, but lack the functionality necessary for graphics design. Toolkits and charting libraries offer more expressive power, but require expert programming skills to achieve custom designs. In contrast, sketching affords fluid experimentation with visual shapes and layouts in a free-form manner, but requires one to manually draw every single data point. In this work, we aim to bridge the gap between these extremes. We propose *Data-Oriented Drawing*, an approach leveraging digital sketching and data manipulation to address the difficulties inherent in creating such visualizations, and present DataInk, a prototype implementing this approach. An evaluation with designers and non-experts demonstrated the expressive flexibility and power of our approach.

In submission to ACM CHI 2018, submission number 2368. Confidential.

Thank you for your consideration.

INTRODUCTION

Visual representations of data are a powerful communication medium for presenting insights and ideas in an understandable form. Beyond traditional information visualization, designers, artists and enthusiasts alike also leverage this mode of expression to craft meaningful, beautiful, and memorable pieces [10, 28, 32, 48, 51]. While compelling, authoring visualizations that embody creative, artistic influences is presently a laborious task, typically requiring one to alternate between different tools to form an enticing, yet factual representation of the data [4].

Tools that aim at supporting design creativity in information visualization should encompass *design expression*, i.e., the crafting of highly customized and stylized visuals and *rigorous execution* of the principles of information visualization to ensure that the visual properties of visuals comply with data they encode [5]. Our goal is to also encompass support for creativity [47], and thus empower authors with *creative exploration* through the fluid and spontaneous experimentation with glyphs and layouts, to facilitate ideation and iterative design.

While several tools exist to assist with the generation of data visualizations, there is a lack of software that supports these three tenants of design creativity. The tools used by visualization authors today typically enable for either rigorous visualization, or creative design expression, but none address our goal of facilitating data visualization as both a medium and a tool for creative expression.

Visualization toolkits (e.g. D^3 [8]), graphical interfaces (e.g. Lyra [45]) and online applications (e.g. RAWGraphs [15]) address the rigorous execution of visualizations by enabling authors to automatically apply visual encodings to their data. These tools offer different trade-offs between power and simplicity of use. Programming toolkits enable for the implementation of any custom data visualization, however often have a steep learning curve. Online applications are limited to the creation of a small set of predefined, mutable data visualizations. The lack of flexibility to customize visuals in a fluid manner makes these tools poorly suited for creative exploration and design expression.

In contrast, graphic design and illustration software such as Adobe Illustrator [1] enables rich visual expression, but lacks support to leverage data bindings to aid in drawing. Further, the lack of support for bindings hampers iterative design, since simple changes require significant work. The visualization research community has recently started to address this by bridging diverse types of interfaces [4] and augmenting illustration software with specific widgets [26]. These latter efforts, however, rely on a rigid, complex workflow that hampers creative exploration. This work thus set out to support design expression and rigorous execution while offering flexibility, plasticity and freedom to manipulate for creative exploration via iterative design.

The inherent freeform nature of sketching enables visualization authors to follow their inspiration and quickly model and remodel different designs, or discard them without second thoughts [6, 41]. Such unconstrained means of expression can also foster an author's expressivity and promote understanding when drawing with data [55]. We thus propose Data-Oriented Drawing (DOD), a novel approach that leverages drawing on a digital canvas and the use of direct manipulation to seamlessly access visual properties of graphics and bind them to data.

This paper articulates a set of five principles for DOD. We realized these principles in DataInk, a pen-and-touch enabled interface (see accompanying video and website http://datainkresearch.github.io). A user study with eight designers and non-experts suggests that DataInk affords rich expressivity via sketching, rapid data binding via direct manipulation, while limiting training.

RELATED WORK

Many authoring tools have been proposed to assist authors in creating visualizations for their data. Creating a visual representation of data requires one to specify a *visual-data mapping*: associating data dimensions to visual variables such as color, shape, size, or position [3, 11]. This encoding produces *marks* that visually represent the data, as a one-to-one correspondence between each data point and its visual representation (i.e. glyph-based representation) or as a groups of data points (e.g. bar charts). Grammel et al. [17] surveyed visualization and HCI publications, and reported on authoring strategies of over 60 visualization creation tools. Online tools also proliferate and spark practitioners' discussions on their merits and limitations [42]. We review visualization authoring tools with regard to the purposes they address, delineating the gap our work fills.

Rigorous Execution

Deeply rooted in scientific practice—where visualizations are seen as functional assets for analysis—the vast majority of tools developed by the research community support the systematic building of a visualization from data. This ensures that the visual properties of marks scrupulously comply to the data they encode (i.e., rigorous execution).

Visualization toolkits and charting libraries (e.g., D3.js [8], Vega [44], ggplot2 [57]), and advanced systems for data analysis (e.g., Microsoft Excel [36], Tableau [49], PowerBi [37], Lyra [45], iVisDesigner [39]) afford some level of design expression. However, creating custom graphics and layouts requires substantial training such as learning to program, or having good command of the vast functionality present in a full-featured graphical user interface. Trading power for simplicity, many online systems (e.g. ManyEyes [53], EasyCharts [16], RAWGraphs [15]) enable non-expert audiences to create visualizations of their data in just a few steps, i.e., load data, choose a template, and select the data dimensions to represent. The range of possibilities these applications support, however, is limited to commonplace visualizations. The integration of custom glyphs and layouts is usually limited or not supported.

In such tools, creative exploration is hindered due to the large number of steps required to generate, evolve, and refine design alternatives. Moreover, most design decisions have to be made before the user can see an actual visual representation of their data. Authors not versed in the art of crafting data visualizations may not fully comprehend what each decision entails or how the series of decisions they made will impact the resulting visual representation [35]. Méndez et al. proposed iVolver [34] to alleviate the need to making decisions beforehand, following the principle of constructive visualization [20]. While this approach affords more expressivity than other systems, it still requires one to learn the workflow language to achieve the envisioned designs, making it difficult to pursue creative exploration.

Design Expression

Beyond their functional purpose, visual representations of data have increasingly been viewed as a means of expression, where pursuits for aesthetics can further yield unique, beautiful pieces [52]. Many designers and artists leverage design expression to craft personally-relevant and evocative visualizations susceptible to provoking emotional responses from the audience. One example demonstrating a high level of design expression is the Dear Data project [32], featuring a collection of hand-drawn data visualizations. Recent studies investigating visual thinking on whiteboards [9] and data sketching [55] highlighted the expressiveness of sketched visuals when working with data.

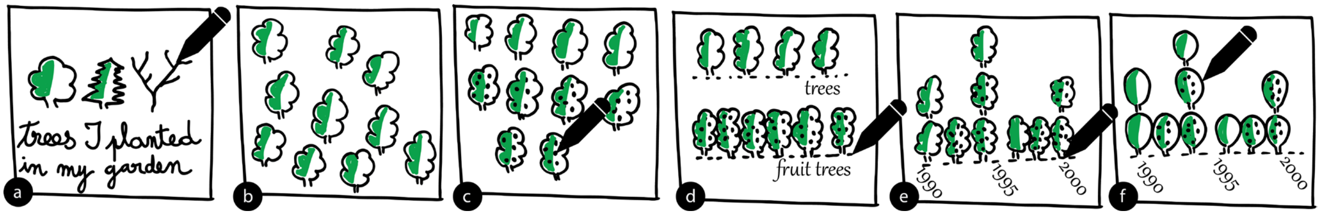

Figure 2. A storyboard illustrating the creation of a visualization using the Data-Oriented Drawing approach. a) Sketching visual designs using digital ink on a canvas enables one to experiment with various shapes. b) Selecting one of the shapes to specify a data-visual mapping automatically populates the canvas with all relevant data points. c) Sketching compound glyphs to represent additional data dimensions. d) Drawing a layout based on a data dimension to structure the data spatially. e) Redrawing the layout to map the data to a different data dimension. f) Redrawing a visual mark for a different data dimension.

Sketching is widely recognized as an excellent instrument for the quick generation of multiple designs due to its inherent freeform and effortless nature [6, 41]. Yet, the lack of automation makes it hard to envision how resulting visuals might be mapped to actual data.

The same issue holds for graphics design and illustration tools such as Adobe Illustrator [1] or Affinity Designer [2], which focus on design expression but lack support for mapping realized visuals to data. Bigelow et al. [4] provide insights on the laborious process in which designers alternate back and forth between graphics authoring and data visualization tools to realize stylized visuals. Recent efforts have tried to simplify this process by providing bridges between tools [5] or integrating data-driven widgets in illustration interfaces [26]. Such efforts are great avenues to reconcile the expressive power of graphic design software to generate data visualizations. Yet, they impose a high threshold [50] on users, as they still require to master complex functionalities, making it laborious to generate alternative designs, which can inhibit creative exploration.

Creative Exploration

Pen and touch enabled interfaces leverage natural human sketching and physical manipulation skills, empowering users to pursue creative tasks in a fluid workflow [19]. The visualization community sees these interfaces as a promising alternative to WIMP UIs, allowing analysts to focus on the data under study, rather than how to operate the interface [29]. Several visualization tools take advantage of pen and touch input for data exploration [9, 14, 22, 43, 60] and its presentation to an audience [30, 31].

However, pen usage is often limited to making simple annotations on customized data views. Only rarely is the expressiveness of sketching exploited to create new simple stroke-based visuals (e.g. [30]). In the context of graphics authoring, direct manipulation approaches such as object-oriented drawing [59] have been introduced to facilitate the creative exploration of visual designs by providing a novel and more direct interaction metaphor to manipulate the visual properties of graphics. The present work continues to pursue this line of research, investigating the power of pen and touch interaction to enable authors to create personalized, expressive data visualizations.

DATA-ORIENTED DRAWING

We propose *Data-Oriented Drawing* (DOD) a novel approach to fluidly experiment with visual designs and create expressive data visualizations that support visual-to-data bindings (Figure 2). Informed by [47], we articulate a set of five design principles illustrated in Figure 3.

Principle 1. Support Freeform Inking & Direct Manipulation

To provide interaction that *supports exploration* [47], authors should be supported to rapidly generate many different alternative designs and explore various ways to bind visuals to data. We draw inspiration from designers' and artists' general workflows [4], where designs are first explored by sketching on pen and paper or whiteboards. A similar workflow should be foster the rapid experimentation with visual marks, by providing freeform sketching and augmentation on a digital canvas (Figure 3:1a), through easy access to the visual properties of graphics via direct manipulation (Figure 3:1b).

To compose compelling visual designs for a given dataset, authors must make many assumptions about the dataset's structure, often introducing errors within a design [4]. To overcome this issue, authors should be able to access data dimensions and visual properties of graphics at any time through direct physical manipulation (Figure 3:1c) [19, 59].

Principle 2. Focus on Glyph Composition

To empower authors to craft expressive and personalized visual representations of their data, we propose a focus on glyph-based visualizations [7]. In these types of visual representations, each data point is associated with a *glyph*, i.e., a compound visual mark whose visual properties are dictated by one or more dimensions of the data point it represents. Glyph-based representations [33, 40] are commonly found in visualization for communication purposes [38] as they afford expressivity in the composition of the marks, while also offering a concise presentation of multivariate information. They also simplify the specification of the visual-data mapping since they do not require complex specification grouping functions.

Our second principle proposes to empower authors to create visually-rich marks portraying multiple facets of data points (Figure 3:2), by iteratively incorporating components to realize compound glyphs.

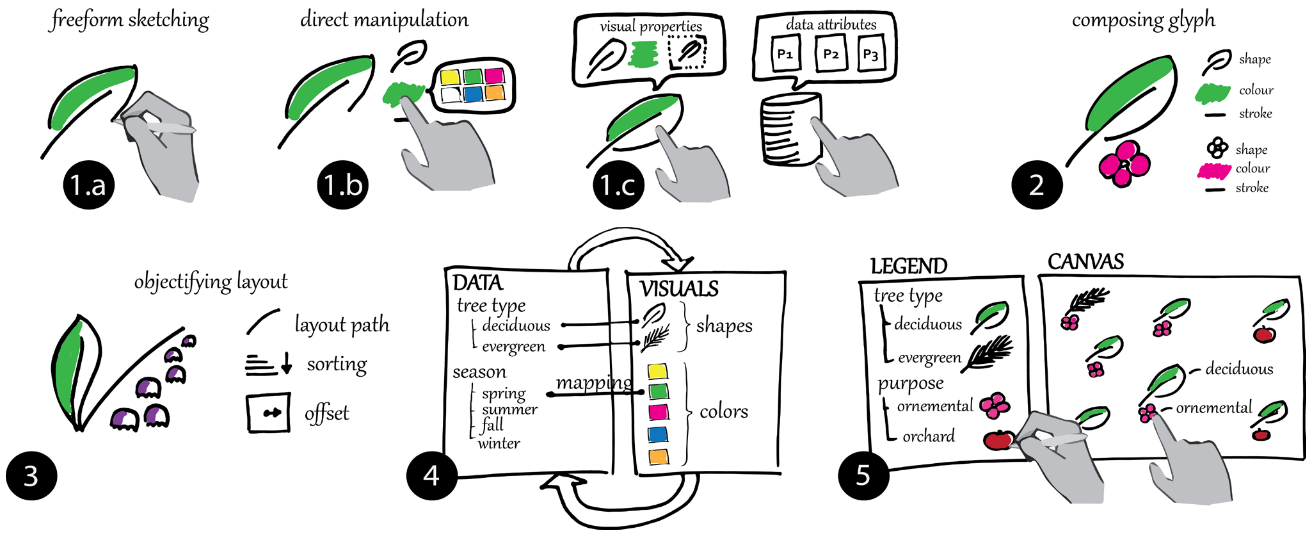

Figure 3. The five principles of Data-Oriented Drawing. 1) Freeform sketching (1.a), direct manipulation (1.b, 1.c) to enable flexible authoring of visual designs, and quick access to both visual properties of glyphs and data attributes. 2) Composing glyphs specifying visual properties to data dimensions. 3) creating layouts and editing them as objects. 4) specifying visual-data mapping from visual variables or data dimensions. 5) supporting multiple workflows from the legend or the canvas.

Principle 3. Objectify the Layout

In addition to encoding data by composing glyphs, we also aim to enable authors to encode data by structuring the layout of these glyphs in space.

To make this specification as direct as possible, our third principle proposes to objectify the concept of layout [12], and treat this reification similarly to glyphs. This suggests considering the layout as a visual mark that an author can directly draw on the canvas, manipulate the properties of, and bind to a data dimension. This visual mark can also serve to select groups of glyphs and move them in space via direct manipulation for example. We propose to treat properties of the layout such as the shape of the distribution path and the order of glyphs along this path similarly to the visual properties of glyphs such as shape and color.

Our third principle proposes to empower authors to create a spatial structure encoding different dimensions of the data by iteratively drawing and composing the outcome they envision (Figure 3:3).

Principle 4. Support Bidirectional Visual-Data Mapping

The realization of a data visualization requires one to specify a set of mappings between data dimensions and visual variables [13].

To provide maximum flexibility, our fourth principle proposes to enable authors specifying visual-data mappings from either direction (Figure 3:4). In simple terms, author may start from a visual property (e.g. shape, color) and experiment with different data dimensions it could encode; or they may start from a data dimension and experiment with different visual encodings that could represent it.

Principle 5. Enable Legend & Visualization Composition

To foster creative exploration, it is essential to support multiple workflows to create data visualizations.

Data visualizations experts, for example, may use a generative approach, making decisions at the data dimension level. They may choose visual encodings for each data dimension and values successively (e.g., map each value of ‘tree type’ to a shape then map each value of ‘temperature’ to a color, etc.). Put differently, they decide on a set of rules to generate glyphs: the *legend*. Designers on the other hand, may take more of a design-by-example approach, making decisions at the data point level, by iterating on the design of the final glyph for individual data points (e.g., draw a glyph representing a tree that grows in a warm environment and produces fruit). Put differently, they decide on the designs of glyphs first: the *visualization,* and derive generative rules later.

Supporting authors to follow either one of these workflows or fluidly switch between them, our last principle proposes to use a legend and visualization canvas metaphor (Figure 3:5). Interaction within the legend enables authors to specify mappings at the data dimension level, while interaction within the canvas enables them to specify mappings at the data point level.

To summarize, these five principles support the creative construction [47] of data visualizations by providing:

- *low threshold*: easy entry for novices,
- *high ceiling*: experts to achieve sophisticated pieces,
- *wide walls*: a wide range of possible explorations,
- *many styles and many ways*: multiple workflows.

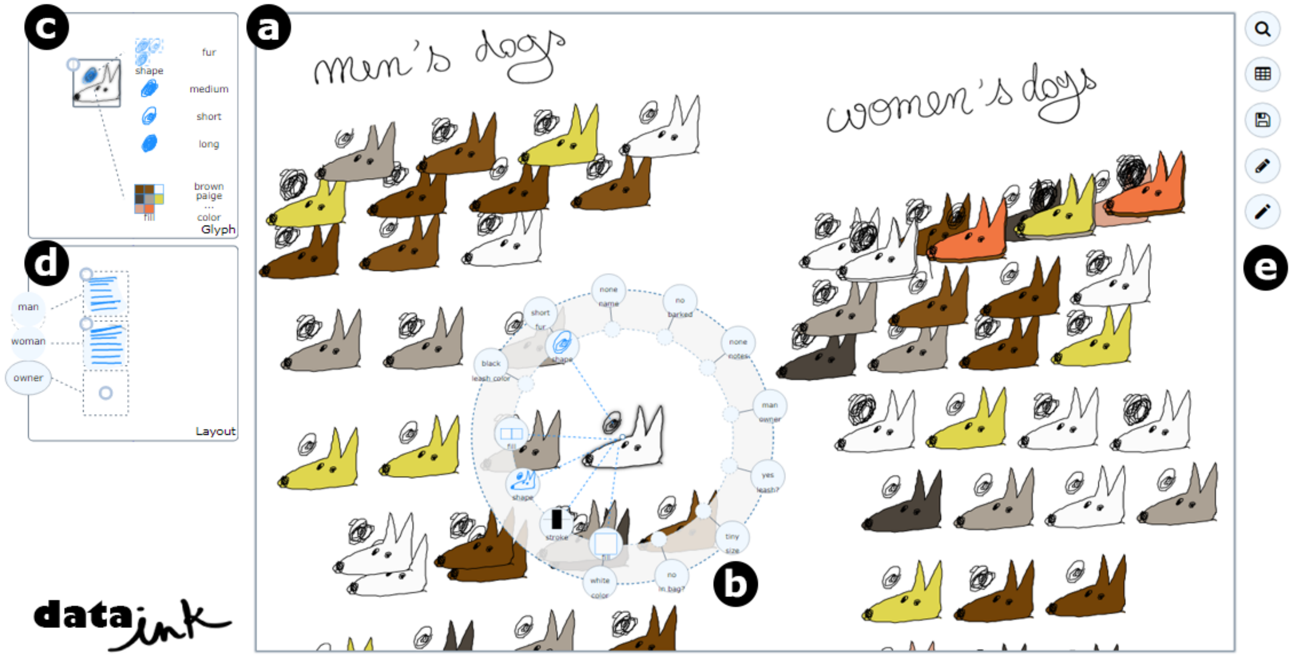

Figure 4. DataInk interface: a) main canvas to create data visualizations. b) the visual-data palette to edit visual properties and bind them to existing data dimensions directly from the canvas. c) glyph panel to define visual-data mappings similar to a legend. d) layout panel to structure data dimensions. e) menu to load data, search for vector graphic visuals and store visualizations.

DATAINK

The DataInk interface (Figure 4) is an application of our five principles. We first describe the workflow to create a data-driven drawing, then delve into the design of each interface component with close-up figures.

DataInk Workflow

To engage with her colleagues at her new job, Emma decided to create a fun, whimsical visual representation of her personal data. Inspired by the Dear Data project, she recorded the dogs she encountered during her daily commute during the week in a spreadsheet, where each row is a dog and each column represents one of its characteristics (e.g. color, length of fur, whether their owner was a man or a woman).

Emma opens DataInk on her pen-and-touch enabled tablet and loads her spreadsheet. She starts by experimenting with different dog shapes by drawing on the canvas (Figure 4:a). Looking at the different shapes she drew, she taps on her preferred design and binds it to her data dimensions. The system instantaneously duplicates the dog drawing, and spread the newly created collection of dogs randomly on the canvas. Each drawing now represents a row in the data table, or data point.

As Emma taps on a dog, a visual-data palette (Figure 4:b) appears depicting the visual properties of the drawing in the inner ring, and the values for each data dimension in the outer ring. Working at the data point level, she can modify the color of her drawing from white to brown because the data value of this dog shows that it was brown. She then can create a mapping from the color visual encoding to the color data dimension by aligning the inner and outer rings.

The legend (Figure 4:c) thus updates to display the color mapping. Emma taps the legend to access the list of colors and sets them successively. She then decides to encode a second data dimension, the fur length, by adding a shape to her dog drawing. She taps a dog, draws on the canvas just above its head and map the shape of this mark to the corresponding attribute using the visual-data palette. She then taps a different dog, and changes the shape by drawing in a contextual interactive panel invoked directly from the touching the visual shape item in the palette or legend.

Satisfied by the design of her current glyph, Emma decides to organize the space. She does not really know what data dimension would provide an interesting structure. Working at data dimension level, she taps the layout legend and browses through the different data dimensions (Figure 4:d). As the data dimensions are in focus, DataInk previews the different groupings in the canvas. Given the number of dogs she has in her visualization, Emma opts for structuring the space by the gender of the dog owners. The dogs appear now into two distinct groups on the canvas. Emma can touch each group and move them in the canvas. As she taps the layout legend, she can select a particular group and simultaneously draw distribution path in the canvas. This enables Emma to experiment with different layout shapes for each group. She concludes by saving her work and sending it to the large format printer (Figure 4:e).

See supplemental material for the video of this scenario.

Figure 5. Object-Oriented Drawing : sketching or reusing visuals and directly editing visual attributes on the canvas.

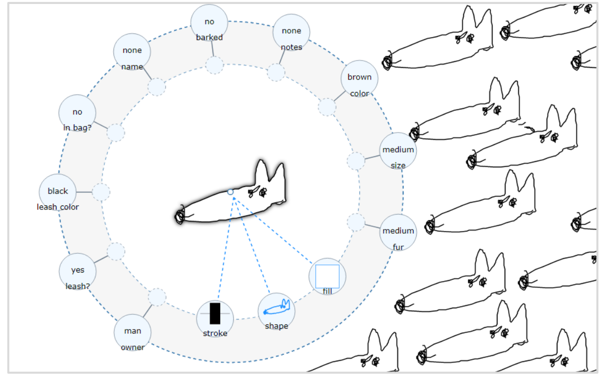

Figure 6. Visual-Data Palette - visual attributes distributed along the inner ring, data attributes along the outer ring.

UI elements

DataInk is composed of five UI components (Figure 4). The visualization canvas (Figure 4a) shows the generated visual representation of the data, as well as any visual elements not bound to data such as handwritten text or illustrative marks. A contextual visual-data palette appears when the authors taps on a visual element, enabling them to set appropriate mappings (Figure 4b). The legend depicting the sets of mapping between visual encodings and data dimensions are grouped in two panels: one for the glyph (Figure 4c) and one for the layout (Figure 4d). A side menu (Figure 4e) allows an author access to functionality such as searching vector graphics, loading data, saving the visuals, as well as providing access to several illustration tools.

Object-Oriented Drawing

DataInk offers *freeform sketching* (Principle 1) to enable authors to ideate on graphical elements that can later serve to compose data visualizations. Authors generate these graphical elements by directly drawing on the canvas with a digital pen or writing in a search box to import vector graphics. Following the principle of *direct manipulation* (Principle 1), the properties of each visual mark on the canvas can be accessed and modified by touch interaction. DataInk follows an object-oriented drawing approach [59], materializing visual properties of the encodings by interactive cards that authors can interact with to adjust parameters (e.g., Figure 5 depicts the interaction to adjust the color of a stroke in a graphical element).

Once the author binds a visual mark to represent its data, this mark effectively becomes a glyph. Each glyph represents a data point (e.g., a row of the data table). By tapping each glyph, the author can now access both visual encodings and data values for that data point via a the visual-data palette contextual menu (Figure 6).

Visual-Data Palette

We designed the visual-data palette to enable *bidirectional mappings* between data dimensions and visual encodings (Principle 4). This palette is composed of two rings (Figure 6), whereby the inner ring depicts a list of visual encodings for the selected visual and the outer ring depicts a list of data values if this visual represents a data point. The authors can specify a mapping by aligning items in both rings (e.g., in Figure 7, the author set the fill color of the shape to encode the data dimension 'color').

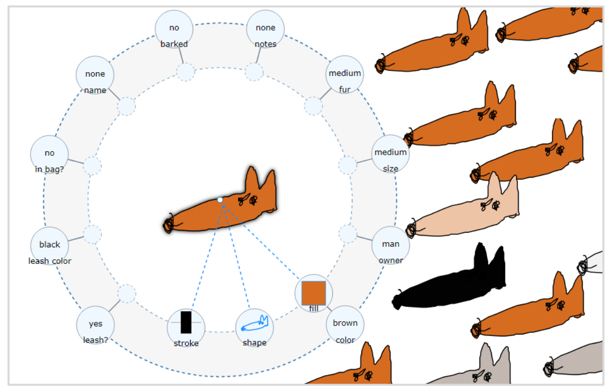

Figure 7. Visual-Data Palette – binding the fill color (visual) to the color of the dog (data).

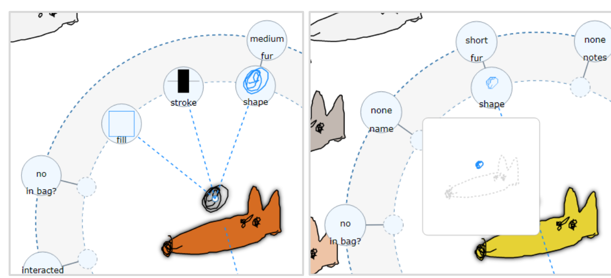

Figure 8. Visual-Data Palette – adding a new shape to create a compound glyph.

To compose compound glyphs, the author can directly draw additional shapes on the canvas. The visual properties of these shapes will appear in the palette for authors to further map to data (Figure 8a). As every item in the palette is interactive, authors can open a card representing the shape of a data point and directly draw the alternative shape they desire. Shapes are propagated to the other data points with the same data value. In our example with Emma, she drew a small fur ball to represent a short haired dog (Figure 8b).

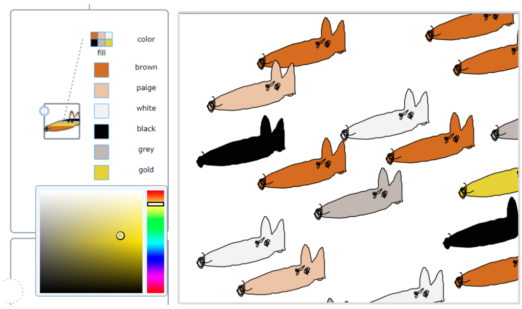

Figure 9. Editing visual properties from the glyph legend.

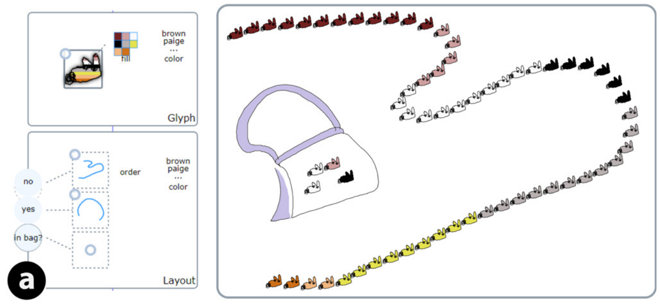

Figure 10. Specifying distribution path for a data dimension from the layout legend (a).

Glyph and Layout Legends

The visual-data palette enables authors to reason at the data point level, inside the visualization canvas. To provide a *flexible workflow* (Principle 5), DataInk also enables authors to work at the data dimension level, inside legends. DataInk provides two types legends: a legend for *glyphs* (Principle 2), and a legend for the *layout*s (Principle 3).

The glyph legend depicts the visual-data mappings of the glyph (Figure 9). The current implementation of DataInk supports shape, fill color, and stroke color for each component of the glyph. Interacting with the legend enables authors to set mappings for each value of a data dimension successively (e.g., assigning a color hue to each data value; Figure 9). Working at the data point level, in the visualization canvas, enables to experiment with visual-data mappings whereas working at the dimension level is effective to map the complete set of values once the author decides on a visual-data mapping.

The layout legend depicts the visual-data mappings for the layout of the visualization (Figure 10). The author can preview the effect of different groupings by browsing the set of data dimensions available from the data items on the left of the legend (Figure 10a). Once a data dimension is selected for the grouping, touching one of the values enables to select the corresponding set of glyphs on the canvas. Authors can then draw a distribution path for each group directly in the canvas. In a similar spirit to the glyph legend, additional structural properties of the layout, such as the sorting order of the glyphs along the distribution path, are visually represented in the legend.

Pen and Touch Interactions

DataInk provides a simple vocabulary of pen and touch interactions for visualization canvas and legend (Figure 11).

Authors directly draw glyphs or layout distributions with their pen. Touch enables them to move their drawing in space and reveals the visual-data mappings, selecting a shape with the pen to edit them.

Interactions with the visual-data palette are based on drag and drop interactions with items located on the rings. To create a binding, authors drag a visual property (inner ring) and drop it into the slot afforded by the data property (outer ring). Once mapped, both items are visible linked. Authors can break this mapping by holding one item and dragging the other away.

Authors add elements to a glyph by selecting it and drawing a new shape in its vicinity. They can also transfer mappings they previously created to different components. For example, they may transfer the color representing the owner's gender from the head of the dog to the leash. Instead of recreating this mapping from scratch, authors can hold it with one finger and use their pen to select the shape that they want to transfer it to.

Touching visual properties items in both canvas and legend brings interactive panels that authors can touch or draw in to modify such as drawing a new shape or picking a color hue. Finally, DataInk provides interaction across legend and canvas. For example, layout is supported by holding a data dimension from the legend to select the subset of glyphs to layout and drawing their distribution on the canvas.

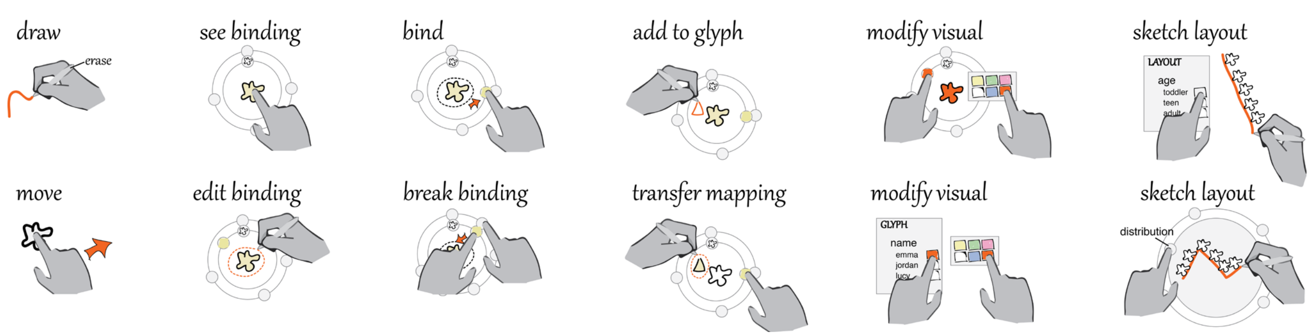

Figure 11. Vocabulary of Pen and Touch Interactions in DataInk

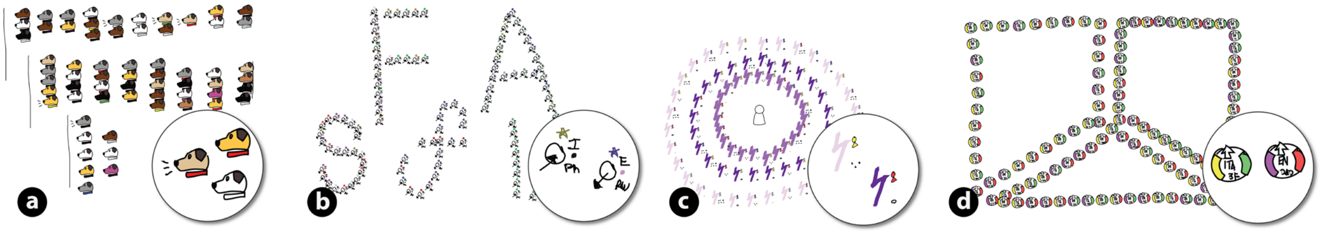

Figure 12. Examples of design realized by participants during the freeform exploration tasks.

USER STUDY

We conducted a user study to gain insights into the potential of DOD in supporting design expression, rigorous execution and creative expression, as well as the usability of DataInk for authoring data visualizations. We targeted designers to understand tradeoffs of DOD compare to graphics design tools, and invited non-experts (limited experience in programming, data analysis and design) to assess the receptiveness of DOD by a general audience.

Apparatus

DataInk was implemented as a web application, running on a Surface Studio, where finger and stylus input can be simultaneously detected and differentiated.

Participants

We recruited 8 participants (3 males; Age:18 to 39, Average: 26). Four of them (D1-D4) were designers who reported having more than 7 years of experience using digital drawing applications. The others were non-experts (P1-P4) who reported having no or very limited experience in digital drawing. All participants but P2 reported having no prior experience in creating data visualizations; P2 had experience using D3 to create charts for his own research. Participants were provided with a $50 (local currency) honorarium for an approximately 90-minute session.

Protocol

Participant first filled a consent form and a demographics questionnaire, then completed three tasks with DataInk, concluding by a feedback questionnaire and an interview.

Demonstration and Training (25 min).

The experimenter demonstrated the underlying concepts of DataInk, including basic sketching, glyph composition, visual-data mapping and layout via a walkthrough of an example. Participants were instructed to replicate the exact same example, seeking guidance whenever necessary. The experiementer also encouraged them to freely explore the interface to try every functionality.

Replication and Iteration Exercises (20 min)

The experimenter then proceeded to the replication task. This task consists in asking participants to replicate a data visualization for a new dataset without guidance. Participants received the target visualization on a paper sheet, along with an accompanying legend describing the visual encodings. We design this exercise to assess the learning curve necessary for people to use DataInk to recreate a design from scratch and bind data to its visual elements. The second task required participants to alter and iterate on their design, requiring them to make four changes to the glyph composition and three changes to the layout given an example provided by the experimenter.

Freeform Exploration (25 min)

After completing the replication and iteration exercises, the experimenter ask participants to realize their own data visualizations (Figure 12). Half of the participants (D1, D2, P1, P2) started with the dataset used in the previous stage to give us an opportunity to assess the expressivity of the system. We provided the other half (D3, D4, P3, P4) with a new dataset to observe the entire creation process.

Questionnaire and Interview (20 min)

Participants concluded the study by a questionnaire, addressing expressive power and usability of our prototype using a 5-point Likert scale (1–strongly disagree, 5–strongly agree). The experimenter then conducted a semi-structured interview to collect qualitative comments on the expressiveness, utility, and usability of DOD and DataInk.

Results

Overall, all participants successfully completed the tasks and created data visualizations for each dataset. We did not observe any notable differences in terms of DataInk usage or quality of the outcomes among designers and non-experts, though our sample size is small. In the following, we report subjective ratings and qualitative comments made by participants that suggest that DataInk enabled them to easily get started (i.e. had a low threshold) and create expressive data visualizations (i.e. had a high ceiling), while supporting a wide range of data-visual mappings (i.e. had wide walls) in a fluid and flexible workflow (i.e. supporting many paths, many styles).

Low Threshold - Learning Curve and Usability

All participants found the interface easy to learn (5/8 strongly agree and 3/8 agree) and easy to use (5/8 strongly agree and 3/8 agree). As D2 noted, "*your tool to Illustrator is like SketchUp to AutoCAD.*" They also found the interface had "*no high-skill cap*" (D3) and was "*fun to use*" (D2, D3, D4, P3, P4).

Notably, the usability of the Visual-Data Palette was a recurrent theme discussed in the open comments. Several participants found the drag-and-drop nature of the visual attributes to data attributes to be "*very easy and intuitive*" (D2, D3, D4, P1, P2, P4). D4 noted "*it's like tether the actual data and the visual representation*".

The locality and the design of the Visual-Data Palette enabled participants to "*quickly pick up the functionalities of the system*" (P1), as "*it doesn't distract [them] at all*" (P1). When compared to traditional graphic editing software, P2 noted: "*I can't remember all the stuff and things are usually several click away*". In contrast, participants found the palette enabled them to "*have an overview of all the available data attributes*" (P1), and that "*the slot on the inner ring is good visual indicator of the mapping*" (D4). D1 and D3 noted the "*new set of tools*" provided by the interface are "*unfamiliar*" and "*different*", but the Data-Oriented Drawing is "*simple, straightforward, consistent, and easy to grasp*".

High Ceiling – Expressive Power

Participants responded positively about the expressive power of the interface, strongly agreeing (6/8) or agreeing (2/8) that it was easy to create the desired visuals, and strongly agreeing (5/8) or agreeing (3/8) that they were satisfied with the range of visualization designs the system is capable of. Participants highlighted the range of possibilities that DataInk affords: *"You have the micro of what you can change, which is the glyph, right? And you have the macro which is the overall layout with the shapes, you can get really crazy funky with it."*

Participants also favored the expressiveness of the interface, which enabled them to achieve sophisticated outcomes with a simple consistent set of interactions to specify the visual-data mapping, e.g., "*The foundation is really good and I can see how it scales up, because you can just populate the functionalities with more and more attributes, but they all work the same way. I can totally see its potential*" (D4). Participants commented that the sketching and rapid generation of data-binded graphics inspired them to become more creative, e.g., *"I feel like I could get more creative. I like that outcome which was easy to do, easy to read, info clearly communicated, and visually appealing." (P3)*

Wide Walls, Many Paths, Many Styles– Creative Exploration

For the Iteration task, all participants successfully completed the requested changes, and rated that it was easy to change glyphs (i.e., strongly agree (4/8) or agree (4/8)) and layout (i.e., all strongly agree).

Participants commented positively on their ability to experiment with different alternatives: "*when you showed me the new one, my first impression is: I will start a new one, because the task is pretty hard. I am surprised I can do it*"(P3). Comments and ratings on the ability for them to experiment with different alternatives (i.e., strongly agree (5/8) or agree (3/8)) suggests that DataInk supported their creative explorations.

Figure 12 features a sample of visualizations realized at the Freeform Exploration stage. Participants were all satisfied by their final creation (3 designers and 2 non-experts responded "Strongly Agree" and 1 designer and 2 non-experts responded they "Agree").

During each of these tasks, we observed different workflows and creation strategies. All participants started by drawing basic shapes to act as a glyph representing a data point. After the glyphs populates the canvas, instead of electing to see the data attribute of a data point, D1 and P3 started with the layout, as they "*want[ed] to know how many different things are in the dataset, like how many data attributes, how many different values of them, and how many data points in each group*". After seeing the different categories, D1 continued exploring different layouts, while P3 switched to exploring how to compose the glyphs. The rest of the participants started by composing the glyphs and then creating the layout. All participants went back and forth between glyph making and layout editing to see whether there were data attributes left to be visualized or to change the mapping.

We also observed participants fluidly switching between visual-data palette and legend. Participants typically experimented with new drawings mappings directly on the canvas. Then, they switched to the legend to set the mapping for every value of a data dimension, e.g., "*the dragging and the coding on the legend make it very simple and fast to do what I want to do" (D3)*. Participants commented that always being able to see the data dimensions was useful for their creative process on the canvas (e.g., "*…This helps me to think about the creation of the glyph, deciding which data attributes to pick and what would be the most appropriate visual form or layout to represent them" (P1)),* and legend *(e.g., "Looking at the attributes, I really have the urge to map them all!!" (P3)).*

Every participant explored different mappings for glyphs and layouts (Figure 12). The ability to propagate changes to other glyphs and setting the mapping on the attribute level in the legend were reported as key factors to support rapid exploration of different alternatives, e.g., *"I like the capability the tools provide for easily duplicating glyphs and mapping data attributes to visual elements. I am happy to be able to create a data visualization containing so much data in such short time."* (P1), and *"It's fast. You can change the drawing and mappings easily. It's something everyone can easily put together." (D3).*

Suggestions for Improvement

We observed a number of usability issues participants encountered such as quickly locating a data dimension of interest. The DataInk palette orders existing mappings in relation to the part of the glyph they apply to. However, this design induces changes in the order of items in the outer ring. D1 and P2 suggested keeping this order persistent instead, sorting them in alphabetical order for easy navigation. D4 also suggested having a search function of data attributes for large complex datasets. Participants also suggested several features such as multiple glyph and layout panels (P3) to easily change and compare different designs, and moving these legend inside the canvas (D3) to take advantage of the entire screen space.

DISCUSSION AND FUTURE WORK

The study suggests that the principles of DOD implemented in DataInk are promising to support creative exploration, design expression, and the rigorous execution of data visualizations. We reflect on each of these goals, outlining challenges and opportunities for future research.

Blurring the Line with Data Exploration

When considering creative exploration, Principles 1, 4 and 5 suggested interactions mechanisms to explore the different dimensions of the data and fluidly experiment with visual-data mappings. Reflecting on our study observations, we believe that we can enrich DataInk to provide more data exploration capabilities. In particular, we plan to enrich the interactions with the legend (Principle 5) to enable authors to browse and manipulate data dimensions and values.

Expanding Design Expression Features

When considering design expression, DataInk only offers a subset of the features available in professional illustration software. While there is a trade-off between low thresholds (i.e., learning curves) and high ceilings (i.e., expressivity), we believe that there are opportunities to support greater expression without compromising the interaction paradigm. We ground our discussion with examples from Dear Data.

Glyphs

DataInk supports a subset of the visual properties available for encoding. Adding support for additional ones such as size or opacity is directly achievable. Another interesting addition is to incorporate procedural drawing techniques such as those demonstrated in Vignette [25] and Para [21]. To reproduce the glyph made of concentric circles in Figure 13:a, DataInk requires authors to draw a different shape for each data value. This process forces authors to provide a categorical visual encoding for continuous data value, while procedural drawing enables to generate continuous ones.

Describing how to generalize DOD to encompass every data visualization imaginable is out of scope, however, we reflect on a few straightforward extensions, as well as more challenging ones. A simple modification to our prototype would enable authors to create visual marks representing aggregates used in standard representations such as bar or bubble charts. DataInk could support this activity by enabling authors to bind properties of visuals to the results of operations on groups of data points. For example, one could create groups of dogs by fur length (via the same mechanism provided to group glyphs for the layout). Then, by providing a set of operations on groups (e.g. count), one could map the size of a drawing to the results of this operation. DataInk would thus generate one drawing per group, rather than one per data point.

Other types of visualizations would require more profound changes. For example, node-link diagrams (Figure 12b) rely on two types of glyphs: nodes and links. This may result in a more complex creation process, as it introduces dependencies between nodes and links glyphs when generating the visualization.

Figure 13. Examples from Dear Data [32] requiring more features: a) shape generation for continuous data; b) node-link diagrams; c) hierarchical or d) nested layouts.

Layout

One of the more challenging features to tackle pertains to the specification of the layout of visual marks in space. Our current prototype enables authors to specify a single level of grouping of glyphs as well as their distribution path and sorting order. Supporting the hierarchical and nested spatial structure that can be seen in several examples of Dear Data (Figure 11c) would require an author to specify multiple levels of groupings, distribution paths, and orders. In our prototype, this could be achieved by specifying multiple layout cards, however, the order of layout cards introduces dependencies and constraints regarding the possibilities of the nested layout structure. Conveying these constraints to authors in a transparent way, and enabling them to fluidly experiment with layout hierarchies, raises new challenges.

Bringing Creativity to the Next Level

DOD goes beyond the generation of static visualizations. We envision many opportunities to support creative and expressive designs for interactive and dynamic data visualizations. The philosophy and general directness of our approach combines well with techniques such as Draco [24] or Kitty [23], enabling authors to specify the dynamic and interactive behavior of graphics through sketching. A direct opportunity for future research would be to explore how interactions afforded by these systems can be integrated into DataInk and augmented to support data binding.

CONCLUSION

This paper proposes Data-Oriented Drawing, an approach for authoring whimsical, personalized visual representations of data. DOD aims at supporting the creative process while affording design expression to craft visually-rich graphics via freeform sketching, and rigorous execution by maintaining a tight coupling between visuals and data via direct manipulation. The principles of DOD are embodied in DataInk, a pen-and-touch drawing application that enables both rich expressivity and effortless execution of visualizations of data. A user study with eight designers and non-experts suggests that this approach is promising. DataInk enabled authors to experiment with visual designs, ultimately creating a diverse set of glyph-based representations with minimal training.

ACKNOWLEDGEMENTS
Left blank for peer review.

REFERENCES

1. Adobe Illustrator 2017, Retrieved September 16, 2017 from www.adobe.com/Illustrator
2. Affinity Designer, Retrieved September 16, 2017 from https://affinity.serif.com
3. Jacques Bertin. 2010. *Semiology of Graphics: Diagrams, Networks, Maps (1st ed.)*. ESRI Press, Redlands, Calif.
4. Alex Bigelow, Steven Drucker, Danyel Fisher, and Miriah Meyer. 2014. Reflections on How Designers Design with Data. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (AVI ’14), 17–24. https://doi.org/10.1145/2598153.2598175
5. Alex Bigelow, Steven Drucker, Danyel Fisher, and Miriah Meyer. 2017. Iterating between Tools to Create and Edit Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 481-490. https://doi.org/10.1109/TVCG.2016.2598609
6. Bill Buxton. 2010. *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann.
7. Rita Borgo, Johannes Kehrer, David H.S. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, Min Chen. 2013. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. *Eurographics State of the Art Reports*, 39-63.
8. Mike Bostock, V. Ogievetsky, and J. Heer. D^3 Data-Driven Documents. 2011. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309. https://doi.org/10.1109/TVCG.2011.185
9. Jeffrey Browne, Bongshin Lee, Sheelagh Carpendale, Nathalie Riche, and Timothy Sherwood. 2011. Data analysis on interactive whiteboards through sketch-based interaction. In *Proceedings of the International Conference on Interactive Tabletops and Surfaces* (ITS ’11), 154-157. https://doi.org/10.1145/2076354.2076383
10. Alberto Cairo. 2012. *The Functional Art: An introduction to information graphics and visualization*. New Riders.
11. Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, eds. 1999. *Readings in information visualization: using vision to think*. Morgan Kaufmann.
12. Marianela Ciolfi Felice, Nolwenn Maudet, Wendy E. Mackay, and Michel Beaudouin-Lafon. 2016. Beyond Snapping: Persistent, Tweakable Alignment and Distribution with StickyLines. In *Proceedings of the ACM Symposium on User Interface Software and Technology* (UIST ’16), 133-144. https://doi.org/10.1145/2984511.2984577
13. William S. Cleveland and Robert McGill. 1987. Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data. *Journal of the Royal Statistical Society. Series A (General)* 150(3):192–229. https://doi.org/10.2307/2981473
14. Andrew Crotty, Alex Galakatos, Emanuel Zgraggen, Carsten Binnig, and Tim Kraska. 2015. Vizdom: interactive analytics through pen and touch. In *Proceeding of VLDB Endow*, 8(12):2024-2027. http://dx.doi.org/10.14778/2824032.2824127
15. DensityDesign Research Lab. RAWGraphs. Retrieved September 16, 2017 from http://rawgraphs.io/
16. Drupal. EasyCharts. Retrieved September 16, 2017 from http://www.easychart.org/
17. Lars Grammel, Chris Bennett, Melanie Tory, and Margaret-Anne Storey. 2013. A Survey of Visualization Construction User Interfaces. In *EuroVis - Short Papers*. http://dx.doi.org/10.2312/PE.EuroVisShort.EuroVisShort2013.019-023
18. Lars Grammel, Melanie Tory, and Margaret-Anne Storey. 2010. How Information Visualization Novices Construct Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 16(6): 943–952. https://doi.org/10.1109/TVCG.2010.164
19. Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddingotn, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. 2010. Pen + touch = new tools. In *Proceedings of the ACM Symposium on User interface software and technology* (UIST ’10), 27–36. https://doi.org/10.1145/1866029.1866036
20. Samuel Huron, Sheelagh Carpendale, Alice Thudt, Anthony Tang, and Michael Mauerer. 2014. Constructive Visualization. In *Proceedings of the Conference on Designing Interactive Systems* (DIS ’14), 433–442. https://doi.org/10.1145/2598510.2598566
21. Jennifer Jacobs, Sumit Gogia, Radomir Mech, and Joel Brandt. 2017. Supporting Expressive Procedural Art Creation through Direct Manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ’17), 6330-6341. https://doi.org/10.1145/3025453.3025927
22. Jaemin Jo, Sehi L'Yi, Bongshin Lee, and Jinwook Seo. 2017. TouchPivot: Blending WIMP & Post-WIMP Interfaces for Data Exploration on Tablet Devices. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI ‘17), 2660-2671. https://doi.org/10.1145/3025453.3025752

23. Rubaiat Habib Kazi, Fanny Chevalier, Tovi Grossman, and George Fitzmaurice. 2014. Kitty: Sketching Dynamic and Interactive Illustrations. *Proceedings of the ACM Symposium on User interface software and technology* (UIST '14), 395–405. https://doi.org/10.1145/2642918.2647375
24. Rubaiat Habib Kazi, Fanny Chevalier, Tovi Grossman, Shengdong Zhao, and George Fitzmaurice. 2014. Draco: bringing life to illustrations with kinetic textures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 351-360. https://doi.org/10.1145/2556288.2556987
25. Rubaiat Habib Kazi, Takeo Igarashi, Shengdong Zhao, and Richard Davis. 2012. Vignette: interactive texture design and manipulation with freeform gestures for pen-and-ink illustration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (CHI '12), 1727-1736. https://doi.org/10.1145/2207676.2208302
26. Nam-Wook Kim, Eston Schweickart, Zhicheng Liu, Mira Dontcheva, Wilmot Li, Jovan Popovic, Hanspeter Pfister. 2017. Data-Driven Guides: Supporting Expressive Design for Information Graphics, *IEEE Transactions on Visualization & Computer Graphics*, 23(1):491-500. https://doi.org/10.1109/TVCG.2016.2598620
27. https://doi.org/10.1145/2207676.2208302
28. Randy Krum. 2013. *Cool infographics: Effective Communication with Data Visualization and Design*. Wiley.
29. Bongshin Lee, Petra Isenberg, Nathalie Henry Riche, and Sheelagh Carpendale. 2012. Beyond Mouse and Keyboard: Expanding Design Considerations for Information Visualization Interactions. IEEE Transactions on Visualization and Computer Graphics 18(12): 2689-2698. https://doi.org/10.1109/TVCG.2012.204
30. Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. 2013. SketchStory: Telling More Engaging Stories with Data through Freeform Sketching. *IEEE Transactions on Visualization and Computer Graphics* 19(12):2416-2425. https://doi.org/10.1109/TVCG.2013.191
31. Bongshin Lee, Greg Smith, Nathalie Henry Riche, Amy Karlson, and Sheelagh Carpendale. 2015. Sketchinsight: Natural data exploration on interactive whiteboards leveraging pen and touch interaction. In *Proceeding of the IEEE Pacific Visualization Symposium* (PacificVis '15), 199–206. https://doi.org/10.1109/PACIFICVIS.2015.7156378
32. Giorgia Lupi, and Stephanie Posavec. 2016. *Dear Data*. Princeton Architectural Press.
33. Eamon Maguire, Philippe Rocca-Serra, Susanna-Assunta Sansone, Jim Davies, and Min Chen. 2012. Taxonomy-Based Glyph Design with a Case Study on Visualizing Workflows of Biological Experiments. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2603-2612. https://doi.org/10.1109/TVCG.2012.271
34. Gonzalo G. Méndez, Miguel A. Nacenta, and Sebastien Vandenheste. 2016. iVoLVER: Interactive Visual Language for Visualization Extraction and Reconstruction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '16), 4073–4085. https://doi.org/10.1145/2858036.2858435
35. Gonzalo G. Méndez, Uta Hinrichs, and Miguel A. Nacenta. 2017. Bottom-up vs. Top-down: Trade-offs in Efficiency, Understanding, Freedom and Creativity with InfoVis Tools. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '17), 841-852. https://doi.org/10.1145/3025453.3025942
36. Microsoft Excel 2016, Retrieved September 16, 2017 from https://products.office.com/en-us/excel
37. Microsoft Power BI, Retrieved September 16, 2017 from https://powerbi.microsoft.com/en-us/
38. Otto Neurath. 2011. *From Hieroglyphics to Isotype: A Visual Autobiography*, eds. Matthew Eve and Christopher Burke, Hyphen Press.
39. Donghao Ren, Tobias Höllerer, and Xiaoru Yuan. 2014. iVisDesigner: Expressive Interactive Design of Information Visualizations. IEEE Transactions on Visualization and Computer Graphics, 20(12):2092–2101. https://doi.org/10.1109/TVCG.2014.2346291
40. William Ribarsky, Eric Ayers, John Eble, and Sougata Mukherjea. 1994. Glyphmaker: creating customized visualizations of complex data. *Computer*, 27(7):57–64. https://doi.org/10.1109/2.299412
41. Yann Riche, Nathalie Henry Riche, Ken Hinckley, Sheri Panabaker, Sarah Fuelling, and Sarah Williams. 2017. As We May Ink? Learning from Everyday Analog Pen Use to Improve Digital Ink Experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '17), 3241-3253. https://doi.org/10.1145/3025453.3025716
42. Lisa Rost. Blog Post. Retrieved 12 September 2017 from https://source.opennews.org/articles/what-i-learned-recreating-one-chart-using-24-tools/
43. Ramik Sadana and John Stasko. 2014. Designing and implementing an interactive scatterplot visualization for a tablet computer. In Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14). ACM, New York, NY, USA, 265-272. https://doi.org/10.1145/2598153.2598163

44. Arvind Satyanarayan, Ryan Russell, Jane Hoffswell, Jeffrey Heer. 2016. Reactive Vega: A streaming dataflow architecture for declarative interactive visualization. *IEEE Transactions of Visualization and Computer Graphics,* 22(1): 659–668. https://doi.org/10.1109/TVCG.2015.2467091

45. Arvind Satyanarayan and Jeffrey Heer. 2014. Lyra: An interactive visualization design environment. *Computer Graphics Forum*, 33(3):351–360. https://doi.org/10.1111/cgf.12391

46. Carlos Scheidegger, Huy Vo, David Koop, Juliana Freire, and Claudio Silva. 2007. Querying and creating visualizations by analogy. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1560–1567. https://doi.org/10.1109/TVCG.2007.70584

47. Ben Shneiderman et al. 2006. Creativity support tools: Report from a U.S. National Science Foundation sponsored workshop. *International Journal of Human-Computer Interaction*, 20(2):61-77. http://dx.doi.org/10.1207/s15327590ijhc2002_1.

48. Julie Steele, and Noah Iliinsky. 2010. *Beautiful visualization: looking at data through the eyes of experts*. O'Reilly Media, Inc.

49. Tableau Software, Retrieved September 16, 2017 from https://www.tableau.com/products/desktop

50. Maryam Tohidi, William Buxton, Ronald Baecker, and Abigail Sellen. 2006. Getting the Right Design and the Design Right: Testing Many Is Better Than One. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), 1243-1252. https://doi.org/10.1145/1124772.1124960

51. Andrew Vande Moere. Infoesthetics blog. Retrieved September 12, 2017 from http://infosthetics.com/

52. Andrew Vande Moere, and Helen Purchase. 2011. On the role of design in information visualization. *Information Visualization,* 10(4): 356-371. https://doi.org/10.1177/1473871611415996

53. Fernanda B. Viegas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. 2007. ManyEyes: a Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics* 13(6):1121–1128. https://doi.org/10.1109/TVCG.2007.70577

54. Jagoda Walny, Sheelagh Carpendale, Nathalie Henry Riche, Gina Venolia, Philip Fawcett. 2011. Visual Thinking In Action: Visualizations As Used On Whiteboards. *IEEE Transactions on Visualization & Computer Graphics*, 17(12):2508-2517. https://doi.org/10.1109/TVCG.2011.251

55. Jagoda Walny, Samuel Huron, and Sheelagh Carpendale. 2015. An Exploratory Study of Data Sketching for Visual Representation. *Computer Graphics Forum*, 34(3). https://doi.org/10.1111/cgf.12635

56. Jagoda Walny, Bongshin Lee, Paul Johns, Nathalie Henry Riche, and Sheelagh Carpendale. 2012. Understanding Pen and Touch Interaction for Data Exploration on Interactive Whiteboards. *IEEE Transactions on Visualization and Computer Graphics,* 18(12):2779-2788. https://doi.org/10.1109/TVCG.2012.275

57. Hadlet Wickham. 2010. *ggplot2: Elegant Graphics for Data Analysis (Use R!). Second edition*. Springer.

58. Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2016. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):649–658. https://doi.org/10.1109/TVCG.2015.2467191

59. Haijun Xia, Bruno Araujo, Tovi Grossman, and Daniel Wigdor. 2016. Object-Oriented Drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '16), 4610-4621. https://doi.org/10.1145/2858036.2858075

60. Emanuel Zgraggen, Robert Zeleznik, and Steven M. Drucker. 2014. PanoramicData: data analysis through pen & touch. IEEE Transactions on Visualization and Computer Graphics (TVCG) 20, 12: 2112-2121. https://doi.org/10.1109/TVCG.2014.2346293