# **NAM WOOK KIM** · RESEARCH STATEMENT

My research vision is to **lower the barriers to understand and communicate complex data**. Data can capture a snapshot of the world and allow us to understand ourselves and our communities better. Yet, a lot of research efforts are geared toward building analytic tools for domain experts to explore the data and extract value from it. Such tools, although important, have are not friendly to a general audience. On the other hand, it is becoming more and more important to communicate data to the public, and ultimately equip everyone with skills to make sense of the data.

In my thesis, I tackle this issue by studying **visualization** within the broad context of **human-computer interaction**. By leveraging our visual perception, visualization provides a powerful means for understanding patterns in data without requiring advanced statistical literacy. It has been widely adopted across various fields to combat the overabundance of information in our society. To go beyond traditional expert systems, my research strives to establish a **research framework for creating visual, data-driven systems for the public** by exploring how to enable individuals to (1) design visualizations for communication and personal data [1, 2] and (2) author engaging stories about the data and evaluate them [3, 4], as well as (3) understanding the cognitive aspects of visualization and storytelling [5, 6].

My research advances the design space of visualization systems by inventing new ways to communicate information, and have led to notable recognition from the community. My approach also goes beyond exploratory analysis to presentation and storytelling, and shifts focus from data-oriented to human-centered systems. In the coming years, I will continue to rethink how we approach data and to further contribute to the **democratization of the data**.

## VISUALIZATION FOR COMMUNICATION AND PERSONAL DATA

The design space of standard charts like bar charts or scatter plots is now well-established and can quickly generate visualizations to facilitate rapid exploration of the data. On the other hand, thoughtfully crafted custom visualizations can be highly engaging and have the power to communicate the semantics of the data, that is not possible with the standard charts. However, existing visualization tools have limited support for customizing visual representations of the data, failing to meet this need.

My work takes a radical approach to devise new ways to enable custom visualizations from data. Instead of being confined by predefined templates, **Data-Driven Guides** (DDG) allows designers to generate guides from data and use the guides to place and measure custom shapes accurately [1]. Following the principles of information encoding, the guides can encode three main visual channels, length, area, and position, that can be combined to create a variety of visual structures and map these structures to data (Figure 1). When a user updates the data, DDG uses a deformation technique to transform custom shapes using the guides as the backbone of the shapes.

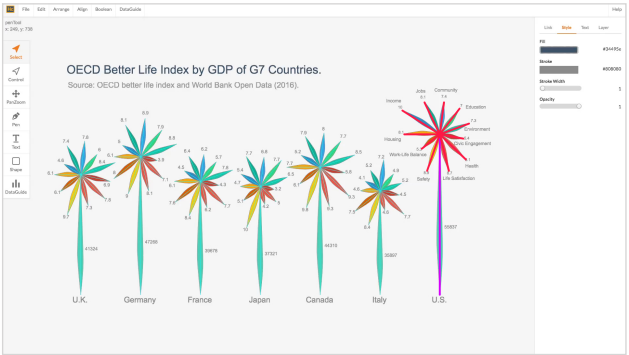


**Figure 1**: A vector graphics editor integrated with DDG, showing a flower chart that users a parent DDG (purple) to encode stems and child DDG (red) to encode flowers.

DDG brings data abstraction into a designer's familiar authoring environment and addresses tedious manual encoding and absence of data binding in the custom design practice. In the evaluation study, users commented that it is "incredibly useful tool for making infographics" and "much easier to try things out and experiment with the graphics." DDG received recognition in the Rising Star category for the Kantar Information is Beautiful Awards [7].

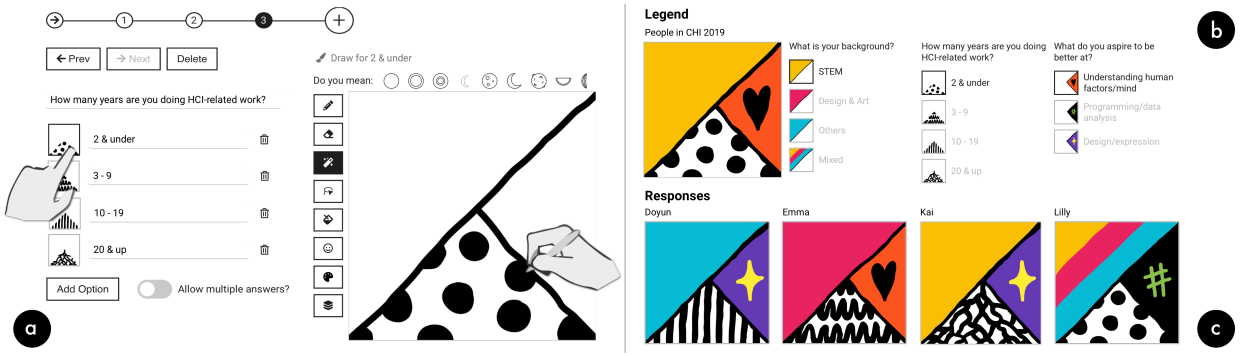


**Figure 2**. The user interface of DataSelfie: (a) user can draw a unique personalized visual for each option in the questionnaire, (b) an interactive legend aids the interpretation of visual mappings, and (c) each form response generates a distinctive visual.

Visualization also often serve casual and personal purposes. Individuals collect data about themselves, analyze the data, and share the visuals to promote self-knowledge and encourage conversation with others. However, most of the existing personal informatics systems 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.

To bring the expressivity of DDG into a personal context, I designed **DataSelfie** [2] by augmenting 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 qualitative and nuanced aspects of personal data (Figure 2). DataSelfie introduces a novel generative framework for producing a visualization, compared to descriptive approaches in existing tools including DDG. As a result, 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.

## TELLING ENGAGING STORIES WITH DATA

Journalists (e.g., from the New York Times and the Guardian), as well as designers, analysts, and casual users, are increasingly using visualization as a storytelling medium to convey messages about data. Most visualization tools like DDG and DataSelfie allow for the production of one visualization at a time. While these tools may be sufficient for conveying simple messages about the data, they cannot support the design of a fuller narrative and thus their ability to produce a comprehensive story is limited.

To address this need, I developed **DataToon** [3], a data storytelling tool for drawing data comics with pen + touch interaction (Figure 3). DataToon focuses on dynamic networks, one of the challenging forms of data from a storytelling standpoint, due to its evolving nature making it difficult to communicate with a single visual representation. DataToon leverages the well-established visual language of comics to tackle

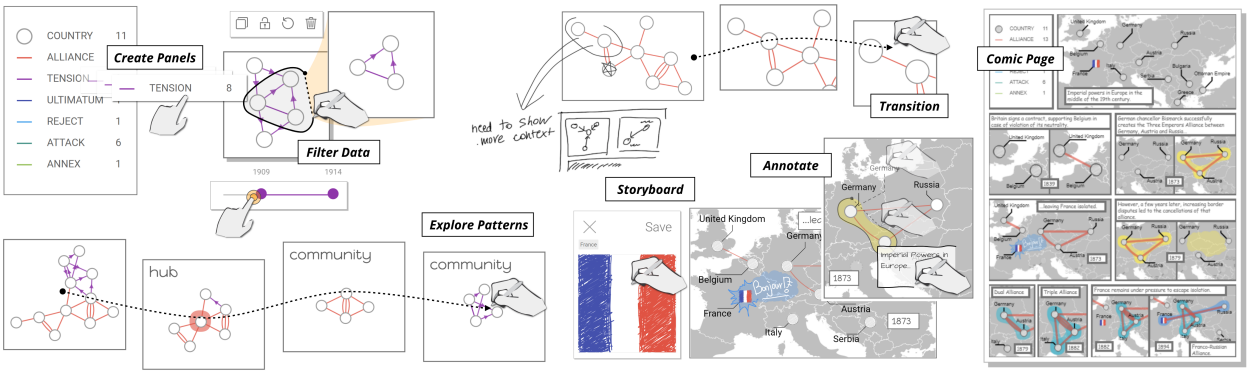


**Figure 3**. DataToon provides a pen & touch environment for producing data comics. A storyteller can rapidly isolate aspects of their data via filtering and pattern detection, as well as assemble a rich narrative via annotation and automatic panel transitions.

the challenge, by integrating captions and annotations with visualization and suppressing the complexity of data by incrementally revealing aspects of the data across multiple panels, arranged in space.

DataToon introduces a new interaction model that supports the full lifecycle of storytelling process from exploring data to discover insights, turning these insights into a narrative, and presenting and sharing the narrative with others. To facilitate rapid data exploration, it suggests recommendations of interesting patterns in the data. To allow for rapid iterations of alternative narrative structures, it leverages direct manipulation to arrange and style panels, and automatic transitions between the panels.

Crafting a story is a creative, open-ended process that involves trial and error. Being able to quickly grasp the overall narrative structure can accelerate this iterative process. I designed **Story Curves** [4], a novel technique that visualizes the narrative structure of a story. It reveals nonlinear narrative patterns by showing the order in which events are told and comparing them to their actual chronological order (Figure 4).

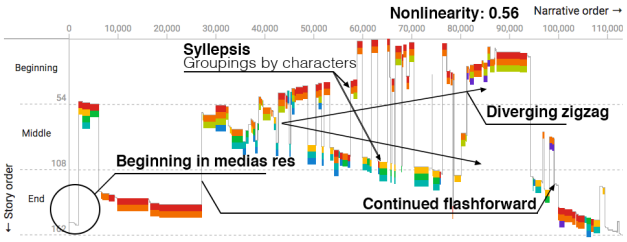


**Figure 4**: A story curve of the movie *Eternal Sunshine of the Spotless of Mind*, showing two diverging narratives within.

The analysis of movie scripts using Story Curves unveiled new narrative patterns, e.g., merging or diverging zigzags, that have not been discovered in the literature. Story Curves has been listed as part of the "Best of the Visualization Web" for the month of October 2017 [8].

## UNDERSTANDING COGNITIVE ASPECTS OF VISUAL DATA STORIES

As visualizations are more frequently used for communication, higher-order cognitive measures such as memorability, engagement, and aesthetics have become essential factors for visualization design. However, most prior research has been focused on the low-level perceptual processing of visual encodings, and as a result, lack of experimental research exists to understand these cognitive dimensions.

To study how people comprehend and recall messages in visualizations, we conducted an eye-tracking experiment investigating what elements of a visualization attract people's attention, and what information is encoded into memory [5]. The findings from the study unveiled several design guidelines, including that 1) titles plays an important role and should convey the message of a visualization, 2) pictograms. If used appropriately, can improve recognition, and 3) redundant encoding helps communicate the message.

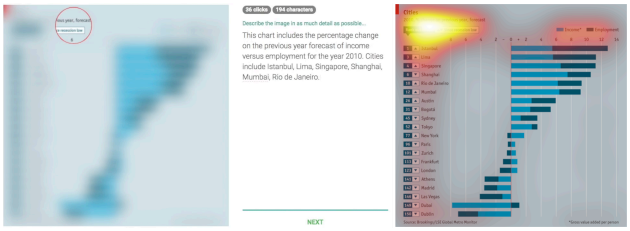

**Figure 5**: The BubbleView interface to collect clicks on an image (left). An importance map generated from the clicks (right).

Although eye-tracking has been proven useful for studying the cognitive processes involved in visual information processing, including which visual elements people look at first and spend the most time on, they often require high-cost equipment and tedious calibrations and thus are difficult to scale to large scale studies beyond controlled lab environments.

To overcome this issue, I developed **BubbleView** [6], a cursor-based, moving-window approach to collect clicks on static images as a proxy for eye fixations (Figure 5). BubbleView blurred images and allows participants to click around to reveal small circular "bubble" regions of the image at the original resolution. BubbleView can provide important insights about human perception and cognition, but at a lower data collection cost than eye tracking. It can easily scale up data collection to many participants and images, and be launched remotely to enable online crowdsourcing. To demonstrate this, based on the data collected using BubbleView, we built a neural network model for predicting the visual importance of elements in a visualization as well as graphic design image [9].

## FUTURE RESEARCH AGENDA: HUMAN-DATA INTERACTION

My research agenda comprises four main themes: 1) deepening our understanding of design practice and underlying mechanisms of visual data stories, 2) leveraging the knowledge gained to design better tools for storytellers and readers, 3) using the tools to promote communication in data science, and 4) improving data and graph literacy of the public to close the loop.

**Science of Data Storytelling.** Even with the widespread adoption of visualizations for data stories, we still do not have much knowledge about what makes an effective data story and how visualizations contribute to it. My approach to this research direction is two-fold: 1) leveraging the well-established, but currently not leveraged, narrative theories in humanities and computational narratives, and 2) building an online platform to crowdsource visualization experiments (e.g., using BubbleView), in which users test their own hypothesis to extract design guidelines.

**Tools for Thought.** Storytelling requires a complex set of skills but not all people have these skills. I plan to explore how to design intelligent interactive tools that amplify human creativity and cognition in order to assist with authoring and reading effective visual data stories. The automatic transition and suggestion of panels in DataToon and autocomplete of drawing in DataSelfie are a step toward this direction but how can we bring further narrative design guidance into the authoring process without hampering the creativity of storytellers? In addition, I also aim to improve the reading experience of data stories by pushing the boundaries of traditional storytelling genres (e.g., transmedia and immersive storytelling).

**Promoting Communication in Data Science**: Data science is an emerging field whose results have important implications of our lives. However, much of the efforts is currently geared towards on exploration and analysis of data. For instance, a recent study of 200 academic computational notebooks, a popular medium for documenting and sharing exploratory analysis, revealed that few describe analytical reasoning and results of the analysis [10]. I plan to investigate ways to bridge the gap between analysis and communication in the data science practice (e.g., providing structural support to scaffold a narrative by linking charts with explanations).

**Improving Graph Literacy of the Public.** Over the past decade, the visualization research community have developed a variety of new visualization techniques and design guidelines beyond simple charts and perceptual rankings of visual variables. However, most people still do not keep up with such recent advancements. The current approach is to develop passive, instructional guidelines. This usually works for people with enough data literacy like analysts but not for novices such as students and children. I envision a framework that takes the constructionism approach to allow learners to actively take part in the learning process (e.g., deconstructing or constructing graphs by themselves).

## REFERENCE


1. **Nam Wook Kim**, Eston Schweickart, Zhicheng Liu, Mira Dontcheva, Wilmot Li, Jovan Popovic, Hanspeter Pfister. Data-Driven Guides: Supporting Expressive Design for Information Graphics. IEEE VIS (InfoVis) 2017.
2. **Nam Wook Kim,** Hyejin Im, Nathalie Henry Riche, Alicia Wang, Krzysztof Z. Gajos, Hanspeter Pfister. DataSelfie: Empowering People to Design Personalized Visuals to Represent Their Data. ACM CHI 2019 (under review).
3. **Nam Wook Kim**, Nathalie Henry Riche, Benjamin Bach, Guanpeng Xu, Matthew Brehmer, Ken Hinckley, Michel Pahud, Haijun Xia, Michael McGuffin, Hanspeter Pfister. DataToon: Drawing Dynamic Network Comics With Pen + Touch Interaction. ACM CHI 2019 (under review).
4. **Nam Wook Kim**, Benjamin Bach, Hyejin Im, Sasha Schriber, Markus Gross, Hanspeter Pfister. Visualizing Nonlinear Narratives with Story Curves. IEEE VIS (InfoVis) 2018.
5. Michelle A. Borkin*, Zoya Bylinskii*, **Nam Wook Kim**, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. Beyond Memorability: Visualization Recognition and Recall. IEEE VIS 2016.
6. **Nam Wook Kim***, Zoya Bylinskii*, Michelle A. Borkin, Krzysztof Z. Gajos, Aude Oliva, Fredo Durand, Hanspeter Pfister. BubbleView: an interface for crowdsourcing image importance maps and tracking visual attention. ACM TOCHI 2017.
7. Kantar Information is Beautiful Awards, November 2017. https://tinyurl.com/kanter2017
8. Best of the Visualization Web. October 2017. https://tinyurl.com/bestvis2017
9. Zoya Bylinskii, **Nam Wook Kim**, Peter O'Donovan, Sami Alsheikh, Spandan Madan, Hanspeter Pfister, Fredo Durand, Bryan Russell, Aaron Hertzmann. Learning Visual Importance for Graphic Designs and Data Visualizations. ACM UIST 2017.
10. Adam Rule, Aurélien Tabard, James Hollan. Exploration and Explanation in Computational Notebooks. ACM CHI 2018.