Data Sketches: An Exploratory Study

Jagoda Walny* and Sheelagh Carpendale†

University of Calgary

ABSTRACT

Hand-drawn sketches, such as those on napkins and whiteboards, are commonly used as part of everyday thinking processes. These types of sketches can be thought of as spontaneously created representations of ideas and concepts, but can also be applied to data. Our question is whether a better understanding of sketched data representations can contribute a new perspective on our visual representation repertoire. To explore this, we conducted a qualitative study in which we asked people to manually sketch representations of a small, understandable dataset and to report on what they learned or found interesting about the data.

Index Terms: *Qualitative evaluation, data transformation and representation, sketching*

1 INTRODUCTION

Writings abound about how readily people use sketches during thinking, e.g. [3,7]. Using hand-drawn sketches and diagrams is known to be effective in promoting innovation, creativity, and thinking in general [1]. However, the literature that praises people's general facility with using sketching for ideation does not talk about sketching data. Given the increased need to think about data and data visualizations, we ask: Is sketching data more difficult than sketching ideas? Is there a relationship between sketching data and people's understanding of that data?

Our goal is to develop a deeper understanding of the answers to these questions with a closer look at data sketching. Here, instead of creating rapid pen-based interfaces for visualization (e.g. [2]), we describe an exploratory study of manually sketched data representations, summarize our classification of these sketches, and discuss select hypotheses that emerge from our study results.

2 DATA SKETCHING STUDY

We asked people to sketch a visual representation of a small dataset and briefly tell us what they learned or found interesting about the data. We ran three formal sessions with 7, 8, and 7 participants in each, for a total of 22 unique participants (13 male, 9 female). All participants had completed some post-secondary education: 6 had a Bachelor's degree; 13 a Master's degree; and 2 a doctorate. Eighteen participants had education in computer-related fields such as computer science or software engineering; of those, 3 also reported experience in graphics or visualization, and 1 in business. Additionally, we had one participant each in degrees related to design, communication-illustration, kinesiology and languages, and one who did not report this information.

2.1 Set-up, Materials, and Dataset

We used a simple classroom with ample tables and chairs and good lighting. We provided blank sheets of standard letter-size paper (participants could use as many as they liked), a variety of colored pencils, and a printout of the dataset to each participant.

Our dataset was chosen to be engaging and easily understandable without need of specialized knowledge. It contains mean appropriateness ratings (ranging from 0 to 9) of 15 behaviors in 15 different social situations extracted from a 1974 social psychology study [4]. We consider our participants experts in understanding this data, not as social psychologists, but because it describes common everyday situations. Thus we were able to gather manually sketched data representations from a group of experts with an understanding of the data and domain.

2.2 Procedure

Each session of the study was administered in a group setting. We scripted the explanation of the dataset and the data sketching rules to ensure consistency across sessions. We told participants to represent the data on the blank paper in any way they chose and that there was no right or wrong method. We encouraged them to draw the data as they explored it and to think about:

- Connections between different pieces of data
- Ways of grouping the data
- Similarities and differences in the data
- Interesting patterns and surprising findings in the data

We planned answers to common questions to avoid influencing the representations. For instance, if asked how to draw something, we would answer, "any way that makes sense to you". Talking was permitted, but tended to be minimal. When a participant handed in a sketch, or after 45 minutes, we administered a simple questionnaire asking for basic demographic information and one important question with a full page for possible answers, worded as follows: *Please describe what you learned or found interesting about this data during the session (there are no wrong answers)*.

2.3 Analysis

We used a qualitative analysis approach. Using careful examination and affinity diagramming, we categorized the sketches, working together until agreement was reached. We analyzed individual statements made in the responses to the questionnaire using an open coding approach [6]. One author performed open coding; both authors performed focused coding independently, then discussed each coding instance until consensus was reached. Lastly, we used a matrix to relate the representation types to the spectrum of reported knowledge.

3 RESULTS AND DISCUSSION

We collected a total of 35 visual representations, with as many as four coming from a single participant. In our analysis, the most important factors in the sketched data representations were the uniqueness of each representation and their wide variety.

3.1 Classifying the Representations

Our affinity diagramming approach yielded a classification of the representations along a spectrum with purely numeric representations of the data at one end, more abstract representations in the middle, and pictorial representations at the other end.

*email: jkwalny@ucalgary.ca
†email: sheelagh@ucalgary.ca

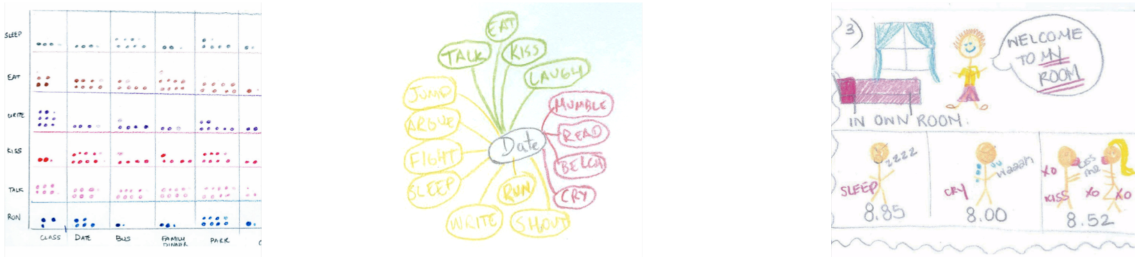

Figure 1: A numeric matrix representation, an abstract graph-like representation, and a pictorial representation of the same dataset.

We now briefly describe each portion of this spectrum:

Numeric representations included those representations from which the original numeric data was readily retrievable (to the degree possible in a hand-drawn representation). This included: matrices and countable tables (7), dot / scatter plots (1), bar charts (13), line charts (1) and coordinate plots (2). (Note that some sketches included more than one representation type; e.g. one chart included lines and bars).

Abstract representations were those less focused on direct representations of numeric values, predominantly incorporating some grouping or binning scheme. This included graphs or graph-like representations (3), ranked lists (2) and Venn diagrams (2).

Storytelling and pictorial representations used pictorial icons such as stick figures and other line drawings to present the data as a story. We collected four such representations, all unique. For example, one representation used icons and line drawings to depict a person moving through various everyday situations and used an accompanying visualization to act as a decision support tool to guide appropriate behavior choices. Another sketch includes a descriptive scene of a setting, icons explaining appropriate behaviors, and is annotated with words, sound effects, and the actual numerical values from the data (see Fig. 1). These sketches show a capability of seeing data as a story, which is interesting in light of recent work on narrative visualization [5].

3.2 Classifying Reported Knowledge

Participants' questionnaire responses provided a glimpse into the knowledge that participants gained about the data. We categorized each statement in these responses on a spectrum starting from low-level single data value retrieval to relatively high-level conjectures and fledgling hypotheses. On one end we have statements with ***information intrinsic to the dataset***, including those referring to specific behavior-situation pairs, low-level summaries about individual situations or behaviors, and low-level pairwise comparisons between behaviors or situations. Next are statements about ***comparisons and trends within the dataset***, noting trends across three or more rows or columns, classifying or grouping items by value, or making global comparisons. Following are statements with ***information extrinsic to the dataset***, classifying values, rows, or columns using named concepts (e.g. "safe", "aggressive", "work-related"); comparing the data against pre-existing expectations; or explaining the data in the domain context (e.g. "*people care a lot in job interviews*"). Lastly, we found statements with ***analytic potential***, in which participants asked questions or made conjectures about the data. For example, one participant hypothesized that *park* and *own room* have similar values due to their relative anonymity. Another speculated that there were more females than males in the initial study due to the approval rating of talking in the bathroom.

3.3 Reported Knowledge vs. Representation

Nine of our participants (41%) reported statements with analytic potential. Of these, one participant drew a numeric representation, one drew a representation on the border of numeric and abstract, three participants drew abstract representations, and four participants drew pictorial / storytelling representations. In fact, all participants who drew pictorial representations also provided statements with analytic potential. This last point surprised us because, at first glance, these representations seemed furthest removed from the actual data and from the numerical patterns in this data. However, with only one exception, the participants who drew abstract and pictorial representations provided us with more analytic statements than those who drew strictly numeric representations. Due to the nature of this study, it is impossible to tell whether the depth of thought shown by these participants was a result of the representations they drew, the process of drawing, their engagement with this particular dataset, their own analytic skills, or some combination of these. Nevertheless, these are interesting observations that bear further investigation.

4 Conclusion

These select results from our exploratory study of data sketching raise some interesting questions ripe for further study. The variety of representations we collected reinforces the myriad representational strategies available. The apparent relationship between abstractness of representation and the analytic potential of the reported knowledge about the data indicates interesting potential avenues for research: investigating whether there is some previously overlooked relationship between story-based visualizations and numerical visualizations; and investigating if the limited expressiveness of pre-defined visualizations also limits creative thinkers' abilities to express their thinking about data.

Acknowledgments

This research was supported in part by NSERC, AITF, Surfnet, GRAND, and SMART Technologies.

References

[1] B. Buxton, *Sketching user experiences: Getting the design right and the right design*. Morgan Kaufmann, 2007.
[2] W. O. Chao, T. Munzner, and M. van de Panne, "Rapid Pen-Centric Authoring of Improvisational Visualizations with NapkinVis," *Posters Compendium InfoVis*, 2010.
[3] D. Kirsh, "Thinking with external representations," *AI & Society*, vol. 25, no. 4, pp.441–454, 2010.
[4] R. H. Price and D. L. Bouffard, "Behavioral appropriateness and situational constraint as dimensions of social behavior," *Journal of Personality and Social Psychology,* vol. 30, no. 4, p. 579, 1974.
[5] E. Segel, J. Heer, "Narrative visualization: Telling stories with data," *IEEE Trans. Visual. Comput. Graphics*, vol. 16, pp. 1139–1148, 2010.
[6] A. L. Strauss and J. Corbin, *Basics of qualitative research: Grounded theory procedures and techniques*, Sage Publications, 3rd ed., 2008.
[7] B. Tversky, "Visualizing thought," *Topics in Cognitive Science*, vol. 3, no. 3, pp. 499–535, 2011.