Adapting Reorderable Matrices for Qualitative Analysis

Melissa Birchfield* and Eunice Jun*
University of Washington

ABSTRACT

We summarize and reflect on two case studies of projects that involved adapting Bertin's reorderable matrices as an interpretative aid during qualitative analysis. These case studies reveal three sites of interaction: between researcher and reader, researcher and data, and researcher and researcher. We find that in both case studies, interactive visualizations highlight the tensions that arise from viewing interaction with data as an individualistic vs. collective practice. In this tension, we identify future work opportunities for using reorderable matrices during qualitative analysis and designing for collective data interactions.

1 INTRODUCTION

As visualization designers, we focus primarily on using visualizations as a communication tool from researcher to the greater public. Yet, interacting with data is integral for constructing meaning from the data and figuring out what to communicate in the first place. Indeed, in La Graphique [1], Jacques Bertin argues for the importance of interaction in visualization for the purpose of sensemaking. In the text, Bertin describes a now prevalent technique for interacting with data and continually extracting insights from data through interaction: **reorderable matrices**. Although reorderable matrices are primarily taught as a technique for representing and manipulating quantitative data, we adapted the technique to aid interpretation during qualitative analysis. In doing so, we have become more acutely aware of two additional sites of interaction and potential future areas of focus for human-data interaction: between researcher and data (researcher-data) and researcher to researcher (researcher-researcher). Below, we summarize two experiences of using reorderable matrices as a tool in analyzing interview notes and published research papers, focusing on the types of interaction in each and postulating how our processes would have differed without interactions with the reorderable matrices.

2 CASE STUDY 1: REORDERABLE MATRICES TO INTERPRET SEMI-STRUCTURED INTERVIEW TRANSCRIPTS

In order to understand the role of data in cross-sector collaborations to combat human trafficking, the first author conducted semi-structured interviews with 12 anti-trafficking stakeholders across four sectors (i.e., victim service providers, funding agencies, law enforcement, and legislation) in Washington State. To convey her findings, she constructed a reorderable matrix for each research question. As shown in Figure 1, the rows represent themes surfaced in interviews and the columns represent each interview participant. Colored cells denote whether a participant mentioned a particular theme, with each stakeholder group assigned a different color. The first author initially intended these matrices to communicate between **researcher and reader**, strengthening confidence in her findings by demonstrating the number of participants who mentioned certain themes.

Beyond this original purpose, however, the matrix creation process proved valuable for another site of interaction: enhancing the researcher's engagement with the qualitative data itself **(researcher-data)**. The ability to see all interview transcripts distilled into themes across stakeholder groups in a *single view* allowed the first author to identify meaningful patterns from color and space that she would have otherwise missed with a strictly text-based qualitative analysis. Based on similar color fill patterns among rows, she identified redundancy in closely related themes and accordingly consolidated codes or called out the correlation between these themes. She also noticed themes mentioned much more frequently within one stakeholder group over another, which furthered her research goals of understanding how each group's specific perspective helps or hinders their cross-sector collaboration. Although the first author was the only *individual* conducting the qualitative analysis, thus precluding researcher-researcher interaction as part of the reorderable matrix approach, the process of creating and comparing matrix rows/columns nevertheless enriched the analysis by shifting the researcher's physical view to enable extraction of meaning from the data.

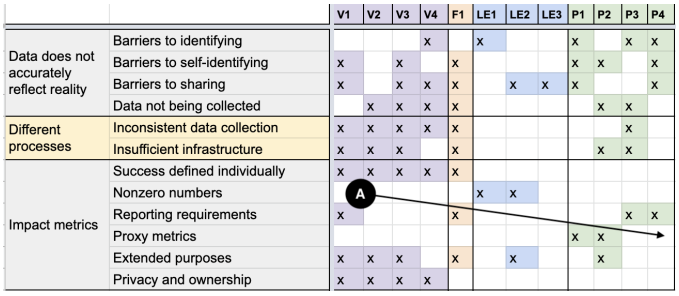

Figure 1: **Thematic analysis of semi-structured interviews.** For each research question, the first author visualized relevant interview content as a "reorderable matrix" [1] to help consolidate codes and identify patterns among themes. This matrix summarizes findings to answer one research question. The rows represent themes surfaced by at least one of the 12 participants, which are represented by columns colored and grouped by sector (V = Victim Service Provider, F = Funder, LE = Law Enforcement, P = Policymaker). A cell marked with an "X" denotes that the given participant mentioned that particular theme during the interview. Based on this matrix, the first author noticed that impact metrics varied by stakeholder group. By rearranging row and column groups, she further observed that the impact metrics became increasingly detached in correlation with a stakeholder's proximity to victims and survivors in their line of work. (A) For example, victim service providers work mostly closely with survivors and view "success" in terms of self-determined goals for each individual client. Policymakers, on the other hand, depend more on proxy metrics, such as the number of hotline calls, as a way to evaluate the impact of their anti-trafficking policies.

3 CASE STUDY 2: REORDERABLE MATRICES TO INTERPRET PUBLICATIONS

In order to identify how researchers translate their domain hypotheses into statistical models, we collected and conducted a content analysis of a corpus of 50 research papers from five different venues including research papers in medicine, psychology, economics, and human-computer interaction (see [2] for more details). As part of our qualitative analysis procedure, we constructed a reorderable matrix for each paper. As shown in Figure 2, the rows represent the codes in our codebook and the columns represent the paragraphs in each paper, fixed in chronological order from left to right. We colored instances of the codes in each paragraph according to the broad categories of codes that comprised our hierarchical codebook.

*equal contributions; mbirch2@uw.edu, emjun@cs.washington.edu

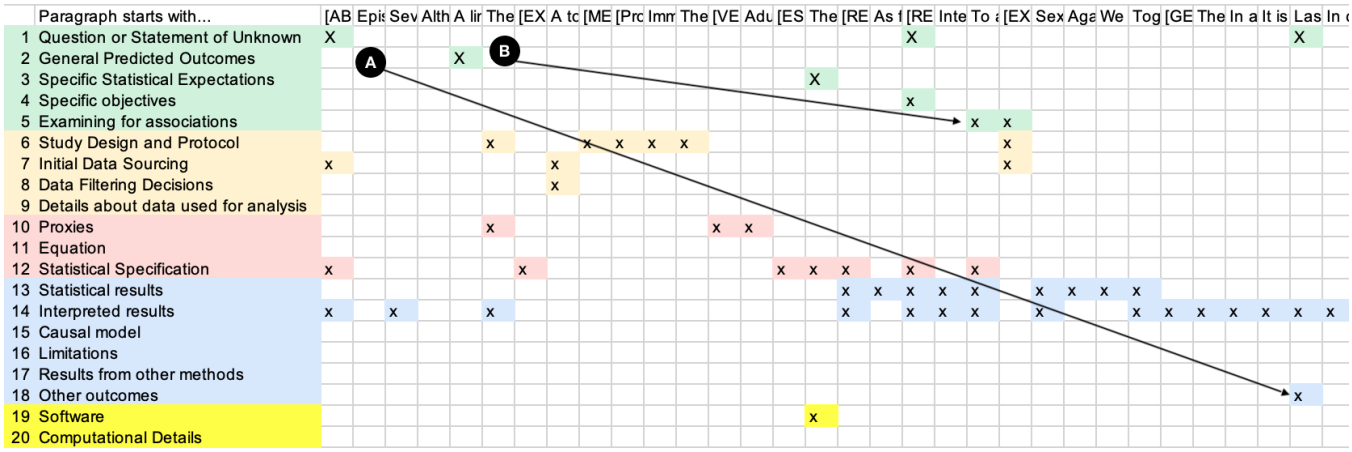

Figure 2: **Content analysis of research publication [3].**

Reorderable matrices, such as the above, helped us detecting patterns in publications' structure and content. The rows represent the codes in our codebook, colored according to the five broad categories of codes: research goals (rows 1-5, green), sample information (rows 6-9, yellow), statistical analysis details (rows 10-12, red), reporting of results (rows 13-18, blue), and computational details (rows 19-20, yellow). The columns are the paragraphs, which are indexed by their first sentences, ordered left to right. In a paragraph's column, there is an "X" for each code the paragraph received. Paragraphs have multiple codes if they contain multiple types of information. Among the ten visual patterns we noticed across our sample and subsequently looked for in each paper, two stand out in this paper. (A) As the paper progresses (visually moving left to right), the paper's focus shifts from research goals to sample information to statistical analysis to results, as indicated by the arrow labeled A. Largely expected, this pattern helps to validate our coding method. Also, there is only one paragraph that discusses statistical software. (B) Researchers discuss research goals and questions throughout the paper. Interestingly, in the middle of the paper, when the researchers discuss their goals in greater detail, the researchers discuss them in increasing specificity, as indicated by the arrow labeled B. We were able to detect this pattern across papers by iterating on how to order the research goal codes (rows 1-5, green) and finally ordering the codes by increasing specificity from top (row 1) to bottom (row 5).

In doing so, we noticed places where codes were under-specified and led to consistent duplicate codes for several paragraphs. This prompted us to revise our codebook and iteratively re-code the papers in our corpus. Additionally, by shuffling the order of codes in the matrices and transposing the matrices, we noticed that some of the codes in our categories could be ordered from broader to more specific statements of research questions, for example, which helped us notice more nuanced patterns of content in the papers.

Interaction with the papers represented as reorderable matrices was central to helping both authors *individually* interpret the data. We found that we each had idiosyncratic ways of interacting with the matrices **(researcher-data)**. For example, we shuffled the order of codes from paper to paper differently in order to do deep dives in a single paper. To compare across papers though, we had to come to a consensus about the and impose a standard order of codes. In hindsight, what became apparent in our discussions around the reorderable matrices was that we had each attuned to different codes, visual patterns, and interactions with the reorderable matrices. In some ways, we were leaving traces of our philosophical or personal biases in our interaction traces with the reorderable matrices. Although the motivation behind constructing reorderable matrices for the papers was to leverage our visual perception to identify any patterns in paper structure and content, we found that another role the visualizations had was in helping to ground our interactions with one another and to more clearly articulate and understand each other's interpretative stance **(researcher-researcher)**. In other words, interacting with the data and visualizations provided context and began to represent the *collective* understanding we constructed.

4 Provocations for Future Work

Taken together, the two case studies of adapting reorderable matrices to aid qualitative analysis show the usefulness of reorderable matrices *across* sites of interaction, especially researcher-to-researcher. In the first case study, the main site of interaction was between the **researcher and data**. Even here, the reorderable matrices facilitated rich engagement with the data, albeit just with an *individual* researcher. In the second case study, the reorderable matrices went a step further to facilitate *collective* sense-making of the data and our interpretative stances **(researcher-researcher)**. The collective aspect of the analysis enabled the researchers to see the data from multiple perspectives, how the perspectives aligned and differed, and areas of collective oversight or bias. Moving from a focus on interacting *with* data to interacting *through* data, our experiences suggest two concrete ways that reorderable matrices could improve collaborative sense-making during qualitative analyses.

First, we found that some words or phrases fell more strongly under a given code than others that we also ultimately assigned to the same code. Being able to attach an additional dimension to each cell (e.g., gradient/saturation) for strength of assignment and/or linking each cell to the text that supports its decision would help convey an added layer of nuance during a thematic or content analysis, as well as provide a quick reference to the text for iterating on codes with greater uncertainty. This would be especially useful for interaction between researchers, where collective analysis may require substantial communication.

Second, we recognized our desire to "force" patterns that may have been tenuously supported in the text; on the other hand, perhaps we missed patterns that we did not anticipate and thus did not actively search for. Although the reorderable matrices helped us engage with the data and recognize patterns in a new way, we also acknowledge that our own expectations and biases could have constrained the extent to which these interactive visualizations expanded our understanding of the data. Tools for detecting and identifying patterns prevalent across texts could help analysts reflect on their biases and how they overlap more explicitly.

Beyond reorderable matrices, we ask *How could HDI acknowledge and facilitate collective sense-making?*

5 Conclusion

Using the reorderable matrices to interpret, and not just summarize the qualitative data, helped us illuminate three sites of interaction: reorderable matrices from researcher to reader, between researcher and data, and researcher to researcher. These sites of interaction serve as intervention points where the human-data interaction community could focus beyond individual interactions with data to acknowledge the impact of interacting with data as a collective.

ACKNOWLEDGMENTS

We thank Nicole de Moura, Jeffrey Heer, and Rene Just for their co-authorship and conversations throughout the project [2] that the second case study summarizes. We thank Rene Just for his advice and insight during the first case study. We thank Daniela Rosner for conversations about the individualistic and collectivistic values in design. Critical Fabulations [4] influenced our reflections.

REFERENCES

[1] J. Bertin. *Graphics and graphic information processing*. de Gruyter, 2011.

[2] E. Jun, M. Birchfield, N. De Moura, J. Heer, and R. Just. Hypothesis formalization: Empirical findings, software limitations, and design implications. 2021.

[3] C. T. Ngo, A. J. Horner, N. S. Newcombe, and I. R. Olson. Development of holistic episodic recollection. *Psychological Science*, 30(12):1696–1706, 2019.

[4] D. K. Rosner. *Critical fabulations: Reworking the methods and margins of design*. MIT Press, 2018.