Analyzing Qualitative Data

Sheelagh Carpendale
University of Calgary
Calgary, AB, Canada
sheelagh@ucalgary.ca

Uta Hinrichs
The University of St Andrews
Fife, Scotland
uh3@st-andrews.ac.uk

Søren Knudsen
University of Calgary
Calgary, AB, Canada
sknudsen@ucalgary.ca

Alice Thudt
University of Calgary
Calgary, AB, Canada
athudt@ucalgary.ca

ISS '17, October 17–20, 2017, Brighton, United Kingdom

ACM ISBN 978-1-4503-4691-7/17/10.
https://doi.org/10.1145/3132272.3135087

Abstract
Evaluation is increasingly recognized as an essential component of HCI research. However, evaluation itself is a changing research area. In particular, the many variations of qualitative research are emerging as important empirical methods. This half-day tutorial is designed for beginning to intermediate audiences. We will focus on the basic methods for analyzing qualitative data using a mixture of talks and hands-on activities. In particular we will consider closed and open coding as well as clustering and categorizing coded data. After completing this tutorial, attendees will have a richer understanding of the benefits and challenges of qualitative empirical research and, more specifically, how to analyze qualitative data.

Author Keywords
Evaluation; qualitative studies; data analysis; coding.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; H.5.2. Information interfaces and presentation (e.g., HCI): Evaluation/methodology.

Introduction
We are increasingly interested in evaluation [4,10,12,15,16,17,17]. However, evaluation is a complex multi-faceted process that involves many skills

Topic Overview

Focus on how to analyze carefully collected rich qualitative data. We consider Bryman's four stages of qualitative analysis:

Stage 1: Looking for the ideas that emerge from your data

Stage 2: Identifying codes and creating a coding schema.

Stage 3: Coding

Stage 4: Relation to existing theories and ideas.

[13]. In a previous tutorial, we provided an overview of qualitative evaluation through talks, discussions and hands-on exercises from the perspective of gathering qualitative data[1]. In this previous tutorial we focused on observation and interviewing as qualitative evaluation data *collection* skills. In this tutorial[2], we will focus on the *analysis* of qualitative data, including the analysis of interview transcripts and video data, drawing from previous work [3,9] and our own experience [8,11,14]. The tutorial will introduce participants to the concepts of grounded theory [5,7] and thematic analysis [2]. We will use intermingled talks, discussions and hands-on exercises focusing on closed and open coding, as well as clustering and categorization. In general, we will focus on qualitative analysis methods, providing some insight into their benefits, exploring what "rigor" in qualitative research can mean, and offering some hands-on activities where people will be able to develop some qualitative evaluation skills.

Topics and objectives of the tutorial

In this tutorial we focus on how to proceed once you have carefully collected your fabulously rich qualitative data. As a good basis from which to start we will consider Bryman's four stages of qualitative analysis [3].

Stage 1: Looking for ideas within your data

Once your qualitative data has been collected, the first process is to decide upon your coding focus. A usual first step is to read the whole text (e.g., interview transcript), or to watch the whole video. Preferably this is initially done without interruptions, that is, activities such as taking notes that stop train of thought are done later. In this first pass; one is looking for a general impression, for what the transcript or video recording is really about. In this stage, the intention is to identify major themes with an open mind for surprises, unexpected or unusual factors. After reading, such themes and unexpected aspects can be written down along with other ideas from which to analyze the text or transcript.

Stage 2: Identify codes and create a schema.

In this stage one is working towards developing an initial set of codes or a schema, which one will use to code the data. Generally this work is done with a subset of the data. The selected subset is thoroughly examined. Different people use different techniques, many of which are akin to close reading of text. For example, one might use highlighting, underlining, adding comments, and marginalia. The purpose is to identify a group of factors that are definable, recognizable, and, separately or in combination are of interest to the research questions. Having identified a list of factors, the usual process is to characterize them with recognizable names and clear definitions.

Stage 3: Coding

For the process of coding, one takes the identified and defined factors from Stage 2 and proceeds carefully and slowly through the text, interview transcript or video clip, marking or *coding* each occurrence of each factor identified and pre-defined in Stage 2. When using an *open coding* approach, one adds codes if something of interest or importance occurs in the data for which there is no code. Then one adds a code, informs other coders and goes back to the beginning to see if there are any missed instances of the new code. When all

[1] Find materials from last year's tutorial here: http://innovis.cpsc.ucalgary.ca/qualeval-vis-tutorial/

[2] Find materials for this tutorial here: http://innovis.cpsc.ucalgary.ca/qualitativeanalysis-iss-tutorial/

Activities Overview

Structure and Materials: We briefly introduce each exercise topic and conclude with group discussions of each. We provide data transcripts for each of the following exercise topics:

Coding Focus: We do a group idea generation session to consider what would be useful foci to code for.

Closed Coding: We discuss the closed coding process and provide a set of codes. Each participant will practice closed coding, followed by pair discussions of results.

Open Coding: We discuss the open coding process and provide an initial set of codes. Each participant will practice open coding, followed by pair discussions of results.

Clustering & Categorizing: Finally, we explore various forms of clustering and categorizing, to make sense of the coded data.

the data is coded, it is still important to work with the codes. Are there important clusters? Are there relationships within the codes? Is there an ordering? If one has coded for more than one major focus, what is the relationship between these coding passes? Sometimes it is important to keep track of variations in codes, which can add richness and depth. Coding is often a collaborative process conducted by several researchers who work with the same data. It involves frequently comparing how they have coded particular data snippets in order to verify and (if necessary) refine the coding scheme. This process also helps to minimize the introduction of personal biases that are inevitably introduced while coding and interpreting the collected data.

Stage 4: Relation to existing theories and ideas.
At the end of Stage 3, it is important to look outside of the current data and consider one's findings in relation to existing theories, and understandings.

Activities

Coding, clustering and categorizing are skills that can be learned and practiced. We will provide hands-on activities to let people gain experience in these skills.

Activity 1: Choosing a Coding Focus
In this activity, we ask participants to engage in Stage 1 as described above to learn techniques on how to develop a coding focus, that is identifying aspects of interest in the data and develop a corresponding coding scheme. We will introduce participants to aspects or open-ended questions that can drive this initial stage of qualitative data analysis. Workshop participants will be divided up into small groups and provided with a brief interview transcript or video clip. They will then go through this qualitative data individually, take notes of potential ideas, discuss these within their group. The activity will conclude with a discussion of the ideas developed by the different groups, focusing on similarities and differences in the emerging coding foci.

Activity 2: Closed Coding
In this activity we will introduce participants to closed coding, often also referred to as analysis with a priori codes [6]. A set of a priori codes can be derived from previous research and theory or directly from the evaluation questions driving the research. We will provide participants with qualitative data (e.g. an interview transcript or video snippet) alongside a coding schema. Each participant will use this schema to code the data individually. We will then compare and discuss results among participants. The activity will be concluded with a group discussion about possible variability in outcomes as well as advantages and limitations of analyses with a priori codes.

Activity 3: Open Coding
Participants will gain experience with open coding by working with an interview transcript that we provide. We may reuse the same transcripts as Activity 1 here, to ensure that participants are already familiar with the text. Participants will open-code the transcript individually, then reconcile their codes with a partner, then re-code the transcript together. Emphasis will be placed on the experience of iterative coding, where the transcript is reviewed and re-coded multiple times as the coding scheme evolves. Codes will be written on sticky notes to facilitate the next activity.

Activity 4: Clustering and Categorizing
In this activity, we ask participants cluster and categorize a set of codes. We do so, based on the open coding

Structure of Tutorial

Type of Activities: We will use a mix of short talks, hands-on activities, and group discussions.

Timeframe: We have planned exercises to cover a half-day tutorial. We allocate between 30 and 45 minutes for each activity, with time for a break midway through the tutorial.

Also at IEEE VIS 20217: We run a similar tutorial two weeks prior to ISS at IEEE VIS 2017. Thus, we plan this tutorial to follow a similar outline.

performed in the previous activity, which resulted in coded data. For the purpose of the activity, we will introduce a lightweight approach to clustering and categorizing coded data (e.g., affinity diagramming [1]). We will also discuss alternative in-depth approaches. In small groups, participants will collaboratively categorize and relate different codes and develop a structure of the relationship between codes on a shared medium.

Conclusions

From this tutorial, participants will learn more about the benefits, nuances and challenges of qualitative empirical research and qualitative data analysis in particular. They will have taken the first steps towards learning more from their interviews, and towards practicing and enhancing their qualitative data analysis skills.

Acknowledgments

This research was supported in part by: Alberta Innovates - Technology Futures (AITF); Natural Sciences and Engineering Research Council of Canada (NSERC); and SMART Technologies.

Tutorial organizers

Sheelagh Carpendale is a Professor in Computer Science at the University of Calgary where she holds a Canada Research Chair in InfoVis and NSERC / AITF / SMART Technologies IRC in Interactive Technologies. She has received many awards including the E.W.R. NSERC STEACIE Memorial Fellowship; a BAFTA (British Academy of Film & Television Arts Interactive Awards); an ASTech Innovations in Technology award; and the CHCCS Achievement Award, which is presented periodically to a Canadian researcher who has made a substantial contribution to the fields of computer graphics, visualization, or human-computer interaction. She leads the Innovations in Visualization (InnoVis) research group and initiated the interdisciplinary graduate program, Computational Media Design. Her research focuses on InfoVis visualization and large interactive displays. She both conducts and publishes about evaluation with a particular focus on qualitative evaluation.

Uta Hinrichs is a Lecturer at the University of St Andrews, Scotland, UK in the SACHI research group. Her research is at the intersection of HCI, visualization, design, the humanities, and art. Her work focuses on designing and studying the use and experience of interactive systems that facilitate the exploration and analysis of (cultural) data collections from academic, leisurely, and artistic perspectives. Studying the use of technology in-situ through qualitative research methods such as field observations, interviewing and video analysis is core to her research. Uta holds a PhD in Computational Media Design from the University of Calgary.

Søren Knudsen is a Postdoctoral Fellow in the InnoVis group at the Interactions Lab at the University of Calgary. He holds a PhD in Computer Science from University of Copenhagen. His research focuses on HCI, large interactive displays, and InfoVis. He is interested in studying technologies in-situ and in bringing parts of reality into lab contexts. He uses a mix of qualitative and quantitative methodology in his approach, and study visualization problems as they occur within and across a range of application domains.

Alice Thudt is a PhD student in Computational Media Design at the University of Calgary. She is interested in how visualization of personal data can support self-reflection and expression. Her research aims to understand how people construct meaning with personal digi-

tal data collections and how both digital and physical visualization can be used for personal storytelling and reminiscing. She has used different qualitative research and analysis methods in her research ranging from observations and interviews to variations of a technology probe method.

References

1. Hugh Beyer and Karen Holtzblatt. 1997. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann.
2. Richard E. Boyatzis. 1998. *Transforming Qualitative Information: Thematic Analysis and Code Development*. SAGE.
3. Alan Bryman. 2001. *Social Research Methods*. Oxford University Press.
4. Sheelagh Carpendale. 2008. Evaluating Information Visualizations. In *Information Visualization*. Springer Berlin Heidelberg, 19–45.
5. Juliet Corbin and Anselm Strauss. 1998. *Basics of qualitative research*. SAGE Publications, Thousand Oaks, CA.
6. Benjamin F. Crabtree and William L. Miller. 1999. *Doing Qualitative Research*. SAGE Publications.
7. John W Creswell. 2007. Qualitative enquiry and research design: Choosing among five approaches. *US: Sage publications Ltd*.
8. 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.
9. Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in Qualitative Research*. SAGE Publications.
10. Trevor Hogan, Uta Hinrichs, and Eva Hornecker. 2015. The Elicitation Interview Technique: Capturing People's Experiences of Data Representations. *IEEE Transactions on Visualization and Computer Graphics*: 1–1.
11. Søren Knudsen and Sheelagh Carpendale. 2016. View Relations: An Exploratory Study on Between-View Meta-Visualizations. *Proc NordiCHI*, ACM.
12. H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics* 18, 9: 1520–1536.
13. Joseph E Mcgrath. 1995. Methodology matters: Doing research in the behavioral and social sciences. *Readings in Human-Computer Interaction: Toward the Year 2000 (2nd ed*, Morgan Kaufman, 152–169.
14. Gonzalo Gabriel Méndez, Uta Hinrichs, and Miguel A. Nacenta. 2017. Bottom-up vs. Top-down: Trade-offs in Efficiency, Understanding, Freedom and Creativity with InfoVis Tools. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ACM, 841–852.
15. Tamara Munzner. 2009. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics* 15, 6: 921–928.
16. Alice Thudt, Bongshin Lee, Eun Kyoung Choe, and Sheelagh Carpendale. 2017. Expanding Research Methods for a Realistic Understanding of Personal Visualization. *IEEE Computer Graphics and Applications* 37, 2: 12–18.
17. Melanie Tory. 2014. User Studies in Visualization: A Reflection on Methods. In W. Huang, ed., *Handbook of Human Centric Visualization*. Springer New York, 411–426.