Analyzing Qualitative Data

Sheelagh Carpendale¹, Uta Hinrichs², Søren Knudsen¹, Alice Thudt¹, Melanie Tory³
¹University of Calgary, ²University of St. Andrews, ³Tableau

“Not everything that can be counted counts and not everything that counts can be counted.”
(usually attributed to Albert Einstein, sometimes to Cameron [4])

Abstract
Evaluation is increasingly recognized as an essential component of visualization 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. 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.

Keywords: Evaluation, qualitative studies, data analysis, data coding.

Introduction
We, as visualization researchers, are increasingly interested in evaluation [5,11,13,16,17,18]. However, evaluation is a complex multi-faceted process that involves many skills [14]. In a previous tutorial, we provided an overview of qualitative evaluation data collection skills. In this tutorial¹, we will focus on the analysis of qualitative data, including the analysis of interview transcripts and video data, drawing from previous work [3,10] and our own experience [9,12,15]. The tutorial will introduce participants to the concepts of grounded theory [6,8] 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.

1. TUTORIAL CONTENT AND SCOPE
In this tutorial we focus on how to proceed once you have carefully collected your fabulous rich qualitative data. As a good basis from which to start we will consider Bryman’s four stages of qualitative analysis [3]. These are:
- 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.

1.1 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 for angles from which to analyze the text or transcript.

1.2 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.

1.3 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

---
¹ email: {sheelagh,athudt,sknudsen}@ucalgary.ca
² email: uh3@st-andrews.ac.uk
³ email: mtory@tableau.com

4 Find materials from last year’s tutorial here: http://innovis.cpsc.ucalgary.ca/qualeval-vis-tutorial/
5 Find materials from this tutorial here: http://innovis.cpsc.ucalgary.ca/qualitativeanalysis-iss-tutorial/
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 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
invariably introduced while coding and interpreting the collected
data.

1.4 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.

2 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.

2.1 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.

2.2 Activity 2: Closed Coding
In this activity we will introduce participants to closed coding,
oxen also referred to as analysis with a priori codes [7]. 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
classparticipant 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.

2.3 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.

2.4 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 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.

3 Tutorial Outline
The half-day tutorial will be a mix of short talks and hands-on
activities. We describe a tentative schedule below.

2:00 Brief introduction of the organizers and qualitative
evaluation approaches and how people work with qualitative
data. Brief introduction of participants.

2:15 Talk 1 (15 minutes): We will talk about the challenges of
analyzing qualitative data.

2:30 Activity 1: Choosing a Coding Focus. Working with
transcripts we will provide, we will hold a group idea
generation session to consider what would be useful foci to
code for. The exercise concludes with a group discussion.

2:45 Talk 2 (15 minutes): The purposes and techniques of closed
coding.

3:00 Activity 2: Closed Coding. We will provide qualitative data
as transcripts. We will start a discussion of closed coding
and provide a set of codes. Each tutorial participant will
individually engage in practicing closed coding (15
minutes). This will be followed by a 10-minute discussion
of the coded results in pairs or small groups. The exercise
concludes with a group discussion.

3:30 Break (30 minutes).

4:00 Talk 3 (15 minutes): The purposes and techniques of open
coding.

4:15 Activity 3: Open Coding. We will work with a different set
of transcribed qualitative data. We will start a discussion of
the open coding process and provide an initial set of codes.
Each tutorial participant will individually engage in practicing open coding (15
minutes). This will be followed by a 10-minute discussion of the coded results in pairs or
small groups. The exercise concludes with a group discussion.

4:35 Talk 4 (15 minutes): Consensus and Agreement.

4:50 Activity 4: Clustering and Categorizing. Having coded
qualitative data, it is still necessary to make sense of the
codes. This is usually approached through various forms of
clustering and categorizing.

5:10 Talk 5 (20 minutes): Clustering and Categorizing.

5:30 Closing Discussion

5:55 End

4 Conclusions
From this tutorial, participants will learn more about the benefits,
uantages 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.
5 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.

6 INSTRUCTOR INFORMATION
The instructors in alphabetical order are:

Sheelagh Carpendale is a Professor in Computer Science at the University of Calgary where she holds a Canada Research Chair in Information Visualization and NSERC/AITF/SMART Technologies Industrial Research Chair in Interactive Technologies. She has many received 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 field 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 information visualization and large interactive displays. She both conducts and publishes about evaluation in information visualization 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 visualization, HCI, 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.

Soren 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 information visualization, HCI, and large interactive displays. 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 digital 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. She also published an article on the benefits of qualitative methods for gaining a more realistic understanding of personal visualizations.

Melanie Tory is a senior research scientist at Tableau. Her research focuses on interactive visual data analysis. This includes intuitive interactions with visualizations and the design and evaluation of tools that support the holistic data analysis process, including sensemaking, analytical guidance, and collaboration. Before joining Tableau, Melanie was an Associate Professor in visualization at the University of Victoria. She is Associate Editor of IEEE Computer Graphics and Applications and has served as Papers Co-chair for the IEEE InfoVis and ACM Interactive Surfaces and Spaces conferences. Melanie has conducted a large number of evaluation studies and contributed a chapter on empirical methods to the Handbook of Human Centric Visualization.

REFERENCES