

17-803 Empirical Methods

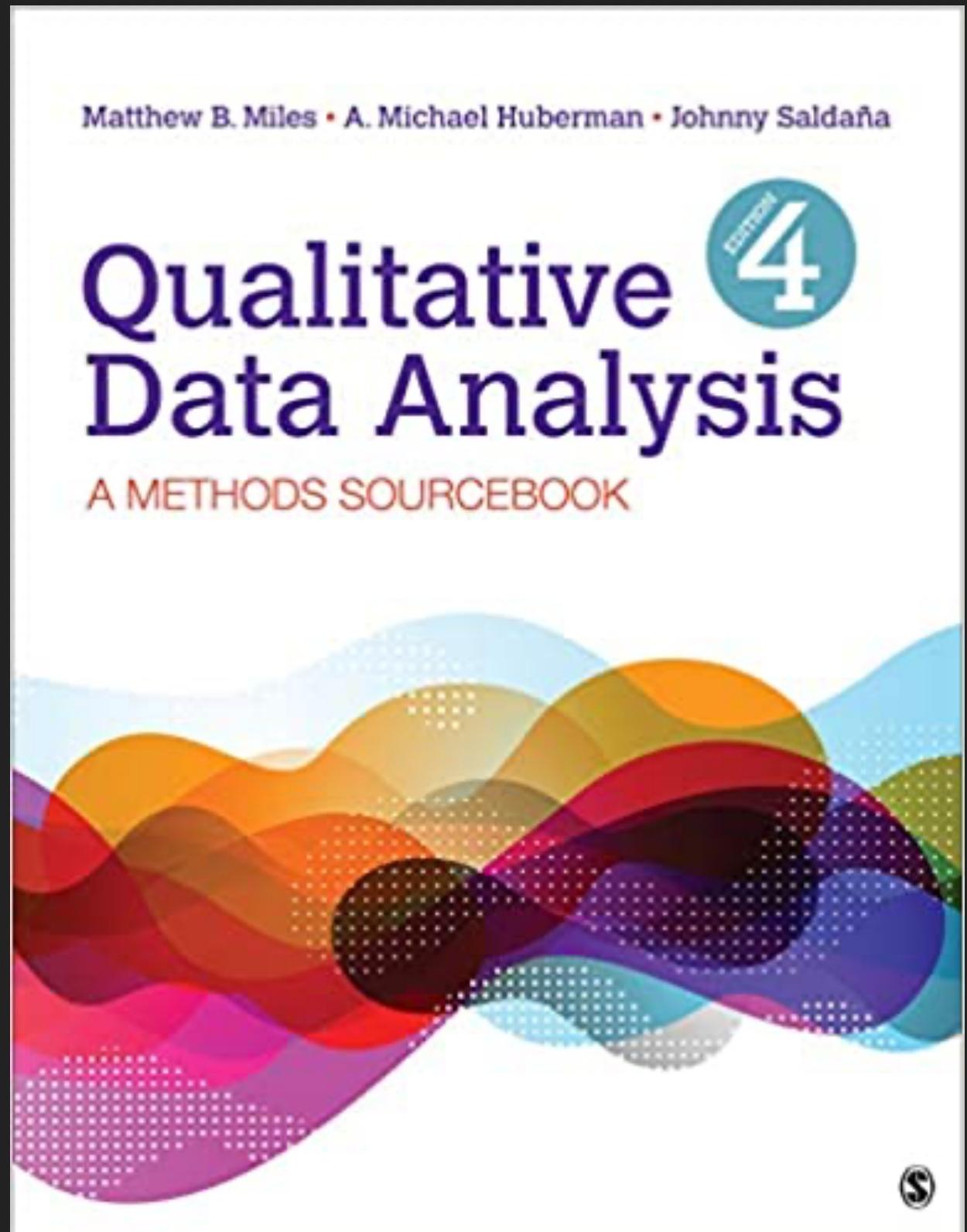
Bogdan Vasilescu, Institute for Software Research

# Qualitative Analysis

Wednesday, February 24, 2021

# Outline for Today

- ▶ Second “half” of interviewing – the analysis
- ▶ Trustworthiness in qualitative research
- ▶ Hands-on coding



# **Part I: Qualitative Analysis**

Miles, Huberman, & Saldaña - Qualitative Data Analysis: A Methods Sourcebook - Chapter 4

# Qualitative Content Analysis

- ▶ Piles of qualitative data, mostly **text**
  - ▶ What to do with it?
  - ▶ From journalism to science - how?
- ▶ **Step 1: Abstraction**
  - ▶ Attach "codes" (labels) to chunks of data
  - ▶ Characterize / summarize the data
- ▶ **Step 2: Finding patterns**
  - ▶ Use these abstractions to find meta-patterns, craft a theory ("grounded theory"), ...
  - ▶ Interpret the data
- ▶ This is difficult, but very doable with practice

# Step 1: Coding

# Types of Coding – Descriptive

- ▶ Code summarizes the basic topic of a passage of text

<sup>1</sup> As I walked toward the school, there was a 7-11 convenience store 1 block away, next to a small professional office building: an optometrist, podiatrist, and other medical/health-related clinics. Directly across the street was an empty lot, but next to that stood a Burger King restaurant.

<sup>1</sup> BUSINESSES

# Types of Coding – in Vivo

- ▶ Short quote as code

I<sup>1</sup> hated school last year. Freshman year, it was awful, I hated it. And<sup>2</sup> this year's a lot better actually I, um, don't know why. I guess, over the summer I kind of<sup>3</sup> stopped caring about what other people thought and cared more about, just, I don't know.

<sup>1</sup> "HATED SCHOOL"

<sup>2</sup> "THIS YEAR'S BETTER"

<sup>3</sup> "STOPPED CARING"

# Types of Coding – Process

## ► Actions ("‐ing" words)

Well, that's one problem, that [my school is] pretty small, so <sup>1</sup> if you say one thing to one person, and then they decide to tell two people, then those two people tell two people, and in one period everybody else knows. <sup>2</sup> Everybody in the entire school knows that you said whatever it was. So . . .

<sup>1</sup> SPREADING RUMORS

<sup>2</sup> KNOWING WHAT YOU SAID

# Types of Coding – Emotion

- ▶ Experienced by participant or inferred by researcher

<sup>1</sup> I just hated it when he got awarded with the honor. <sup>2</sup> I mean, we're praising mediocrity now. Never mind that what you've accomplished isn't worth squat, it's all about who you know in the good ol' boys network.

<sup>1</sup> "HATED IT"  
<sup>2</sup> BITTERNESS

# Types of Coding – Values (V), Attitudes (a), Beliefs (B)

<sup>1</sup> Government regulation of women's health issues has gotten out of hand. It's not about "protecting" us, it's about their need to control and dominate women <sup>2</sup> through covert religious ideology. White Christian men are deciding what's law and what's moral and what's, how it's supposed to be. <sup>3</sup> They can say, "It's not a war on women" all they want, but trust me—it's a war on women.

<sup>1</sup> B: GOVERNMENTAL CONTROL

<sup>2</sup> B: COVERT RELIGIOUS MOTIVES

<sup>3</sup> A: MISOGYNIST MOTIVES

# Types of Coding – Provisional Coding

- ▶ Begin with a “start list” of researcher-generated codes
- ▶ Revise, delete, expand as needed

Q: When would you do this?

A: appropriate for qualitative studies that build on or corroborate previous research and investigations

PRESCRIPTION MEDICATION  
NICOTINE PATCHES  
NICOTINE GUM/LOZENGES  
“ELECTRONIC” CIGARETTES  
PROFESSIONAL COUNSELING  
PEER SUPPORT SYSTEM  
“COLD TURKEY”

# Types of Coding – Hypothesis Coding

- ▶ Apply predetermined list of codes specifically to assess a hypothesis
- ▶ The codes are developed from a theory/prediction about what will be found in the data before they have been collected or analyzed.

Used when searching for rules, causes, and explanations in the data.

# Coding Process - Summary Considerations

- ▶ Deductive (“start list”) vs inductive coding
- ▶ Analysis concurrent with data collection
  - ▶ Helps identify blind spots / which new data to collect
- ▶ Clear operational definitions are indispensable
  - ▶ Apply consistently over time / by different researchers
- ▶ Level of detail
  - ▶ Any block of text is a candidate for more than one code
  - ▶ Not every portion of the transcripts must be coded

## Step 2: “Pattern Coding”

- ▶ Categories / Themes
- ▶ Causes / Explanations
- ▶ Relationships among people
- ▶ Theoretical constructs

# There Is More Than One Way To Find Patterns

Consider these codes related to the first month of withdrawal symptoms described by a participant in a smoking cessation treatment program:

- ANXIETY  
[emotion]
- NERVOUSNESS  
[emotion]
- RESTLESSNESS  
[emotion]
- DEEP BREATHING  
[process]
- THROAT BURNING  
[process]
- “FELT LIKE CRYING”  
[in vivo/emotion/process]
- “HURT SOMEONE BAD”  
[in vivo/emotion]
- ANGRY  
[emotion]
- “EATING A LOT MORE”  
[in vivo/process]
- WANDERING AROUND  
[process]
- HABITUAL MOVEMENTS  
[descriptive]
- MEMORIES OF SMOKING  
[descriptive]
- SMELLING NEW THINGS  
[process]

# One Way

Pattern by code type:

- **EMOTIONS** (ANXIETY, NERVOUSNESS, “HURT SOMEONE BAD,” RESTLESSNESS, “FELT LIKE CRYING,” ANGRY)
- **PROCESSES** (DEEP BREATHING, THROAT BURNING, “FELT LIKE CRYING,” “EATING A LOT MORE,” WANDERING AROUND, SMELLING NEW THINGS)
- **DESCRIPTORS** (HABITUAL MOVEMENTS, MEMORIES OF SMOKING)

Q: Do these make sense?

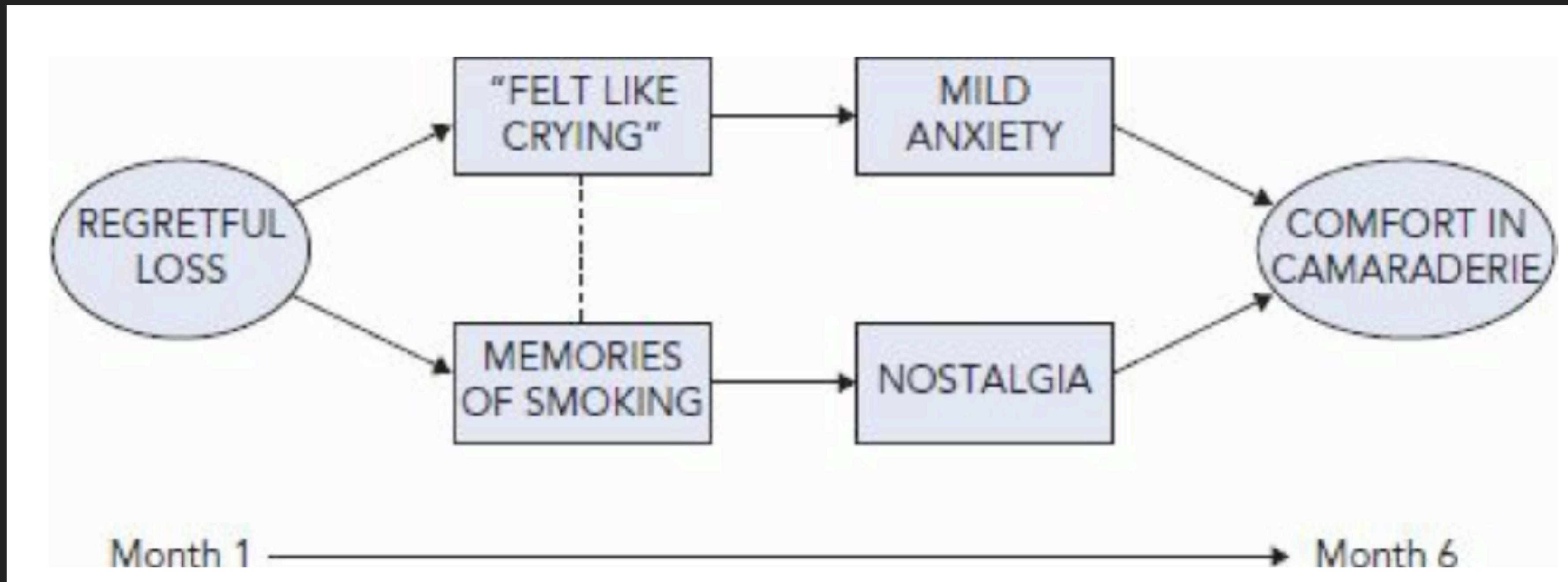
# A Better Way?

Recategorize PROCESSES and DESCRIPTORS:

- NEGATIVE EMOTIONS (ANXIETY, NERVOUSNESS, “HURT SOMEONE BAD,” RESTLESSNESS, “FELT LIKE CRYING,” ANGRY)
- PHYSICAL CHANGES: DEEP BREATHING, THROAT BURNING, “EATING A LOT MORE,” SMELLING NEW THINGS
- RESTLESS JOURNEY: WANDERING AROUND, HABITUAL MOVEMENTS
- REGRETFUL LOSS: “FELT LIKE CRYING,” MEMORIES OF SMOKING

Note: inherently subjective process

# A Possible Next Step on the Way to a Theory: Network Display



# Analytic Memoing

- ▶ Narrative that documents **reflections** and **thinking processes** about the data.
  - ▶ Not just descriptive summaries but attempts to synthesize **higher level analytic meanings**.
- ▶ Generate and memo **assertions** and **propositions**
  - ▶ Assertions – descriptive, broad-brushstroke facts
    - ▶ “Overall, the participant seemed engaged with the NL2Code tool”
  - ▶ Propositions – higher level interpretations about the meanings of the study
    - ▶ “Having pull requests rejected can be demotivating for contributors already demoralized by low self confidence in their programming expertise”
    - ▶ gets closer to prediction or theory

# **Part II: Establishing Trustworthiness in Qualitative Research**

Miles, Huberman, & Saldaña - Qualitative Data Analysis: A Methods Sourcebook - Chapter 11

# A Few Possible Sources of Analytic Bias

- ▶ The holistic fallacy:
  - ▶ finding patterns where there aren't any
- ▶ Elite bias:
  - ▶ overweighting data from high-status participants
- ▶ Personal bias
- ▶ Going native:
  - ▶ losing your outsider perspective

# Confirmability

- ▶ Concerned with establishing that the researcher's interpretations and findings are clearly derived from the data.
  - ▶ demonstrate how conclusions and interpretations have been reached.
- ▶ Confirmability is established when **credibility**, **transferability**, and **dependability** are all achieved (Guba and Lincoln, 1989).
- ▶ Strategies:
  - ▶ articulate the reasons for the theoretical, methodological, and analytical choices throughout the entire study, so that others can understand how and why decisions were made.

# Credibility

- ▶ The credibility of a study is determined when co-researchers or readers are confronted with the experience, they can **recognize it**.
- ▶ Credibility addresses the “**fit**” between respondents’ views and the researcher’s representation of them.
- ▶ Strategies:
  - ▶ prolonged engagement
  - ▶ persistent observation
  - ▶ data collection triangulation
  - ▶ researcher triangulation
  - ▶ member checking

# Strategy To Increase Credibility: Prolonged Engagement

- ▶ Conducting a study for a sufficient period of time to obtain an adequate representation of the “voice” under study.
- ▶ ‘Hawthorne Effect’: participants’ alteration of behavior when observed
  - ▶ What participants **want us** to see vs **what really goes on** when no one is watching
  - ▶ But:
    - ▶ “Evidence of a Hawthorne Effect is scant, and amounts to little more than a good story.” (Paradis & Sutkin, 2017)

# Strategy To Increase Credibility: Persistent Observation

- ▶ Identify the characteristics, attributes, and traits that are most relevant to the phenomena under investigation and focus on them extensively.
  - ▶ separate relevant from irrelevant observations.
  - ▶ prolonged engagement – scope; persistent observation – depth.

# Strategy To Increase Credibility: Triangulation

- ▶ Using multiple and different methods, investigators, sources, and theories to obtain corroborating evidence.
- ▶ Reduces the possibility of chance associations, as well as of systematic biases prevailing due to a specific method being utilized
- ▶ Four types:
  - ▶ Data triangulation: use of a variety of sources in a study
  - ▶ Investigator triangulation: use of several different researchers
  - ▶ Theory triangulation: use of multiple perspectives to interpret the results of a study
  - ▶ Methodological triangulation: use of multiple methods to study a research problem

# Strategy To Increase Credibility: Member Checking

- ▶ Recall the Bogart et al “breaking changes” paper
  - ▶ “We presented interviewees with both a summary and a full draft of Sections 4-7, along with questions prompting them to look for correctness and areas of agreement or disagreement (i.e., fit), and any insights gained from reading about experiences of other developers and platforms (i.e., applicability).”
- ▶ Recall the Barwulor et al “sex workers” paper
  - ▶ “After we drafted the interview protocol, we hired a sex worker as a consultant to review our protocol for appropriateness and to ensure a member of the community under study was involved in the research to the extent that they desired to be involved [68]. The consultant was paid market rate for their work.”

# Transferability

- ▶ Transferability refers to the generalizability of inquiry
  - ▶ case-to-case transfer.
- ▶ Strategies:
  - ▶ Provide thick descriptions (**quotes**), so that those who seek to transfer the findings to their own site can judge transferability.

# Dependability

- ▶ The research process is logical, traceable, and clearly documented.
- ▶ Strategies:
  - ▶ Leave audit trail

# Strategy To Increase Transferability and Dependability: Audit Trails

- ▶ A study and its findings are auditable when another researcher **can clearly follow the decision trail** regarding theoretical and methodological issues throughout the study.
- ▶ Could another researcher with the same data, perspective, and situation **arrive at the same or comparable, but not contradictory, conclusions?**

# Leaving an Audit Trail

- ▶ Maintaining extensive documentation of records and data:
  - ▶ raw data (e.g., videotapes, written notes, survey results);
  - ▶ data reduction and analysis products (e.g., write-ups of field notes, summaries, unitized information, quantitative summaries, theoretical notes);
  - ▶ data reconstruction and synthesis products (e.g., structure of categories, findings and interpretations, final reports);
  - ▶ process notes (i.e., methodological notes, trustworthiness notes, audit trail notes);
  - ▶ materials related to intentions and dispositions (e.g., research proposal, personal notes, reflexive journals, expectations);
  - ▶ instrument development information (e.g., pilot forms, preliminary schedules, observation formats, and surveys).

# See Also

- ▶ Roller, M. R., & Lavrakas, P. J. (2018). A total quality framework approach to sharing qualitative research data: Comment on Dubois et al. (2018). *Qualitative Psychology*, 5(3), 394-401.

# Aside: Should You Share Qualitative Data?

- ▶ Pros:
  - ▶ Transparency, verifiability
    - ▶ e.g., failures to reproduce key findings of seminal studies in social psychology
  - ▶ Enables new research with existing data
    - ▶ Recall, many possible ways to code the same data
  - ▶ Useful for teaching
    - ▶ e.g., this class
- ▶ Cons:
  - ▶ Threat to privacy or a breach of trust within the interviewer-interviewee relationship
    - ▶ Might be ok if data are adequately de-identified?
    - ▶ Get consent!
  - ▶ Policy / legislation

# Summary

# Three Approaches to Qualitative Content Analysis

**TABLE 4: Major Coding Differences Among Three Approaches to Content Analysis**

Type of Content Analysis	Study Starts With	Timing of Defining Codes or Keywords	Source of Codes or Keywords
Conventional content analysis	Observation	Codes are defined during data analysis	Codes are derived from data
Directed content analysis	Theory	Codes are defined before and during data analysis	Codes are derived from theory or relevant research findings
Summative content analysis	Keywords	Keywords are identified before and during data analysis	Keywords are derived from interest of researchers or review of literature

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.

# Three Approaches to Qualitative Content Analysis

**TABLE 4: Major Coding Differences Among Three Approaches to Content Analysis**

Type of Content Analysis	Study Starts With	Timing of Defining Codes or Keywords	Source of Codes or Keywords
Conventional content analysis	Observation	Codes are defined during data analysis	Codes are derived from data
Directed content analysis	Theory	Codes are defined before and during data analysis	Codes are derived from theory or relevant research findings
Summative content analysis	Keywords	Keywords are identified before and during data analysis	Keywords are derived from interest of researchers or review of literature

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.

# Three Approaches to Qualitative Content Analysis

**TABLE 4: Major Coding Differences Among Three Approaches to Content Analysis**

Type of Content Analysis	Study Starts With	Timing of Defining Codes or Keywords	Source of Codes or Keywords
Conventional content analysis	Observation	Codes are defined during data analysis	Codes are derived from data
Directed content analysis	Theory	Codes are defined before and during data analysis	Codes are derived from theory or relevant research findings
Summative content analysis	Keywords	Keywords are identified before and during data analysis	Keywords are derived from interest of researchers or review of literature

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.

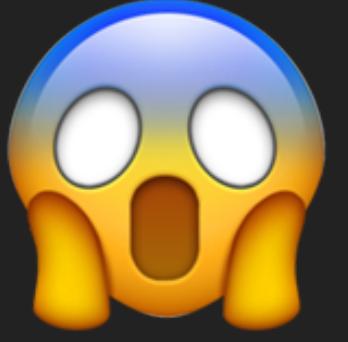
# Three Approaches to Qualitative Content Analysis

**TABLE 4: Major Coding Differences Among Three Approaches to Content Analysis**

Type of Content Analysis	Study Starts With	Timing of Defining Codes or Keywords	Source of Codes or Keywords
Conventional content analysis	Observation	Codes are defined during data analysis	Codes are derived from data
Directed content analysis	Theory	Codes are defined before and during data analysis	Codes are derived from theory or relevant research findings
Summative content analysis	Keywords	Keywords are identified before and during data analysis	Keywords are derived from interest of researchers or review of literature

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative health research*, 15(9), 1277-1288.

# Next Time



“Grounded theory’ is often used as rhetorical sleight of hand by authors who are unfamiliar with qualitative research and who wish to avoid close description or illumination of their methods. More disturbing, perhaps, is that it becomes apparent, when one pushes them to describe their methods, that many authors hold some serious misconceptions about grounded theory.”

Suddaby, R. 2006. From the editors: What grounded theory is not. *Academy of Management Journal*, 49, 4, 633-642.

# Activity

- ▶ 2 groups
- ▶ Read interview excerpts (5 minutes)
- ▶ Why participate in corporate hackathon?
  - ▶ Develop codes (10 minutes)
  - ▶ Apply codes to transcript, compare notes in group (10 minutes)
  - ▶ Report out (10 minutes)

# Activity Results

- ▶ Group 1 codes:
  - ▶ Enjoyable
    - ▶ Creative
    - ▶ Passionate
    - ▶ Fun
    - ▶ Self-ownership
    - ▶ Low-stakes
    - ▶ Meeting new people
    - ▶ Something new
  - ▶ Self-Improvement
    - ▶ Learning
    - ▶ Improving skills
    - ▶ Self-ownership
  - ▶ Innovation
    - ▶ Exploration
    - ▶ Test/implement new ideas
    - ▶ Iteration
  - ▶ Professional Development
    - ▶ Planning / Preparation
    - ▶ Self-ownership
- ▶ Group 2 codes:
  - ▶ Creative
  - ▶ Iterate
  - ▶ Exploration
  - ▶ Project Fit/Interest
  - ▶ Meet new people
  - ▶ Low Stakes
  - ▶ Out of ordinary
  - ▶ Fun
  - ▶ Passionate
  - ▶ Interests

# Credits

- ▶ **Graphics:**
  - ▶ Dave DiCello photography (cover)
- ▶ **Content:**
  - ▶ Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). Qualitative data analysis: A methods sourcebook. Sage publications. (Chapters 4 & 11)
  - ▶ Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1609406917733847.
  - ▶ Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.
  - ▶ Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron?. *Quality & Quantity*, 41(2), 233-249.
  - ▶ DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data?. *Qualitative Psychology*, 5(3), 380.
  - ▶ Roller, M. R., & Lavrakas, P. J. (2018). A total quality framework approach to sharing qualitative research data: Comment on Dubois et al. (2018). *Qualitative Psychology*, 5(3), 394-401.
  - ▶ Paradis, E., & Sutkin, G. (2017). Beyond a good story: from Hawthorne Effect to reactivity in health professions education research. *Medical Education*, 51(1), 31-39.