17-803 Empirical Methods

Bogdan Vasilescu, Institute for Software Research

Qualitative Analysis

Tuesday, September 20, 2022

Outline for Today

▶ Second "half" of interviewing – the analysis

▶ Trustworthiness in qualitative research

▶ Hands-on coding

Matthew B. Miles • A. Michael Huberman • Johnny Saldaña

Qualitative Data Analysis

A METHODS SOURCEBOOK

4

▾ 📁 qualitative analysis
  📄 activity - Hackathon interviews anonymized transcript excerpts.pdf
  📄 Chapter 11 from Lazar, Feng, and Hochheiser –...arch Methods in Human Computer Interaction.pdf
  📄 Corbin & Strauss – Chapter 8 – Analyzing Data for Concepts.pdf
  📄 Corbin & Strauss – Chapter 9 – Elaborating the Analysis.pdf
  📄 Corbin & Strauss – Chapter 10 - Analyzing Data for Context.pdf
  📄 Corbin & Strauss – Chapter 11 - Bringing Process into the Analysis.pdf
  📄 Corbin & Strauss – Chapter 12 - Integrating Categories.pdf
  📄 Miles, Huberman, & Saldaña – Qualitative Data Analysis_ A Methods Sourcebook - Chapter 4.pdf
  📄 Miles, Huberman, & Saldaña – Qualitative Data Analysis_ A Methods Sourcebook - Chapter 11.pdf
  📄 qual content analysis - Three Approaches.pdf
  📄 thematic - Using thematic analysis in psychology
  📄 trustworthiness - A Total Quality Framework Approach to Sharing QualitativeResearch Data.pdf
  📄 trustworthiness - Beyond a good story/ from Ha...vity in health professions education research.pdf
  📄 trustworthiness - Is It Time to Share Qualitative Research Data?.pdf
  📄 trustworthiness - Onwuegbuzie-Leech2007_Article_ValidityAndQualitativeResearch.pdf
  📄 trustworthiness - Thematic Analysis- Striving to Meet the Trustworthiness Criteria.pdf
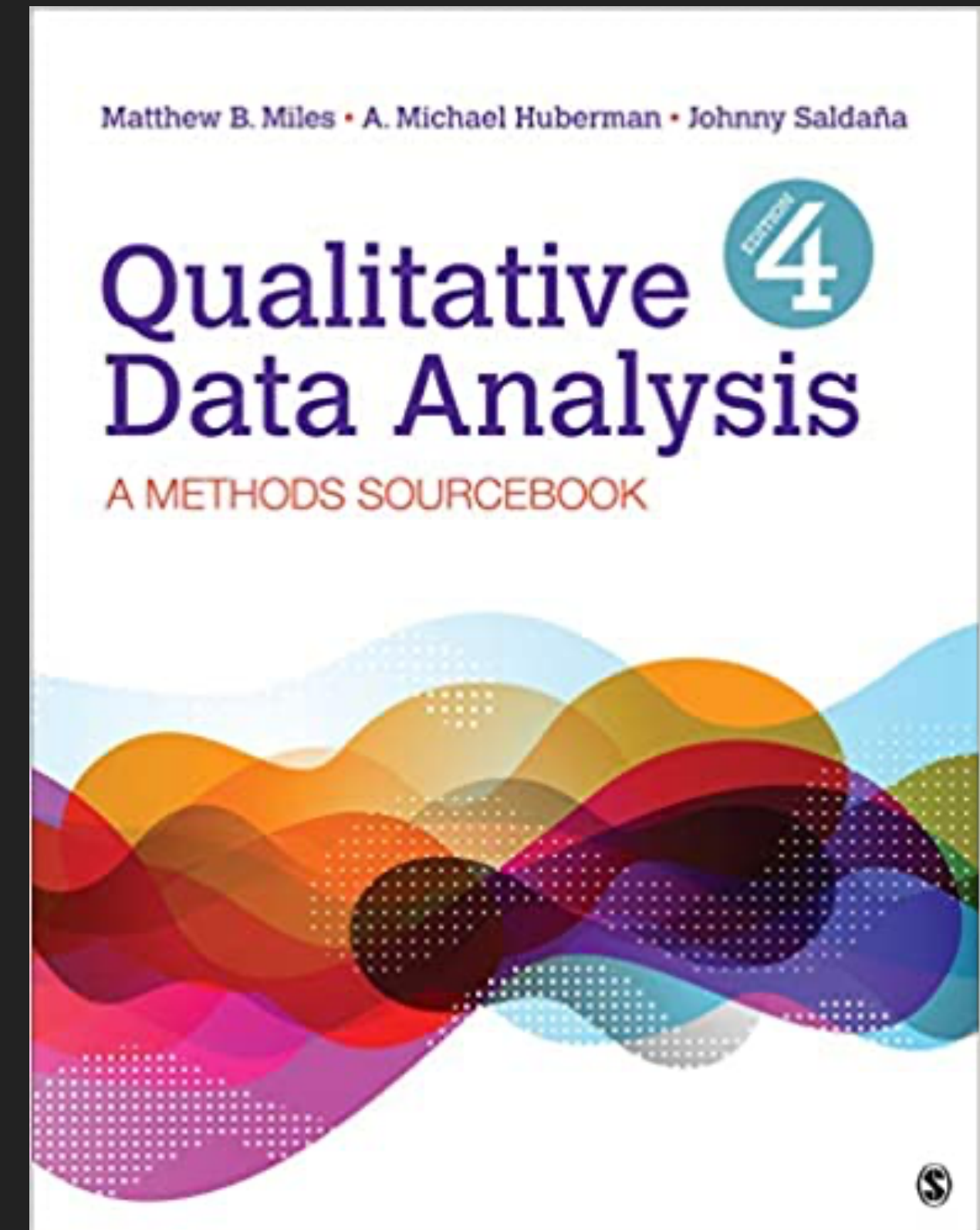
Part I: Qualitative Analysis

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

Types of Coding — in Vivo

▸ Short quote as code

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

[1] "HATED SCHOOL"

[2] "THIS YEAR'S BETTER"

[3] "STOPPED CARING"

Types of Coding – Process

▸ Actions ("-ing" words)

Well, that's one problem, that [my school is] pretty small, so [1] 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. [2] Everybody in the entire school knows that you said whatever it was. So. . . .

[1] SPREADING RUMORS

[2] KNOWING WHAT YOU SAID

Types of Coding — Emotion

▸ Experienced by participant or inferred by researcher

[1] I just hated it when he got awarded with the honor. [2] 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.

[1] "HATED IT"
[2] BITTERNESS

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

[1] 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 [2] 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. [3] They can say, "It's *not* a war on women" all they want, but trust me—it's a war on women.

[1] B: GOVERNMENTAL CONTROL

[2] B: COVERT RELIGIOUS MOTIVES

[3] 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 <u>before</u> 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

Example: Coding
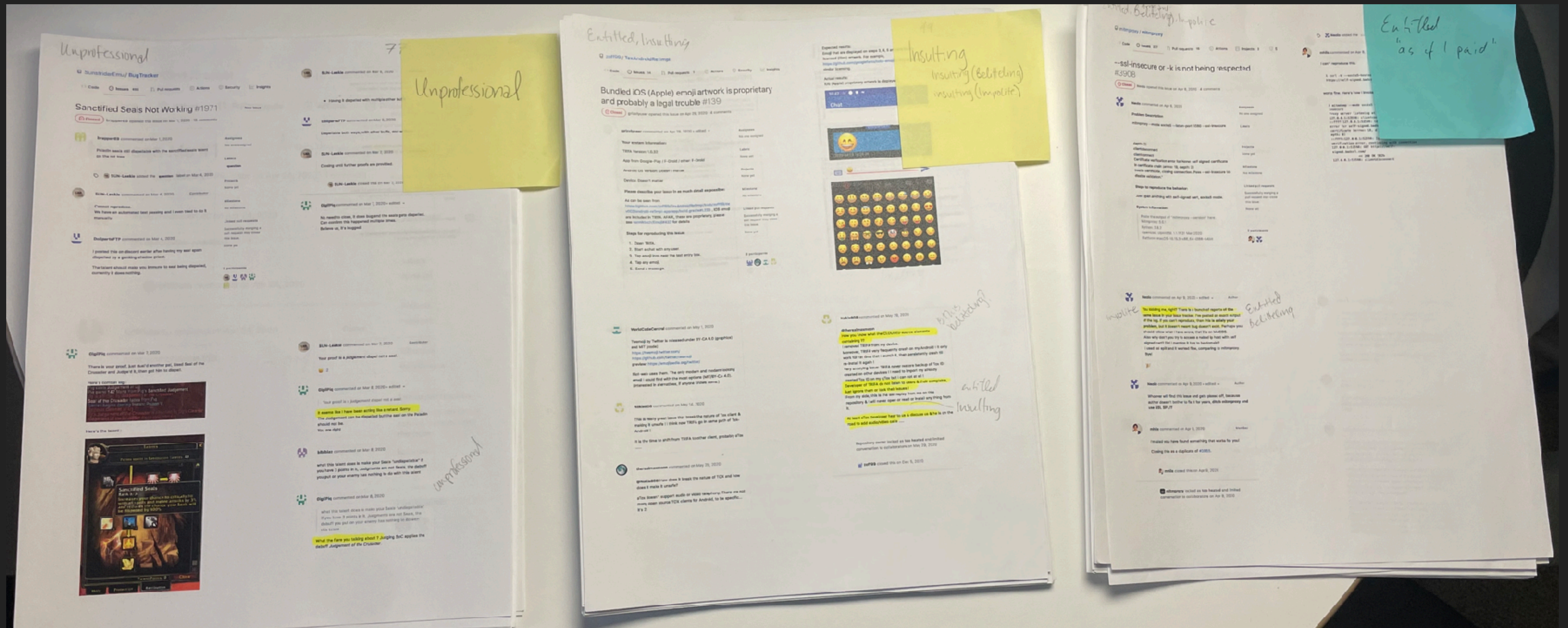
Miller, C., Cohen, S., Klug, D., Vasilescu, B., & Kästner, C. (2022). "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In In 44th International Conference on Software Engineering (ICSE'22).

Example: Codes and Definitions

	Category	Label	Definition
1	Category	Label	Definition
2	position	opening with toxic comment	Toxicity appears in the issue
3		emerging from discussion	Toxicity comes about within the thread
4	what triggered toxicity	failed use of tool/code or errormsg	A problem or technical difficulty with the project
5		politics/ideology	Expressing one's beliefs or general values on product or process
6		past interactions	The users have corresponded in some way before (on GitHub or elsewhere)
7		technical disagreement	Differing views on some technical component
8	target of toxicity	undirected	No real reason except to add emphasis to what is being said
9		at code	Targeted at the project or a specific component of the project
10		at people	Targeted at another user
11	nature of the comment	entitled	A condescending or arrogant tone in a comment or request, as if the author is above the others in the thread
12		troll	Nonsense or an unrelated comment, no actual substance
13		joking	The use of humor that is clearly received as such
14		complaining	Expressing annoyance or dissatisfaction about something (sometimes in very unprofessional language)
15	severity of language	colloquial to offensive	Slang to rude or aggressive comments
16		cursing vs softer	Swear words vs colloquial expressions (acronyms or abbreviations)
17		professionalism	
18	who?	troll (new account)	The user has not opened an issue before and has essentially no other activity on GitHub
19		serial issue reporter	Merging of repeated troll and repeat offender. Users who repeatedly open issues on GitHub with little other activity either in general or within a given project
21		experienced developer	
22		project member	Someone who is part of the project
23		friends	It is clear that the users know each other personally
24		first interaction with project	The user has not participated in this project ever before
26	where	big active project	hundres of followers, still regular commits/issue discussions
27		corporate project	
28		small project	

Miller, C., Cohen, S., Klug, D., Vasilescu, B., & Kästner, C. (2022). "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In In 44th International Conference on Software Engineering (ICSE'22).

Step 2: "Pattern Coding"

▸ Categories / Themes

▸ Causes / Explanations

▸ Relationships among people

▸ Theoretical constructs

Six Phases of Thematic Analysis

Phase	Description of the process
1. Familiarizing yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Card Sorting To Identify Themes

▸ This is when you start thinking about the relationship between codes, between themes, and between different levels of themes.
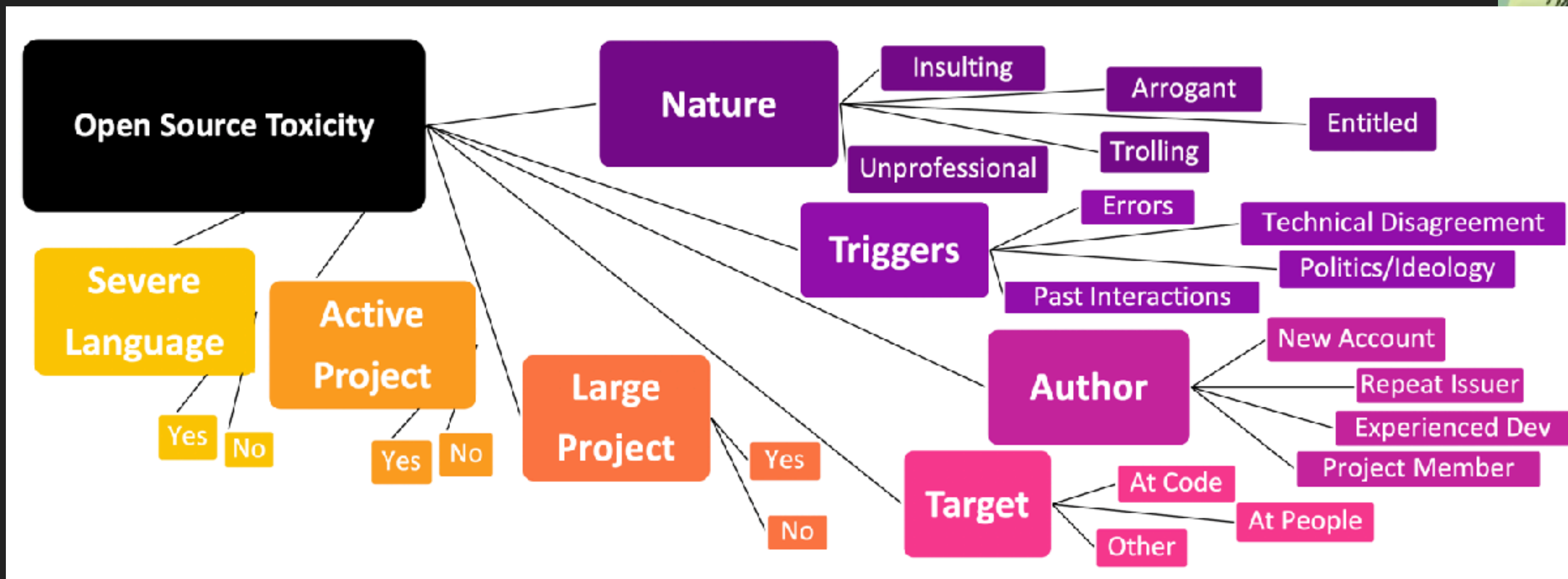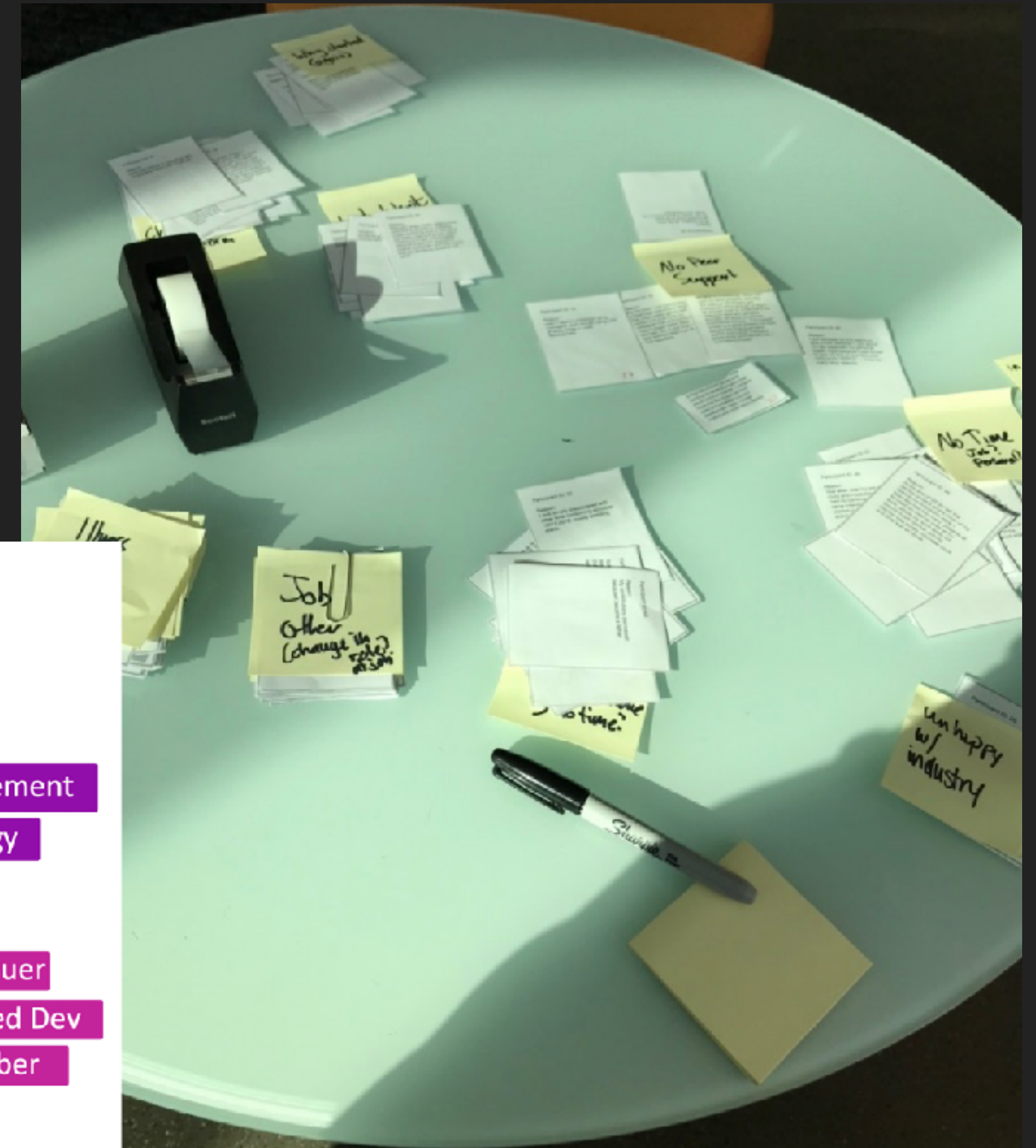
Miller, C., Cohen, S., Klug, D., Vasilescu, B., & Kästner, C. (2022). "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In In 44th International Conference on Software Engineering (ICSE'22).

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]
- RESTLESNESS
  [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
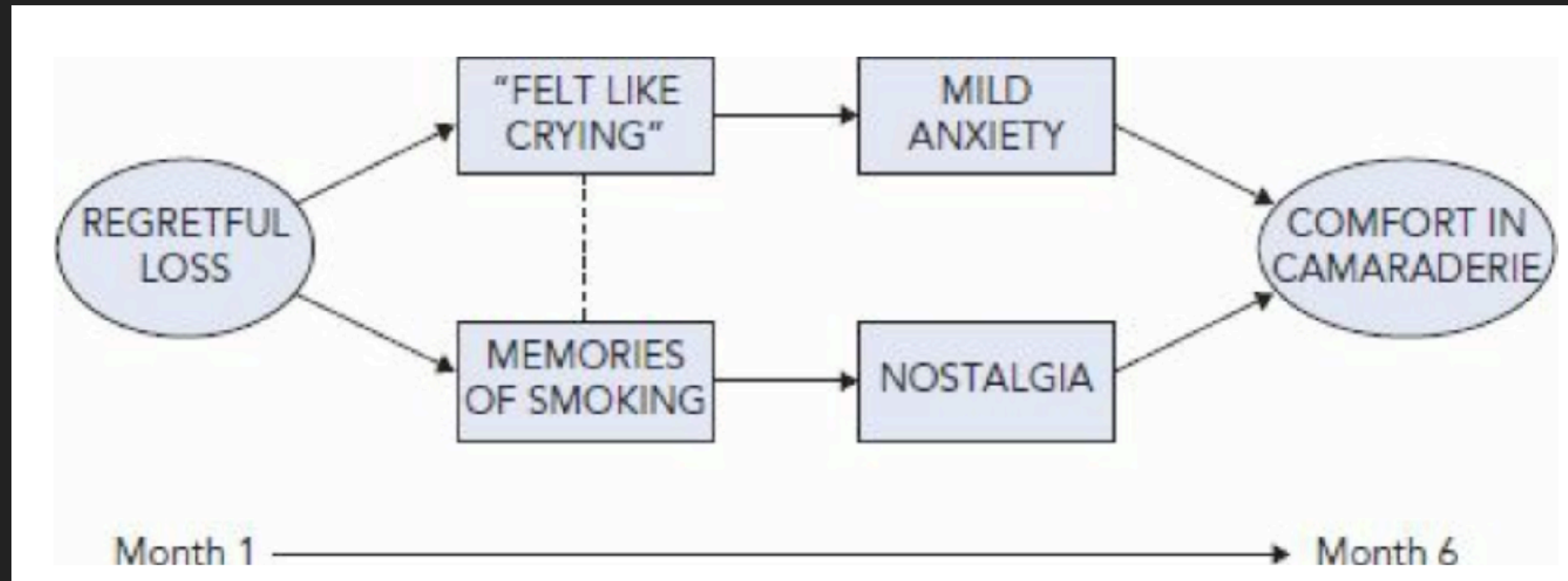
Qualitative Analysis vs Grounded Theory 😱
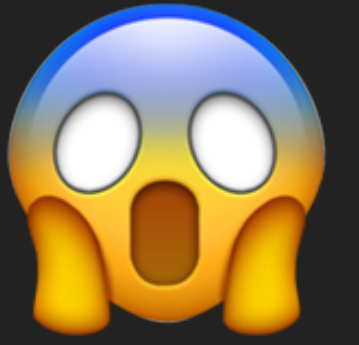
"'Grounded theory' is often used as rhetorical sleight of hand by authors who are unfamiliar with qualitative research and who wish to avoid close description or illumination of their methods. More disturbing, perhaps, is that it becomes apparent, when one pushes them to describe their methods, that 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.

Grounded Theory 'Lite'

"However, in our experience, grounded theory seems increasingly to be used in a way that is essentially grounded theory 'lite' – as a set of procedures for coding data very much akin to thematic analysis. Such analyses do not appear to fully subscribe to the theoretical commitments of a 'full-fat' grounded theory, which requires analysis to be directed towards theory development (Holloway and Todres, 2003). We argue, therefore, that a 'named and claimed' thematic analysis means researchers need not subscribe to the implicit theoretical commitments of grounded theory if they do not wish to produce a fully worked-up grounded-theory analysis."

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

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

Example: Memoing

Miller, C., Cohen, S., Klug, D., Vasilescu, B., & Kästner, C. (2022). "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In In 44th International Conference on Software Engineering (ICSE'22).

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

Seek To Establish 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.

Confirmability Through 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."

Confirmability Through 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.

Confirmability Through 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.

Example: Codes and Definitions

	Category	Label	Definition
1	Category	Label	Definition
2	position	opening with toxic comment	Toxicity appears in the issue
3		emerging from discussion	Toxicity comes about within the thread
4	what triggered toxicity	failed use of tool/code or errormsg	A problem or technical difficulty with the project
5		politics/ideology	Expressing one's beliefs or general values on product or process
6		past interactions	The users have corresponded in some way before (on GitHub or elsewhere)
7		technical disagreement	Differing views on some technical component
8	target of toxicity	undirected	No real reason except to add emphasis to what is being said
9		at code	Targeted at the project or a specific component of the project
10		at people	Targeted at another user
11	nature of the comment	entitled	A condescending or arrogant tone in a comment or request, as if the author is above the others in the thread
12		troll	Nonsense or an unrelated comment, no actual substance
13		joking	The use of humor that is clearly received as such
14		complaining	Expressing annoyance or dissatisfaction about something (sometimes in very unprofessional language)
15	severity of language	colloquial to offensive	Slang to rude or aggressive comments
16		cursing vs softer	Swear words vs colloquial expressions (acronyms or abbreviations)
17		professionalism	
18	who?	troll (new account)	The user has not opened an issue before and has essentially no other activity on GitHub
19		serial issue reporter	Merging of repeated troll and repeat offender. Users who repeatedly open issues on GitHub with little other activity either in general or within a given project
21		experienced developer	
22		project member	Someone who is part of the project
23		friends	It is clear that the users know each other personally
24		first interaction with project	The user has not participated in this project ever before
26	where	big active project	hundres of followers, still regular commits/issue discussions
27		corporate project	
28		small project	

Miller, C., Cohen, S., Klug, D., Vasilescu, B., & Kästner, C. (2022). "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In In 44th International Conference on Software Engineering (ICSE'22).

Example: Memoing

Miller, C., Cohen, S., Klug, D., Vasilescu, B., & Kästner, C. (2022). "Did You Miss My Comment or What?" Understanding Toxicity in Open Source Discussions. In In 44th International Conference on Software Engineering (ICSE'22).

Example: Detailed Description of Methodology

Qualitative Analysis of Sampled Issues. To understand the important characteristics of open source toxicity, we qualitatively analyzed the 100 issues in our sample and their context, using *thematic analysis* [8], following the trustworthiness criteria by Lincoln and Guba [56] as demonstrated by Nowell et al. [69]. As we describe next, this was an iterative and reflective process, constantly moving back and forth between stages of engagement with the data, coding, memoing, searching for themes, and refining, as recommended during qualitative analysis [18].

Overall, our analysis consisted of several phases. We started by immersing ourselves in the data, carefully reading issue threads to understand the problems discussed, the project context, and the relationship of the authors of toxic comments with those projects as well as their past public activities (including past issues) on GITHUB. This typically took 15 to 30 minutes per issue, was conducted in groups of two or three researchers, and involved exploring, besides the issue threads themselves, also the project homepages and user profile pages of the discussants. As we were engaging with the data, we kept posing sensitizing questions regarding what we were observing (what, who, how, where, why, what for, etc.) and, guided by these, generated an initial set of codes to describe the toxic interactions through an otherwise inductive, data-driven process [43]. We also wrote down brief analytic memos [61] summarizing emergent patterns and the possible connections among the codes.

After analyzing 35 issues this way, we paused to sort and collate the codes we had assigned so far into a coding manual with detailed definitions and examples. In the manual, we organized the codes into higher-level categories following the sensitizing questions above to cover the key characteristics identified up to this point—nature of toxicity, nature of the comment, language severity, triggers, authors, position in discussion, project size, and domain, resolution, and identifiable harms. We then revisited all 35 issues using focused coding [22] to make sure our codes were applied consistently. A single researcher then analyzed and coded the remaining 65 issues in the same way, involving other researchers for difficult and ambiguous cases, and extending the coding frame iteratively for new observations, when needed.

For the interpretive categories,[7] in particular those related to the nature of the toxicity, we then searched for themes using *card sorting* [88]. This involved printing each issue discussion and sorting and resorting them into different themes, discussing theme boundaries and subthemes, again in groups of two or three researchers.

Finally, we systematically searched for relationships between the emerging themes across combinations of all nine categories of our coding frame, further exploring observations we had written down in our analytic memos, e.g., whether toxicity by experienced developers tends to use less severe language than toxicity by new accounts. We used exploratory data analysis and visualization to help with this process, like those in Figure 3.

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.

▸ Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

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.

You'll Probably Do a Lot of Thematic Analysis (Conventional CA)

▸ What counts as a theme?

  ▸ Retain flexibility, rigid rules do not work.

  ▸ The 'keyness' of a theme depends on whether it captures something important in relation to the overall research question.

▸ Rich description of the data vs detailed account of one particular aspect.

▸ Progression from description to interpretation (theorize the significance of the patterns and their broader meanings and implications).

▸ Inductive (data-driven) vs theoretical thematic analysis.

Potential Pitfalls To Avoid When Doing Thematic Analysis

‣ Failure to actually analyze the data at all

‣ Using the data collection questions (such as from an interview guide) as the 'themes' that are reported.

‣ Weak or unconvincing analysis (the themes do not appear to work):

   ‣ there is too much overlap between themes
   ‣ the themes are not internally coherent and consistent

‣ Mismatch between the data and the analytic claims that are made about it.

A 15–Point Checklist of Criteria for Good Thematic Analysis

Process	No.	Criteria
Transcription	1	The data have been transcribed to an appropriate level of detail, and the transcripts have been checked against the tapes for 'accuracy'.
Coding	2	Each data item has been given equal attention in the coding process.
	3	Themes have not been generated from a few vivid examples (an anecdotal approach), but instead the coding process has been thorough, inclusive and comprehensive.
	4	All relevant extracts for all each theme have been collated.
	5	Themes have been checked against each other and back to the original data set.
	6	Themes are internally coherent, consistent, and distinctive.
Analysis	7	Data have been analysed – interpreted, made sense of – rather than just paraphrased or described.
	8	Analysis and data match each other – the extracts illustrate the analytic claims.
	9	Analysis tells a convincing and well-organized story about the data and topic.
	10	A good balance between analytic narrative and illustrative extracts is provided.
Overall	11	Enough time has been allocated to complete all phases of the analysis adequately, without rushing a phase or giving it a once-over-lightly.
Written report	12	The assumptions about, and specific approach to, thematic analysis are clearly explicated.
	13	There is a good fit between what you claim you do, and what you show you have done – ie, described method and reported analysis are consistent.
	14	The language and concepts used in the report are consistent with the epistemological position of the analysis.
	15	The researcher is positioned as *active* in the research process; themes do not just 'emerge'.

Activity

▸ In 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)

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.