



Qualitative Data Analysis

Nicole Meng-Schneider and Dr Tara Capel

Quantitative Data

- Numeric
- countable and measurable

CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CI
Device Priv_1	Device Priv_2	Device Priv_3	Device Priv_4	Device Priv_5	Age	Gender	Gender_4	Ethnicity	Education	Occupation
Please indicate your level of concern about the use of your device for the following purpose.	Please indicate your level of concern about the use of your device for the following purpose.	Please indicate your level of concern about the use of your device for the following purpose.	Please indicate your level of concern about the use of your device for the following purpose.	Please indicate your level of concern about the use of your device for the following purpose.	How old are you?	How do you identify your gender?	How do you identify your gender?	What is your ethnicity?	What is your highest level of education?	What is your current occupation?
{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6	{\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6 {\"ImportId\":\"QID6
Very Concerned	Very Concerned	Very Concerned	Very Concerned	Very Concerned	35-44 year	Female		White	Post-grad	Employed
neutral	neutral	neutral	neutral	neutral	35-44 year	Male		White	College or	Employed
Not Concerned At	neutral	Very Concerned	Very Concerned	Very Concerned	55-64 year	Female		White	College or	Retired
Somewhat Not Cc	Somewhat Not Cc	Somewhat Not Cc	Somewhat Not Cc	Somewhat Not Cc	35-44 year	Male		White	Higher or s	Employed
Not Concerned At	Somewhat Conce	neutral	Very Concerned	Somewhat Not Cc	35-44 year	Female		White	Post-grad	Employed
neutral	Somewhat Conce	Somewhat Conce	Somewhat Conce	Somewhat Conce	25-34 year	Female		White	College or	Employed
Somewhat Not Cc	neutral	Somewhat Conce	Somewhat Conce	Somewhat Conce	25-34 year	Female		Black/Afric	College or	Employed
neutral	Somewhat Conce	Somewhat Conce	Somewhat Conce	Somewhat Conce	25-34 year	Female		Asian (Indi	Post-grad	Employed
Very Concerned	Very Concerned	Very Concerned	Very Concerned	Very Concerned	35-44 year	Female		White	College or	Self-emp
Somewhat Not Cc	Somewhat Conce	Somewhat Conce	Somewhat Conce	neutral	25-34 year	Female		White	Prefer not	Other
Somewhat Not Cc	neutral	Somewhat Conce	Somewhat Conce	Somewhat Conce	45-54 year	Female		White	Higher or s	Employed
Somewhat Conce	Very Concerned	neutral	Somewhat Conce	neutral	25-34 year	Female		White	College or	Employed
Very Concerned	Very Concerned	Very Concerned	Very Concerned	Very Concerned	35-44 year	Female		Mixed two	College or	Employed
Very Concerned	Very Concerned	Very Concerned	Very Concerned	Very Concerned	25-34 year	Male		Black/Afric	College or	Employed
Not Concerned At	neutral	Somewhat Not Cc	Somewhat Not Cc	neutral	25-34 year	Male		Black/Afric	College or	Employed

Qualitative Data

- descriptive, usually text-based
- interpretable

73	104	We found that there exist specific situations in which users do not desire voice interaction. Some participants stated that this was due to an awkward feeling when talking to a computer, which can be perceived as “admitting to being lazy” (P10) because a user does not carry out tasks themselves, delegating them to a computer instead.	Concerned to be overheard using device and be judged for what tasks are delegated to smart speaker.
3	60	And AO19 explains they would be uncomfortable if a visitor requested inappropriate services such as “a racist joke” or do anything criminal.	Owner considers possibility that bystander could use their device for something inappropriate
118	93	Our participants also expressed concerns about visitors, including their friends and less familiar people, such as party attendees. P16 expressed her concerns about party guests: “If I have a party with a whole bunch of acquaintances or strangers, I unplug it and put it away ... because if I leave [the speaker] out for strangers, [they] might steal our WiFi, and then weird things [will] happen.” [...] In addition, several participants reported putting the device away as a strategy to mitigate their concerns about visitors.	Concerns that strangers may use the speaker to gain access to wifi.
73	102	In our specific scenarios, interviewees showed awareness of several attack vectors, namely eavesdropping, Replay Attacks, and brute-force attacks on voice codes. Participants expected attackers to employ readily available devices such as microphones to capture a user’s interaction with a voice assistant. While most thought such an attack would need to happen in situ, a few interviewees were aware of voice sampling techniques using arbitrary audio of a user to produce adversarial samples.	Concerns that criminals may attack the voice assistant and try to exploit it.

Several ways to analyse qualitative data

- Thematic analysis
 - Affinity diagram
- Content coding
- Topic modeling

Thematic Analysis

Thematic Analysis

- Go through the data and identify “themes”, these themes become your outcomes.
- Affinity diagrams are one of the easiest ways to do thematic analysis with a group or by yourself
- Pros
 - Pulls the main concepts of the data out
 - Easy for someone else to understand
 - Themes are grounded in the data with clear examples
- Cons
 - Only works with a small amount of data
 - May require more than one person to improve validity

11

11

11

NOT OK
for visitors

OK for
visitors

spaces OK
to have
speaker

shared

bad

Spa
OK
Spa

Smart Speakers
in shared spaces

Experience
with sharing
Smart speakers

Social Norms

everything OK to do

OK for visitor

70k for visitors

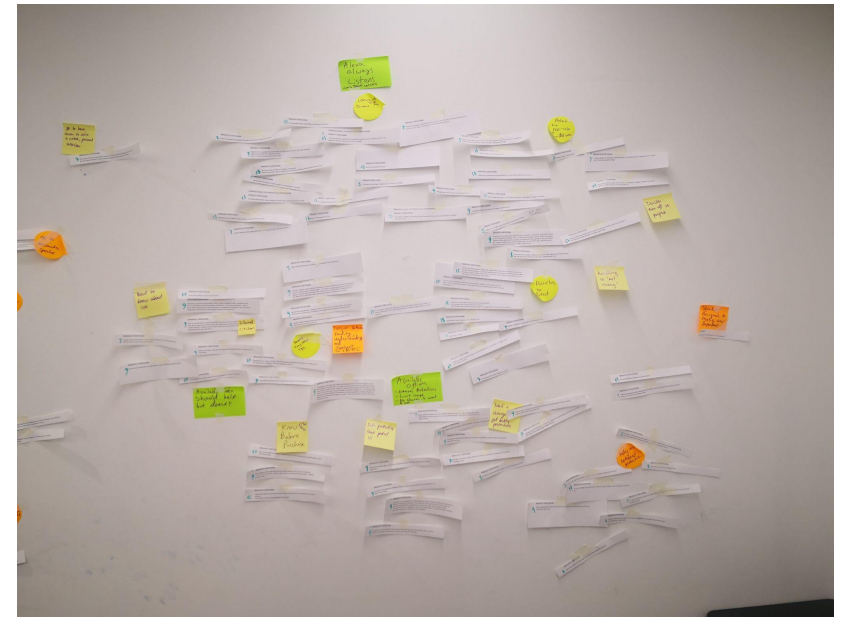
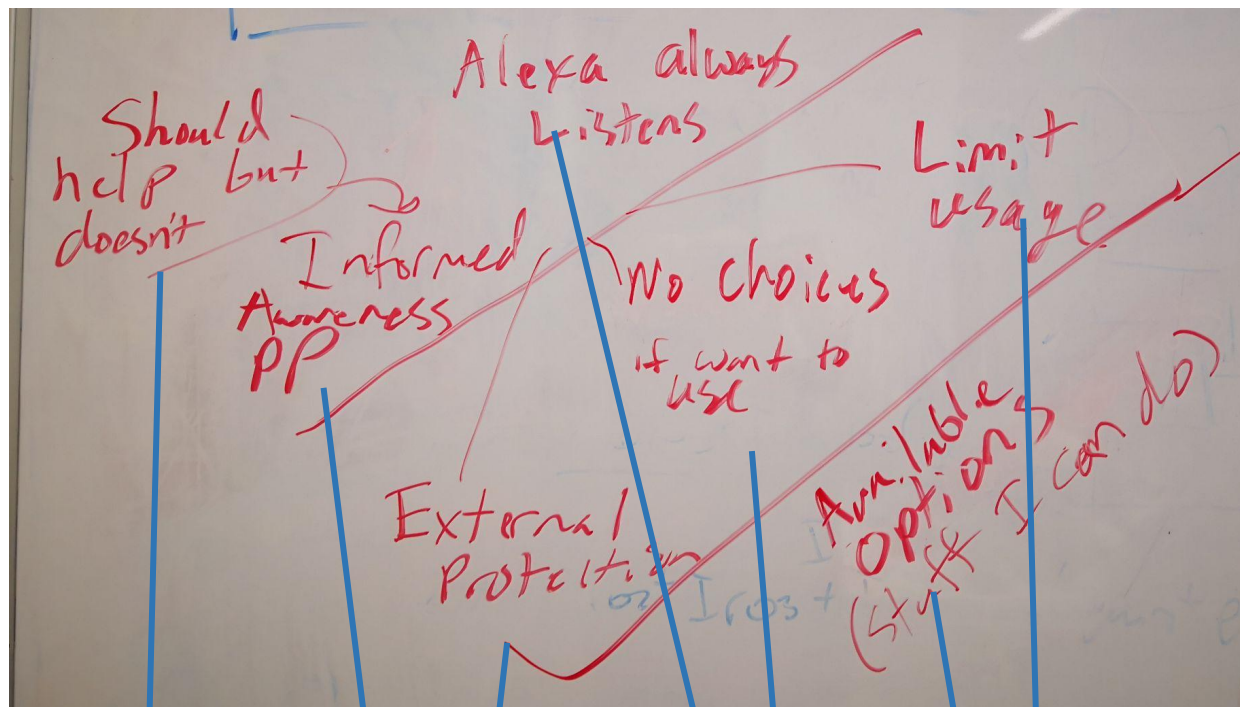
OK with person's spec

→ OK for other
person's speakers

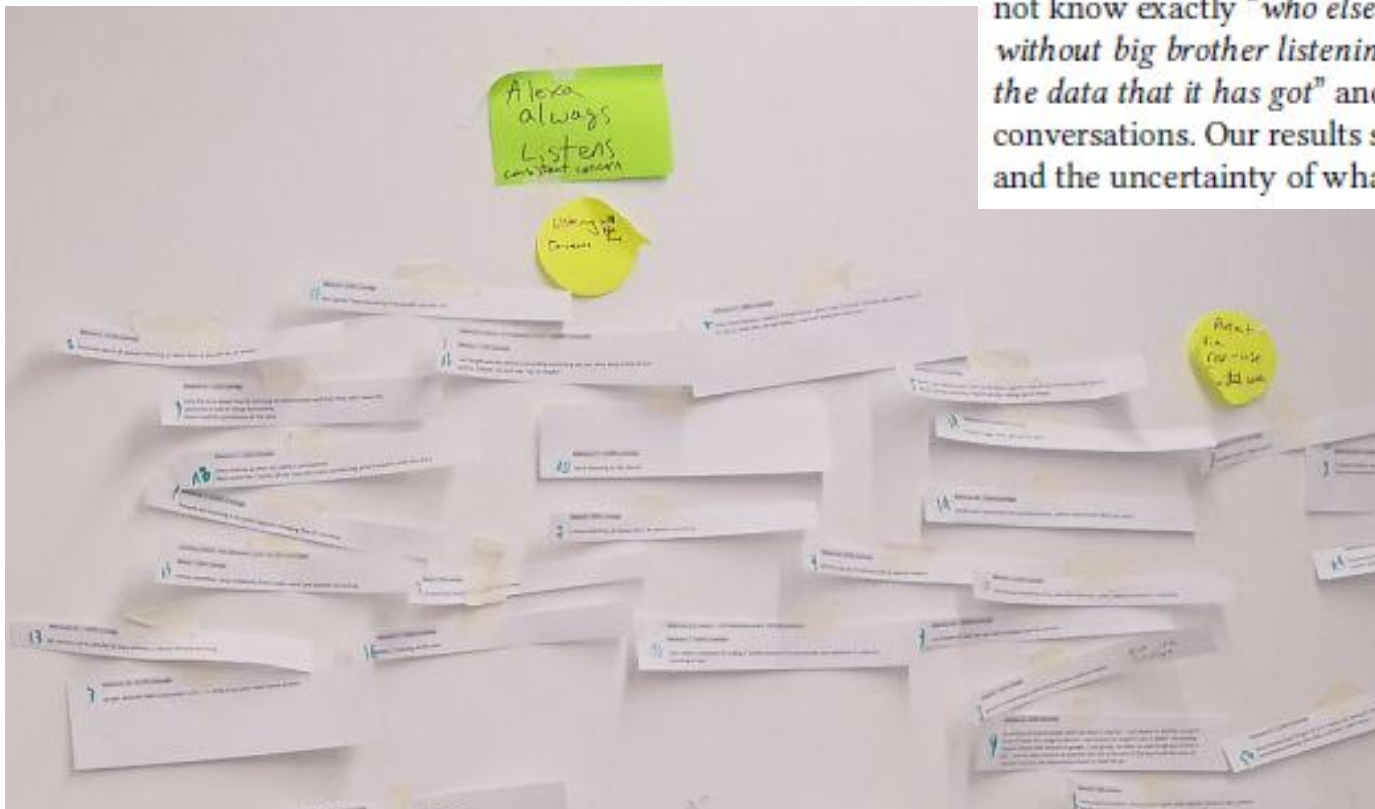
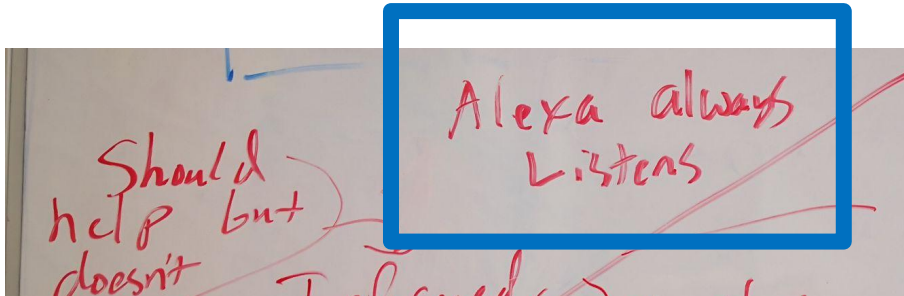
nothing ok $\rightarrow$ private space

- not ok bc:
 - own protection
 - social norms
 - politeness

not ok for visitors
→ examples



Results



Monitoring. A consistent concern emerging in most interviews and in prior work [7, 20, 30, 37, 42] was that the device is always recording. Almost all visitors and half of in-house users were sure that a smart speaker “*hears everything that people actually say*” (RO10) and is “*spying on everyone*” (RO1). Two owners were not sure whether it is listening when not activated through a wake word, but AO19 was convinced that their device always listens out for keywords. A main concern was that the device will intrude on private conversations. For V16 “*it crosses a line that they could be listening and could potentially see what you do*”. Whereas V7 was concerned that they as a user do not know exactly “*who else is monitoring, listening*” and that “*there isn’t any sort of privacy anymore without big brother listening*”. V13 wonders “*if it is listening all the time where is that information, the data that it has got*” and whether “*people [can] tap into people’s Alexa*”, eavesdropping on their conversations. Our results show that participants felt uncomfortable with potential audio recording and the uncertainty of what happens with the data.

Nicole Meng, Dilara Keküllüoğlu, and Kami Vaniea. 2021. Owning and Sharing: Privacy Perceptions of Smart Speaker Users. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 45 (April 2021), 29 pages.
<https://doi.org/10.1145/3449119>

Content Coding

Content Coding

- Different ways of summarising/labelling qualitative data to describe or infer key information
- “codes” are arranged in a codebook and can be used to pull out themes
- Top down vs bottom up approach
- Pros
 - Good way to “listen” to data and mitigate bias (esp. bottom up)
 - Systematic way to analyse unstructured qualitative data
 - Applicable for a wide variety of qualitative data (interviews, emails, secondary data, poems etc)
 - scalable for large amounts of text, but will take longer
- Cons
 - Depending on amount of data, might be a lengthy procedure
 - Best to do with a second coder to add validity

Brilliant overview

Miles, Matthew B. Qualitative Data Analysis : A Methods Sourcebook. International student edition. London: SAGE Publications, 2019. Print.

Available online in library.

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

Qualitative Data Analysis

A Methods Sourcebook



Example

“I decided that I wasn’t going to install the update because I have heard all the reviews online about how it generally makes your phone slower in every respect.”

Example

“I decided that I wasn't
going to install the
update because I have
heard all the reviews
online about how it
generally makes your
phone slower in every
respect.”

Not updating

Recommendations

Phone Speed

Coding types

- Different types of coding suitable for different key aspects
 - Descriptive Coding
 - In Vivo Coding
 - Process Coding
 - Emotion Coding
 - Concept Coding
 - Dramaturgical Coding
 - many more, check Miles et al.*
- Coding Units may range between phrases to complete documents

*Miles, Matthew B. Qualitative Data Analysis : A Methods Sourcebook. International student edition. London: SAGE Publications, 2019. Print.

Descriptive Coding (also Open Coding)

- Goal: Get a sense of how the subject expresses the key ideas in their own words
- Use the participants own words to summarize the text
- Go through and identify key phrases

¹ I asked the principal what the need for the new program was, and he responded that the students coming into the 9th grade were two years below grade level and that the old curriculum was ineffective. Through testing [the Nelson Reading Test] it was determined that students were growing academically only 5 or 6 months during the 10-month school year.

¹ MOTIVATION
FOR ADOPTION

In Vivo Coding

- Goal: label the data to understand what is there in a very general flexible way
- Label data to summarize it
- Typically a word or short phrase, often a noun

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 A LOT BETTER"

³ "STOPPED CARING"

Process Coding

- Goal: Understand the process, actions, or steps people engage in
- Code using only “-ing” words.

Well, that's one problem, that [my school is] pretty small, so ¹ if you say one thing to one person, and 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.

¹ SPREADING
RUMORS

² KNOWING WHAT
YOU SAID

Emotion Coding

- Goal: Identify feelings or emotions
- Identify emotion words and phrases

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

Concept Coding

- Goal: Identify the “bigger picture”
- Codes are nouns that symbolically represent a broader meaning or idea
- Concepts and conceptual processes are not something tangible
 - a clock can be seen, but the conceptual attribute is *time*
 - a church can be touched, but the conceptual attribute might be *spirituality* or *religion*
 - visiting Starbucks every morning could be a *morning ritual* or *feeding an addiction*

¹ What I love about America is the excess. You know, I'll walk into a coffee shop here [in Ireland] and all there'll be is milk. I go into a coffee shop in America, and there's 1%, 2%, skim milk, whole milk, half and half, you know? And the highways! There's two lanes at most here goin' one way, and over there you've got five, six lanes of highway all goin' in the same direction. We just don't have the cars or the land for that much highway. And the malls—you'll never see things like that over here. It's usually one shop at a time, but there you've got it all under one roof. Big stores, too, big stores.

¹ AMERICAN
“EXCESS”

Dramaturgical Coding

- Goal: Identify relationships, intra/interpersonal experiences, and human motives
- Identify objectives (obj), conflicts (con), tactics (tac), attitudes (att), emotions (emo), and subtexts (sub)

¹ There was a lot of pressure this year to “do more with less.” And that always ² frustrates me, because you don’t “do more with less”—you do *less with less*. So ³ if they’re expecting me to do more with less money and less resources, they’re not going to get it. And it’s not because I’m being snotty or passive-aggressive about this; ⁴ it’s simply that you can’t squeeze blood out of a turnip. There’s only so much you can do with what you have. ⁵ And yes, I’m spending some of my own money this year on classroom supplies because we don’t have enough to last us through the end of the year. ⁶ That’s just the way it is these days.

¹ CON: LESS
RESOURCES

² EMO:
FRUSTRATION

³ TAC:
RESISTANCE

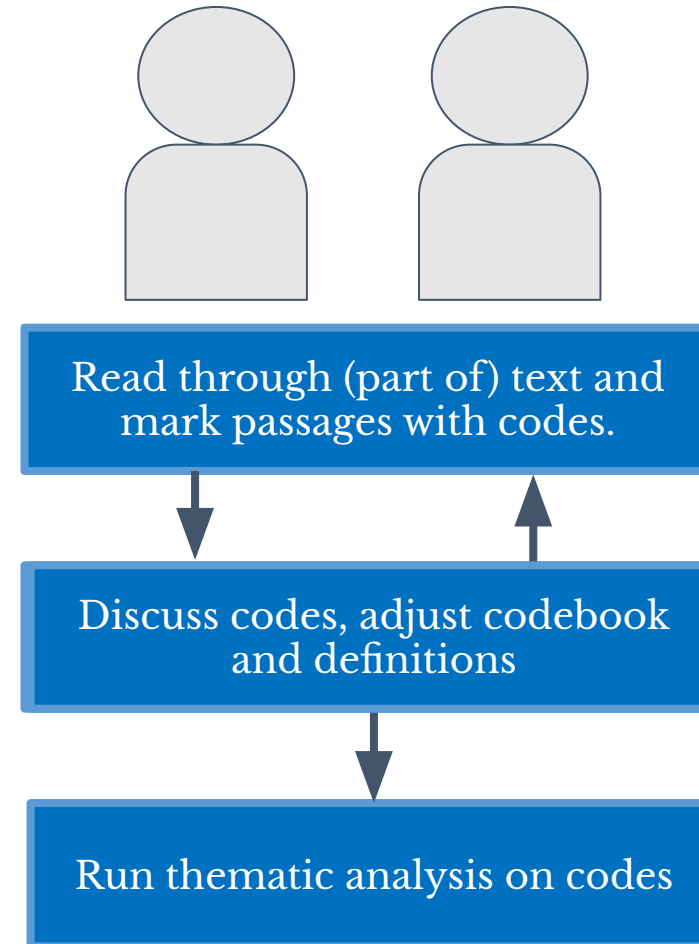
⁴ ATT:
LIMITATIONS

⁵ TAC:
SACRIFICING

⁶ ATT:
ACCEPTING “THE
WAY IT IS”

Coding Process

Typically done with more than one coder for validity



Example

“I decided that I wasn't
going to install the
update because I have
heard all the reviews
online about how it
generally makes your
phone slower in every
respect.”

Not updating

Recommendations

Phone Speed

Coding Process

Code Book

- List of all codes that have been used, often becomes more and more structured during the coding process.
 - Sometimes we go through a second coding cycle with a different coding method
1. Awareness
 2. Deciding
 3. Preparation
 4. Installation
 5. Troubleshooting
 6. Expected post state
 7. Post state
- Installation
 - Time
 - Cost
 - Resources
 - Problems
 - Failure
 - Restart
 - Bundled software

Example

“I decided that I wasn't
going to install the
update because I have
heard all the reviews
online about how it
generally makes your
phone slower in every
respect.”

Not updating

Recommendations

Phone Speed

Example

“I decided that I wasn't
going to install the
update because I have
heard all the reviews
online about how it
generally makes your
phone slower in every
respect.”

Not updating

Deciding: did not
update

Recommendations

Deciding: research

Phone Speed

Expected post
state: performance
: worse

Results

Deciding to update or not update

- Updating is important
- The old version had problems
- Reviews
- Features, performance, resources, bugs
- Wait out the problems
- Worth the bother?
- Updates could contain viruses



Topic Analysis

Topic Modelling

- a form of text mining which provides a simple way to analyse large volumes of unlabelled text.
- e.g. computer goes through the text and creates an word cloud



Topic Modelling

- Always clean data:
 - Remove common words/ stop words (e.g. “the”, “in”, “to”...) which may cover every part of the text
 - Removing punctuation, digits, and white space
 - Lower Case all words
 - Potentially stem words to identify common roots
- Connects words in clusters “topic” – a cluster of words that frequently occur together
- Groups of words that tend to happen very frequently together
- What percentage of the documents are dealing with each topic
- Common method: Latent Dirichlet allocation (LDA)*
 - User specifies number of topics and algorithm creates associations between documents to create clusters

Topic Modelling Output

Topic		Keywords	Example tweet
Having a baby	a	family, congratulation, news, blessing, member, addition, baby, girl, mommy, daddy, pregnancy, delivery, healthy_pregnancy, motherhood, boy, shower, gender_reveal	@username I'm so happy for you! Can't wait for the baby shower!
Travel		enjoy, time, rest, trip, weekend, summer, travel, visit, vacation, relax, holiday, london, japan, flight, safe_travel, korea, europe, germany, chicago, italy	@username so happy for you enjoy your trip [...]
Relationship start		person, relationship, boyfriend, girlfriend, bf, partner, keeper, gf, long_distance	A boyfriend like this is a keeper I'm happy for you sis @username
Cancer (in family)	(in family)	mom, family, sister, dad, fight, brother, cancer, die, beat, lose, stage, grandma, uncle, cousin, warrior, nephew, aunt, fighter, niece, battle, survivor, monster, treatment, grandpa	@username So happy for you! May God keep you cancer free.
Birthday		day, birthday, today, celebrate, gift, bday, present, belated_birthday	@username CONGRATS AND HAPPY BIRTHDAY IM SO HAPPY FOR YOU
Surgery		hope, continue, pray, stay, recovery, health, take_care, heal, recover, speedy_recovery, rest, improve, surgery	@username [...] hope you have ease in your recovery.
Graduation		congratulation, success, work, achievement, accomplishment, celebrate, future, earn, cheer, graduation	[...] Happy graduation @username
Mental health		deal, pain, struggle, problem, doctor, issue, anxiety, mental_health, fear, therapy, overcome, depression, surgery, brain, stress, relief, med	@username I really hope you overcome your anxiety [...]
Familial matters		parent, kid, son, child, daughter, mother, family, mom, father, dad, bear, wife, raise, age, miracle, birth, husband, awareness, sibling, carry, grandmother, adopt	@username [...] that you have a grandchild too. [...]
Marriage		congratulation, wedding, marry, wife, husband, marriage, invite, day, engage, dress, bride, honor, ring, engagement, hubby, anniversary, party, honeymoon, honour, propose, divorce, fiancé	@username [...] The wedding will be fantastic though! [...]
Moving		move, place, home, house, leave, visit, fall, room, space, settle, city, town, area, apartment, land, pack	@username Everyone needs to leave home at one point. I feel you sis. [...]
LGBTQ-related		speak, people, woman, part, trust, realize, process, lie, power, figure, faith, community, truth, accept, idea, pride, gay, doubt, gender, embrace, tran	Congratulations to my favorite lesbians! [...]

Table 1: Life event topics from *HFY-LE* and keywords selected from the 30 most probable words for each topic. Example tweets shown with usernames blinded and some content removed for privacy.

Qualitative Data Analysis

Nicole Meng-Schneider