Lecture 8: Qualitative Analysis

CS 181W   Fall 2022

Review: Types of studies

- **Interviews:** conversations with individuals
- **Focus groups:** discussions with groups
- **Surveys:** asynchronous questions

- **Experimental Studies:** randomized multi-condition studies
- **Usability Testing:** observations of tool use
- **Cognitive Walkthrough:** expert evaluation

- **Diary Studies:** contemporary record of real-world behavior
- **Observational Studies:** records of behavior in the wild

- Mixed-methods studies

Examples of qualitative data

- Interviews

- Open-ended survey responses

- Drawings

- Photos

- Videos

- Social media posts

- Diary entries

- Comments or reviews posted in online forums

- Chat transcripts

- Think-aloud transcripts

- Researcher's notes

Goals of qualitative data analysis

- Make sense of unstructured data

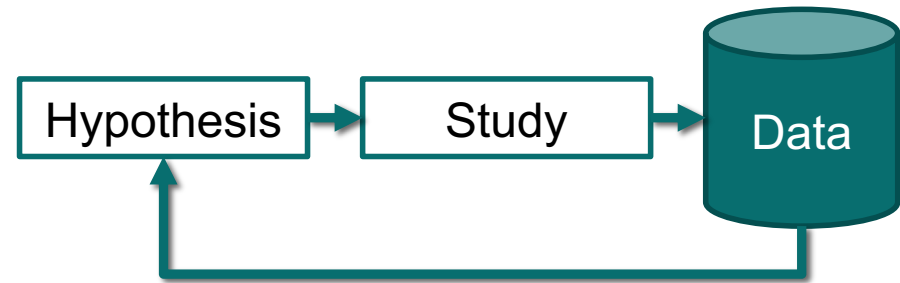- Develop consistent and reliable interpretations of data

Thematic Coding Approaches

Grounded Theory

```
┌─────────┐      ╔═════════╗      ┌─────────┐
│  Study  │ ───▶ ║  Data   ║ ───▶ │ Theory  │
└─────────┘      ╚═════════╝      └─────────┘
```
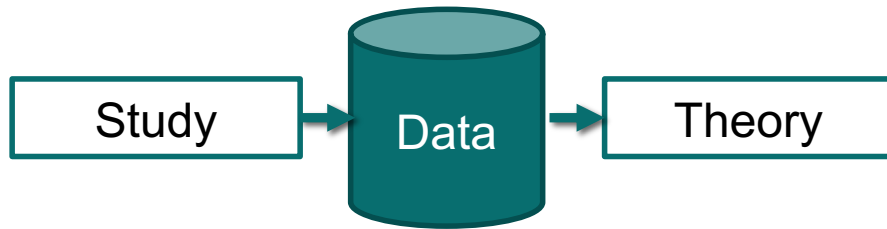
- Emergent coding
  - don't start with a theory or hypothesis
  - look for interesting concepts or ideas and refine as you go

Experimental Research

```
┌────────────┐      ┌─────────┐      ╔═════════╗
│ Hypothesis │ ───▶ │  Study  │ ───▶ ║  Data   ║
└────────────┘      └─────────┘      ╚═════════╝
      ▲                                    │
      └────────────────────────────────────┘
```

- A priori coding
  - start with established theory or hypothesis
  - theory typically from prior work

Creating a codebook

- Emergent coding
  - Read through (subset of) the data
  - Read through again and this time identify interesting things to code
  - Combine and refine codes, perhaps in a hierarchy
  - Update iteratively as you code data

- A priori coding
  - Make a list of all the possible categories you might have based on prior theory

- Codebooks are often hierarchical

- Too many codes can make it hard to code reliably and difficult to make sense of your data

Example: Emergent coding

- Let's using emergent coding to develop a codebook for drawings of privacy
  - labels
  - definitions
  - examples
  - categories/concepts

Example: A priori coding

collection	surveillance	people watching	eyes, law enforce.
	interrogation	coerced acquisition	theft, stealing
processing	aggregation	gathering info	
	identification	link data to person	ID card
	insecurity	data not secure	passwords, enc
	secondary use	data reuse	
	exclusion	failed notice	blindfolded user
dissemination	breach of conf	data breach	hacker
	disclosure	data sharing	arrows
	exposure	embarrassing data	bathroom
	incr. accessibility	easier access to data	
	appropriation	id theft	
	distortion	manipulation of data	
invasion	intrusion	incursion into life	personal bubble
	decision interfere.	impose on decisions	

Exercise: Qualitative Coding

Reliability

- Stability
  - also called **intra-coder** reliability
  - examines whether the same coder rates the data in the same way throughout the coding process

- Reproducibility
  - also called **inter-coder** reliability or investigator triangulation
  - examines whether different coders code the same data in a consistent way

Reliability measures

- Percent agreement

$$\%\text{agreement} = \frac{\text{the number of cases coded the same way by multiple coders}}{\text{the total number of cases}}$$

- Cohen's Kappa:

$$K = \frac{P_a - P_c}{1 - P_c}$$

Example: Cohen's Kappa

		Coder 2			
		Surveillance	**Insecurity**	**Intrusion**	**Marginal Total**
Coder 1	**Surveillance**	.26	.07	.04	.37
	Insecurity	.04	.12	.01	.17
	Intrusion	.09	.02	.35	.46
	Marginal Total	.39	.21	.40	

$$P_a = .26 + .12 + .35 = .73$$

$$P_C = (.37 * .39) + (.17 * .21) + (.40 * .46) = .364$$

$$K = \frac{.73 - .364}{1 - .364} = .575$$

Increasing reliability

- Multiple coders should each code a small number of items independently, compare answers, and resolve conflicts
  - This may lead to changes in the codebook
  - Process may need to be repeated several times until reliability is sufficient
  - For any items double coded, the resolved codes are the ones that should be reported and used in analysis
- Coders can collaboratively discuss and code everything together
- One person can code everything and another can review the codes and suggest changes for the two coders to discuss

Exercise: Measuring Reliability

Qualitative Analysis