Qualitative analysis

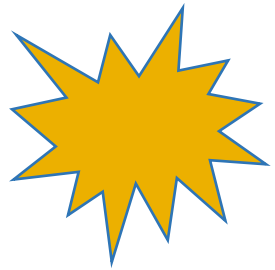
Introduction to Data Science

4th lecture

Izv. prof. dr. sc. Ana Sović Kržić

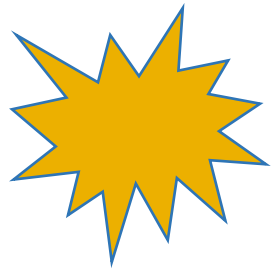
2025/2026

Contents



- Qualitative analysis
- Types of qualitative analyses
- Software for qualitative analysis
- Visualization of qualitative data

What is qualitative analysis?



- analysis of non-numerical data (texts, video, audio)
- to gain in-depth insights into a problem or generate new ideas for research
- it is used to understand how people experience the world

- research bias, Hawthorne effect, observer bias, recall bias, social desirability bias

When is qualitative analysis used?

- **Documentary research** – collection of existing data in the form of text, images, audio and video material
- **Interview** – personal questioning of one participant
- **Focus group** – asking questions and leading a discussion among a group of participants
- **Ethnography** – immersing yourself in a group to understand its culture
- **Observational research** – recording what you see, hear, detailed notes from the field
- **Questionnaire** – open questions

Steps of qualitative analysis

1. **Preparation and organization of data** → e.g. transcribing interviews, writing notes from the field
2. **Data review and exploration** → looking for patterns and repeated ideas in the data
3. **Developing a data coding system** → based on initial ideas, establish a set of codes that can be applied to categorize data
4. **Assigning codes to data** → e.g. in a survey, going through the responses of each participant and marking them with codes in a table. It is allowed to create new codes "on the fly" that will be added to the system later
5. **Identifying recurring themes** → linking codes together into overarching themes

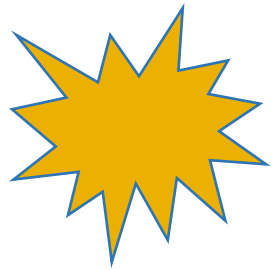
Advantages and disadvantages

Advantages	Disadvantages
Flexibility → the data collection and analysis process can be adjusted as new ideas emerge	Unreliability → possible uncontrolled factors affecting the data
Natural settings → data collection takes place in the real world	Subjectivity → the research cannot be repeated, the researcher decides what is important → interpretations of the same data may differ
Meaningful insights → descriptions of experiences, feelings and perceptions → can be used in designing, testing or improving products	Limited generalizability → small samples → difficult to draw generalized conclusions → data may be biased and unrepresentative
Generating new ideas → discovering new problems or opportunities	Labour-intensive → data analysis often needs to be performed manually

Types of qualitative analysis

Approach	When to use
Content analysis	Describing and categorizing common words, phrases and ideas
Discourse analysis	How language is used to achieve effects in specific contexts
Thematic analysis	Identifying and interpreting patterns and themes
Textual analysis	Examining the content, structure and design of the text

Content analysis (1)



- **a research method used to identify patterns in recorded communication**
- data is collected from written, oral or visual texts: books, newspapers, magazines, speeches and interviews, web content, social networks, photos and movies
- **analysis approach**: categorization (coding) of words, topics and concepts within the texts and analysis of the results

Content analysis (2)

- **quantitative content analysis**: the number of occurrences of a word, phrase, object or concept

research: the importance of the unemployment problem in the pre-election campaign

analysis of the number of times the words "unemployment", "job", "work" appear in articles during the campaign to detect changes over time or differences between candidates

- **qualitative content analysis**: analysis of meaning, semantic links of words and concepts

locate the word "unemployment" in the texts

study which words and phrases are found near this word (e.g. "inequality", "emigration")

analyze the meaning of the relationship of these words in order to understand the objectives of different campaigns

Implementation of content analysis (1)

1. Ask a direct research question

Is there a difference in how the media presents younger politicians compared to older ones in terms of honesty?

2. Select texts to be analyzed

- medium: newspaper articles, speeches, websites...
- type: opinions, political campaign speeches
- inclusion or exclusion criteria: only articles that mention a specific event, speeches by a specific politician
- additional parameters: time range, location...
- if necessary, do not use all articles, but a selected sample

Texts from the web portal will be analyzed between 2021 and 2024.

Considering the large number of such articles, only the 3 largest portals will be used.

Implementation of content analysis (2)

3. Define units of meaning and categories of analysis

- **Units of meaning** to be coded:
  - the frequency of selected words and phrases
  - characteristics of the people who wrote or appear in the texts
  - presence and positioning of images
  - treatment of topics and concepts
- **The set of categories** for coding:
  - objective characteristics (e.g. 50-60 years old, doctor, parent)
  - conceptual (eg reliable, corrupt, conservative)

Units of meaning: politicians appearing in articles; words/phrases used to describe them

Categories: age, honesty

Additional categories: political party, marital status

Implementation of content analysis (3)

4. Develop a set of coding rules

- organizing units of meaning into defined categories
- it is important to clearly define the rules about what will and will not be included in a particular category → ensuring coding consistency and transparency

category "younger politician" → younger politician by age or seniority in politics?

category ”honesty" → words and phrases that will represent honesty, e.g. trust, credence

Implementation of content analysis (4)

5. Encode text based on rules

- go through all the texts
- record all relevant data in the appropriate categories

Read each article from the web portal selected in the sample → note the characteristics of each mentioned politician → note all words and phrases related to “honesty”, which are related to the mentioned politician

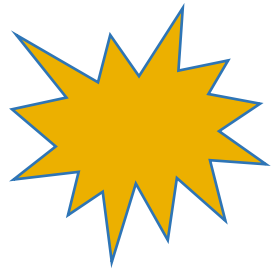
6. Analyze the results and draw conclusions

- examine the collected data → look for patterns from which conclusions can be drawn, which answer the research question
- possible to use statistical analysis (eg correlation), use discussion to interpret the results

Results: words and expressions related to „honesty” appeared more often in the same sentence as ”younger politicians"

Conclusion: web portals present younger politicians as more honest → this could influence readers' perception of younger people in politics

Discourse analysis (1)



- the study of written or spoken language in relation to its social context
- **understand how language is used in real life situations**
- examining how language works and how meaning is created in different social contexts
- gain an understanding of social groups and how they communicate
- often used in: humanities and social sciences (linguistics, sociology, anthropology, psychology, cultural research)
- discourse analysis **emphasizes the contextual meaning of language → the study of larger parts of language: whole conversations, texts or collections of texts**

how people use language to achieve certain effects (e.g. to build trust, create suspicion, evoke emotion or manage conflict)

Discourse analysis (2)

- **Text**:
  - Books, newspapers, magazines
  - Marketing material (brochures and ads)
  - Business and government documents
  - Websites, forums, posts on social networks
  - Interviews and conversations
  - Non-verbal aspects of communication (tone, gestures)

Discourse analysis (3)

	What is analyzed	What they reveal
vocabulary	words and phrases	ideological associations, formality, euphemistic and metaphorical content
grammar	the way sentences are constructed: verb tenses, active or passive construction, use of imperatives and questions	intended meaning
structure	text structure	the way in which emphasis is created or a narrative is constructed
genre	political speeches, tabloid newspaper articles	conventions and communicative goals
non-verbal communication	non-verbal aspects of speech: tone of voice, pauses, gestures, sounds like "um"	speaker's intentions, attitudes and emotions
conversational codes	interaction between people in a conversation: taking turns, interruptions, listener responses	cultural conventions and social roles

Implementation of discourse analysis (1)

1. Define the research question and choose the content of the analysis

- can be applied to smaller or large quantities of material

Research question: how the change of regime from dictatorship to democracy affected the public relations rhetoric of companies in the country

Content of the analysis - texts: marketing material of the 10 largest companies within five years of regime change

Implementation of discourse analysis (2)

2. Gather information and theory about the context

- determine the social and historical context in which the material was produced and intended for reading
- gather factual details about when and where content was created, who wrote it, who published it and to whom it was distributed
- review the literature and build a theoretical framework that will guide the analysis

Collecting information about: politics and history of the country, the affairs of the companies studied, the theory of democratic transitions and the relationship between government and business

Implementation of discourse analysis (3)

3. Analyze content for themes and patterns

- examination of different elements: words, sentences, paragraphs, overall structure
- relating them to attributes, themes and patterns relevant to the research question

Analyzing the material in search of formulations and statements that refer to authoritarian and democratic political ideologies: attitudes towards government, popular opinion

Implementation of discourse analysis (4)

4. Review the results and draw conclusions

- reflect on the results to examine the function and meaning of the used language
- consider the analysis in relation to the wider context → draw conclusions that answer the research question

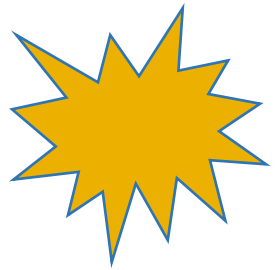
Results:

- material published before the regime change used language that emphasized the quality of its products
- material published after the transition to a democratic regime emphasizes the needs and values of consumers

Analysis of the wider context: comparison of results with research on the ideology of political regimes

Conclusion: the changing political context has shaped the communication strategies of national companies

Thematic analysis



- examining the data to **identify common themes, ideas, and recurring patterns of meaning**
- goal: to find out about people's attitudes, opinions, knowledge, experiences or values
- data collection from a group of relevant participants
- **texts**: a set of texts, interviews, transcripts, profiles from social networks, survey answers
- enables easier access to large data sets by sorting them into broad themes

Examples of research questions:

- How do patients perceive doctors in the hospital environment?
- What are the experiences of young women on dating sites?
- What are opinions of non-experts about climate change?

Approaches to thematic analysis

- **Approaches:**
  - **Inductive approach** → involves letting the data determine the themes
  - **Deductive approach** → involves arriving at data with some preconceived themes expected to emerge, based on theory or existing knowledge

Does theoretical framework give a strong idea of what kind of themes are expect in the data (deductive), or will new framework will be develop based on new founding (inductive)?

- **Approaches:**
  - **Semantic approach** → includes analysis of explicit data content
  - **Latent approach** → involves reading the subtext and assumptions underlying the data

Am I interested in people's opinions (semantic) or what their statements reveal about their assumptions and social context (latent)?

Implementation of thematic analysis (1)

1. Getting to know the data

- a detailed **overview of all the collected data**
- includes: audio transcription, text reading, initial note taking, general data browsing

We investigate the perception of climate change among residents over 50 years old. The data was collected through a series of interviews.

Implementation of thematic analysis (2)

2. Coding

- **highlighting parts of the text** (phrases or sentences) with different colors or marks
- devising abbreviated labels (**codes**) to describe their content (ideas, feelings,...)
- it is allowed to add new codes while going through the text
- **put all the data into groups** identified by the code
- the codes allow to get a concise overview of the main points and common meanings that recur throughout the data

Respondent's answer	Code
Personally, I'm not sure. I think the climate is changing, of course, but I don't know why or how. People say you should trust the experts, but who's to say they don't have their own reasons for pushing this story? I'm not saying they're wrong, I'm just saying there are reasons not to trust them 100%. The facts keep changing - it used to be called global warming.	Uncertainty Confirmation of climate change Distrust in experts Change in terminology

Implementation of thematic analysis (3)

3. Generating themes

- reviewing the codes that have been produced → **identifying a pattern**
- **themes** → broader than codes → usually a combination of several codes in one theme
- **rejecting codes** → if some codes are too vague or insufficiently relevant (because they do not appear often enough in the data)
- create potential themes that say something useful about the data for the research

Codes	Themes
• Uncertainty • Leave it to the experts • Alternative explanations	Uncertainty
• Change of terminology • Distrust of scientists • Resentment towards experts • Fear of state control	Distrust in experts
• Incorrect facts • Misunderstanding of science • Biased media sources	Disinformation

Implementation of thematic analysis (4)

4. Reviewing themes

- checking that **the themes are useful and accurate** representations of the data
- returning to the data set and comparing the themes with the original data
- asking questions: is something missing? Are the mentioned themes really present in the data? Can anything be changed to make the themes fit the data better?
- if there is a problem with themes → combine, split, discard, create new themes

Looking at the data, we conclude that ”change of terminology" fits better under the theme of "uncertainty" than under "distrust in experts" → because the data marked with this code includes confusion, not necessarily distrust

Implementation of thematic analysis (5)

5. Defining and naming themes

- articulating exactly **what we mean by each theme** and understanding how this helps understand the data
- coming up with a **concise and easy-to-understand name** for each theme

Who exactly do we mean by "experts" in the theme "distrust in experts"?

Is a better name for this theme "distrust in authority" or "conspiracy thinking"?

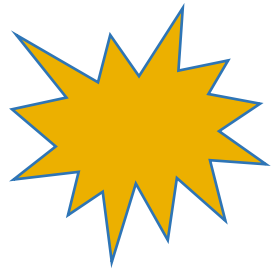
Implementation of thematic analysis (6)

6. Writing

- introduction → research question, objectives, approach
- methodology → how we collected data, how we conducted the thematic analysis
- results → each theme separately: how often the themes appear and what they mean, include examples from the data as evidence
- conclusion → explains the main conclusions, shows how the analysis answered the research question

Conclusion: Conspiratorial thinking about climate change is widespread among residents over 50. This indicates the uncertainty with which respondents view the issue. In the results, discuss the role of misinformation in respondents' perceptions.

Textual analysis



- **research method used to describe, interpret and understand text**
- literal meaning of the text, symbolism, assumptions, values
- often the goal is to connect the text with a wider social, cultural, political or artistic context
- **text**: book, e-mail, transcribed conversation, film, painting, sculpture, space

- Short story analysis → images, narrative perspective, text structure
- Film analysis → dialogues, cinematography, use of sound
- Building → architectural features, the way visitors move around it
- Game rules → a type of behavior designed to motivate players

Textual analysis in cultural research

- elucidating something from the social context of the cultural object being investigated
- **existing theory is linked to new texts**
- **text**: music videos, data from social networks, advertisements on billboards...
- **analysis**:
  - Choice of words
  - Design elements
  - Text location
  - Target audience
  - Relationship with other texts

Textual analysis in social research

- **empirical conclusions about social relationships**
- **text**: transcripts of interviews, surveys
- **methods of analysis**: content analysis, thematic analysis and discourse analysis
- often a quantitative approach → features of texts are measured numerically

how often certain words are repeated in posts on social networks

which colors appear in product ads for different demographics

Textual analysis in literature

- almost all works in literature involve in-depth analysis of texts
- to understand **how certain text elements contribute to the meaning of the text**
- explores the intended meaning of a text, potentially unintended connections between different texts, what the text reveals about the context in which it was written, analyzes an old text in a new way...
- **text**: novel, poem, story, drama
- emphasis on **deliberately constructed text elements**: rhyme, narrative perspective in a novel...

Examples of qualitative analysis

TABLE II. QUALITATIVE COMMENTS OF THE STUDENTS. NUMBERS IN BRACKETS REPRESENT HOW MANY STUDENTS GAVE SIMILAR COMMENTS.

	Question (1) Game vision	Question (2) Robot vision	Question (3) Game upgrade	Question (4) Robot upgrade
Taw choice	-	-	choose taw when the game starts	-
Collecting	collecting opponents' marbles	-	collecting opponents' marbles; choosing marbles at the start of the game; new marbles with levels	-
Aesthetics	own marble design; marble holder appearance; marble colors; sounds (cheering and marble collision)	non-humanoid robot (5); humanoid robot (2); color; robot-skin; glass fence	the different look of the marbles (4)	clear and modern design; color; a humanoid robot
Perspective	first-person, third-person (side-view), top view (2D)	-	top view	smartphone (camera) should be closer to the marbles
Rules	the round ring is drawn by a player; billiard-like holes; barriers; marbles formation (circle, cross, random)	ring-fence and four spots for marbles	changing position after shooting; billiards-like with holes; drawing the ring; terrain choice	more marbles
Detecting and scoring points	points for marble knocking; winning the game; exchanging points for marbles; double-points marbles; the scoreboard	buying better marbles for points; hit and knock detection: metal detector and metal marbles; a chip in a marble; touch detector in the holes	winning points; more levels; scoreboard; "end of the game" mark (5)	positive feedback when the marble is knocked, like a sound "Great!" or a screen that shows which marbles were knocked
Control	billiard, slingshot, or mouse control; marble and terrain physics; moving around the ring	different control: auto control (5), manual control (8), joystick (5), buttons (4), remote control, smartphone app; ways of shooting: cannon, slingshot, auto-reloading marbles, marbles slide, more precision, air-pump; moving (2); aiming: laser, camera, target on a digital map	more physics depending on shooting power; level-depending shooting precision and difficulty	more precise movements (3); more power and faster shooting (3); faster response (3); auto-control (1); change of the turn for another player with sound, led light, or just to lock another player while the first one is shooting; aiming: laser (2), more realistic gunpoint
Multiplayer	human vs. human (10); human vs PC (5)	human vs. autonomous robot (4); human vs. human controlling robots (3); autonomous robot vs. autonomous robot; no human (1), solo (1)	multiplayer (4)	-

Software for qualitative analysis

- nVivo: https://lumivero.com/products/nvivo/
- Atlas.ti: https://atlasti.com
- Deedoose: https://www.dedoose.com
- MaxQDA: https://www.maxqda.com
- Taguette: https://www.taguette.org
- QualCoder (Python): https://qualcoder.wordpress.com/reports/

Qualitative analysis in Excel

- counting the most frequently repeated words
- searching for the most important words and phrases
- determination of themes using K-means
- sentiment analysis

https://medium.com/@steflaihk/how-to-use-python-and-ai-in-qualitative-research-in-excel-96343f4dbdb6

Visualization of qualitative data

- word cloud
- bubbles
- graphic timelines
- icon next to the description
- heat map
- mind map
- illustrative diagrams
- display of open questions next to closed ones
- photos of the participants
- color coding

Word cloud

- word size indicates the frequency of occurrence of that word
- known software: Wordle, Tagxedo

Bubbles and spider webs

- word frequency is represented by the size of the bubble

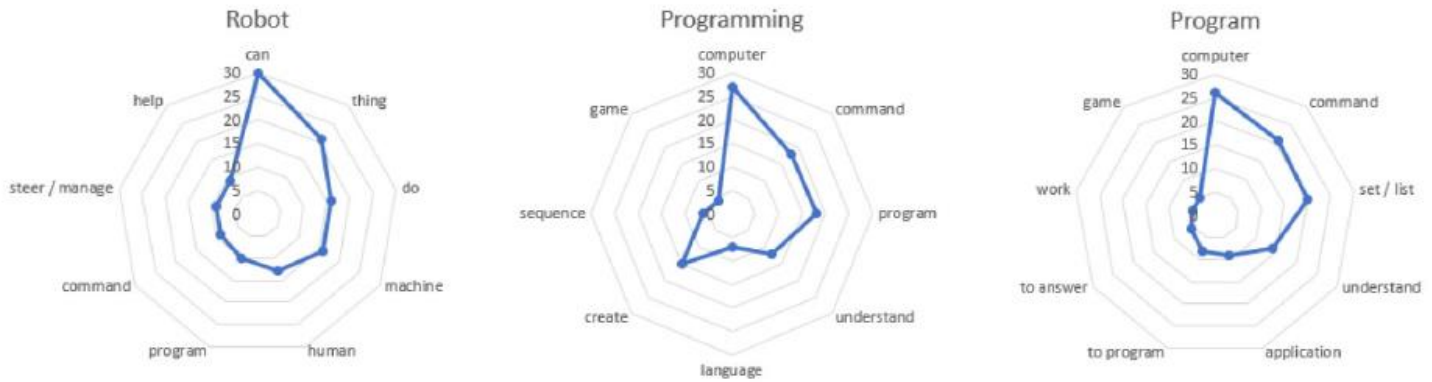
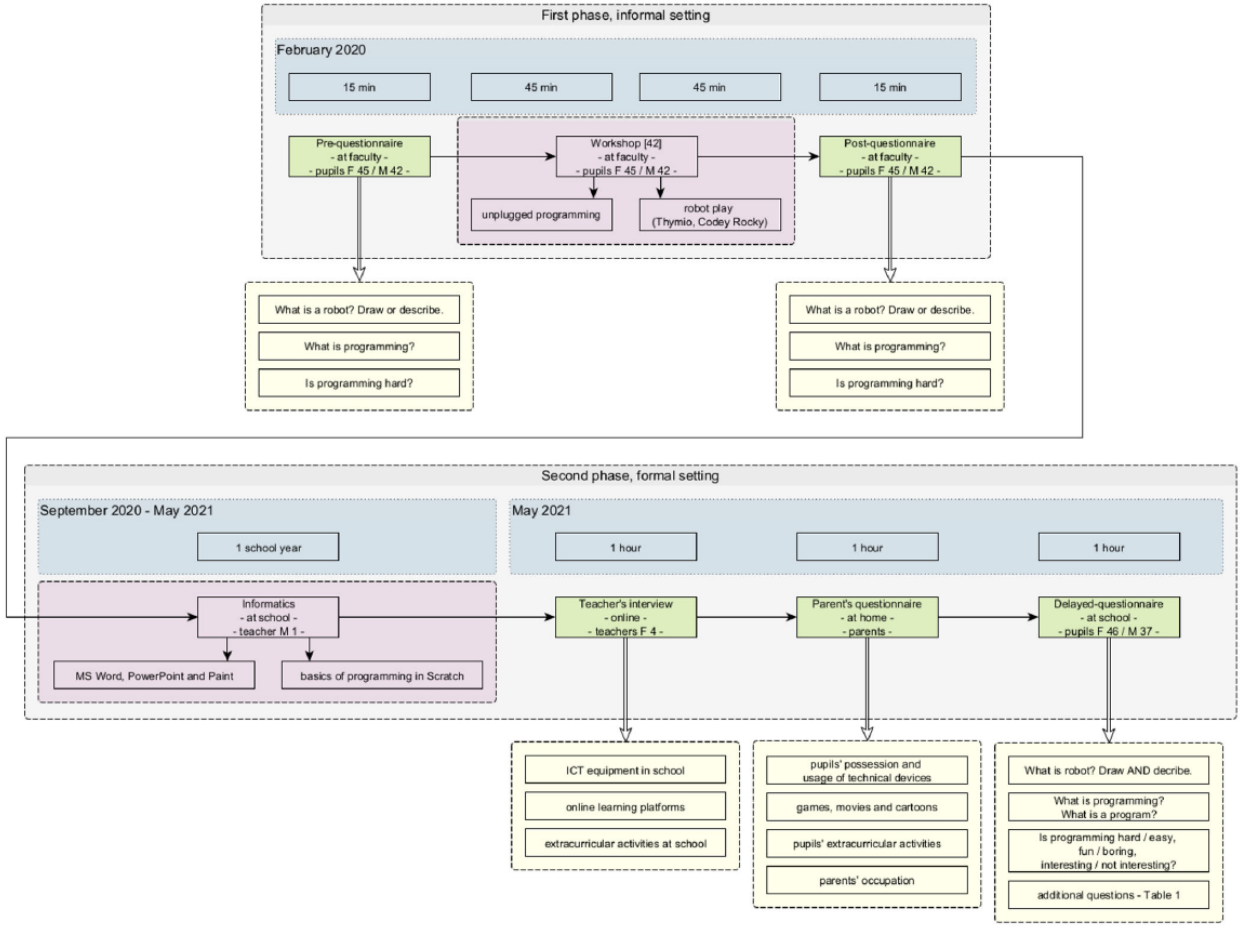


Fig. 3. Most common words in each of the definitions in the delayed questionnaire.

Graphical timelines

- timelines with the addition of pictorial illustrations, diagrams, photos...



Icons next to the description

- using icons instead of long descriptive texts

Heat map

- the intensity or frequency of the data is additionally represented by color
- add a legend that explains the meaning of individual colors
- helps to discover trends in data

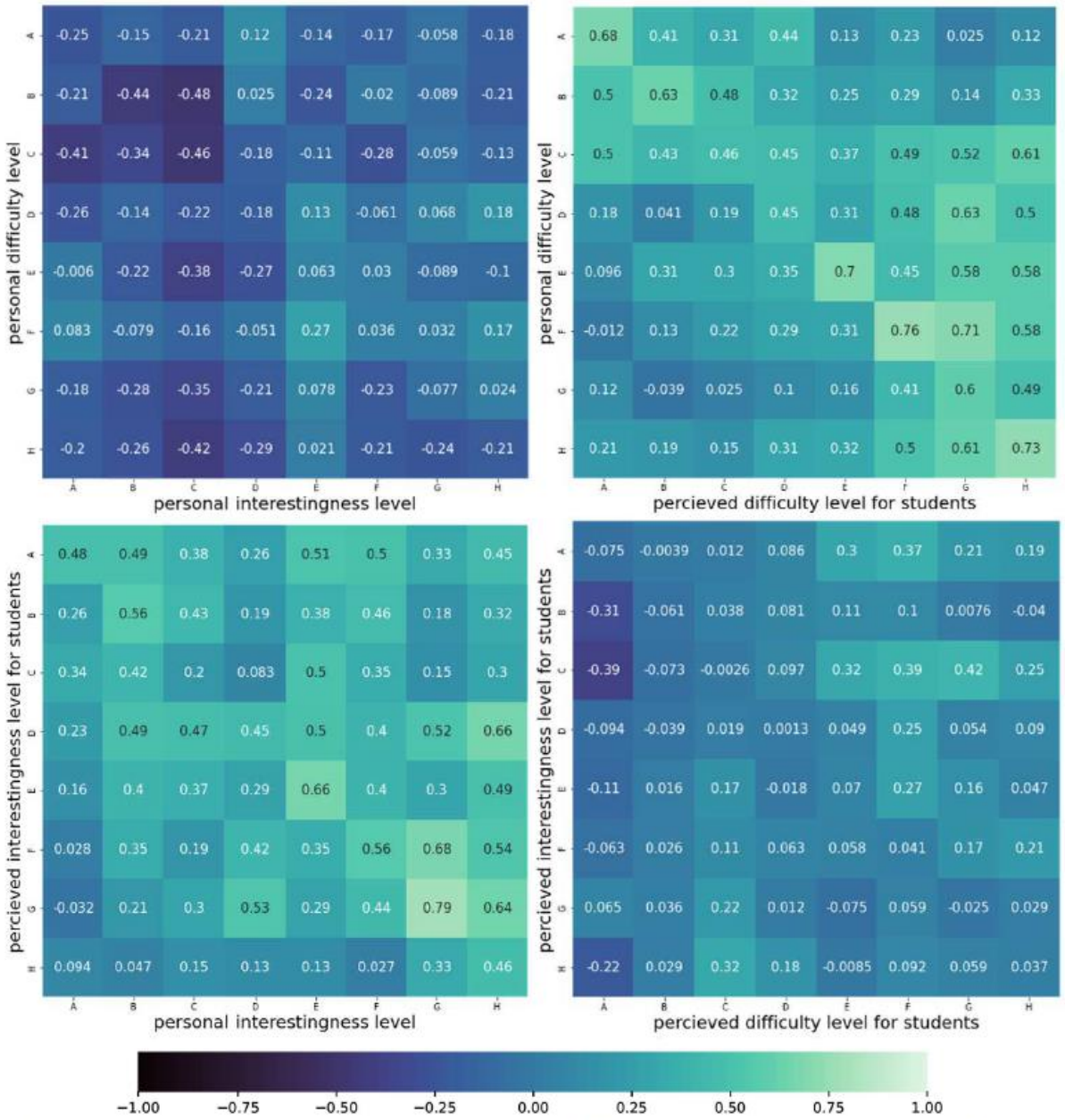


Fig. 4. Spearman's correlation matrices depicting correlation coefficients for the four Likert scales between difficulty and interestingness level. The legend shows the colors depicting the correlation coefficient values, where -1 shows strong negative correlation, +1 shows strong positive correlation, while values around 0 show weak or no relationship. The labels on x- and y- axis (letters A to H) denote observed lessons as given in TABLE I

Mind map

- helps explain concepts and ideas related to the central idea
- visual structure of ideas without too much text

https://images.app.goo.gl/isLVoawfwzexPcp87

Illustrative diagrams

- Image with characteristic parts that are shown in the image

Display of open questions next to closed ones

- selected answers to open questions are written next to the group statistics

Photos of the participants

- add their photo next to the participant's answer (for non-anonymous research)

Color coding

- reduced image in which the amount of a certain color is visible + a fragment of the real text

"Our convention occurs at a moment of crisis for our nation. The attacks on our police, and the terrorism in our cities, threaten our very way of life."

"How great are our police? And how great is Cleveland?"

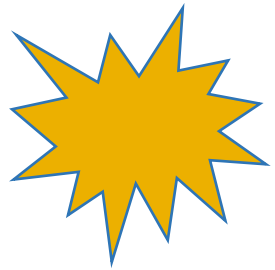
"The irresponsible rhetoric of our president, who has used the pulpit of the presidency to divide us by race and color, has made America a more dangerous environment than frankly I have ever seen and anybody in this room has ever watched or seen."

"There's a lot of work to do. Too many people haven't had a pay raise since the crash. There's too much inequality. Too little social mobility. Too much paralysis in Washington. Too many threats at home and abroad."

"We have the most dynamic and diverse people in the world. We have the most tolerant and generous young people we've ever had."

"So don't let anyone tell you that our country is weak. We're not."

References



https://www.scribbr.com/methodology/qualitative-research/

https://guides.library.jhu.edu/QDAS

https://www.geeksforgeeks.org/qualitative-data-analysis-methodologies-and-methods/#5-grounded-theory-analysis

https://www.enago.com/academy/data-visualization-techniques-qualitative-research/

https://depictdatastudio.com/how-to-visualize-qualitative-data/