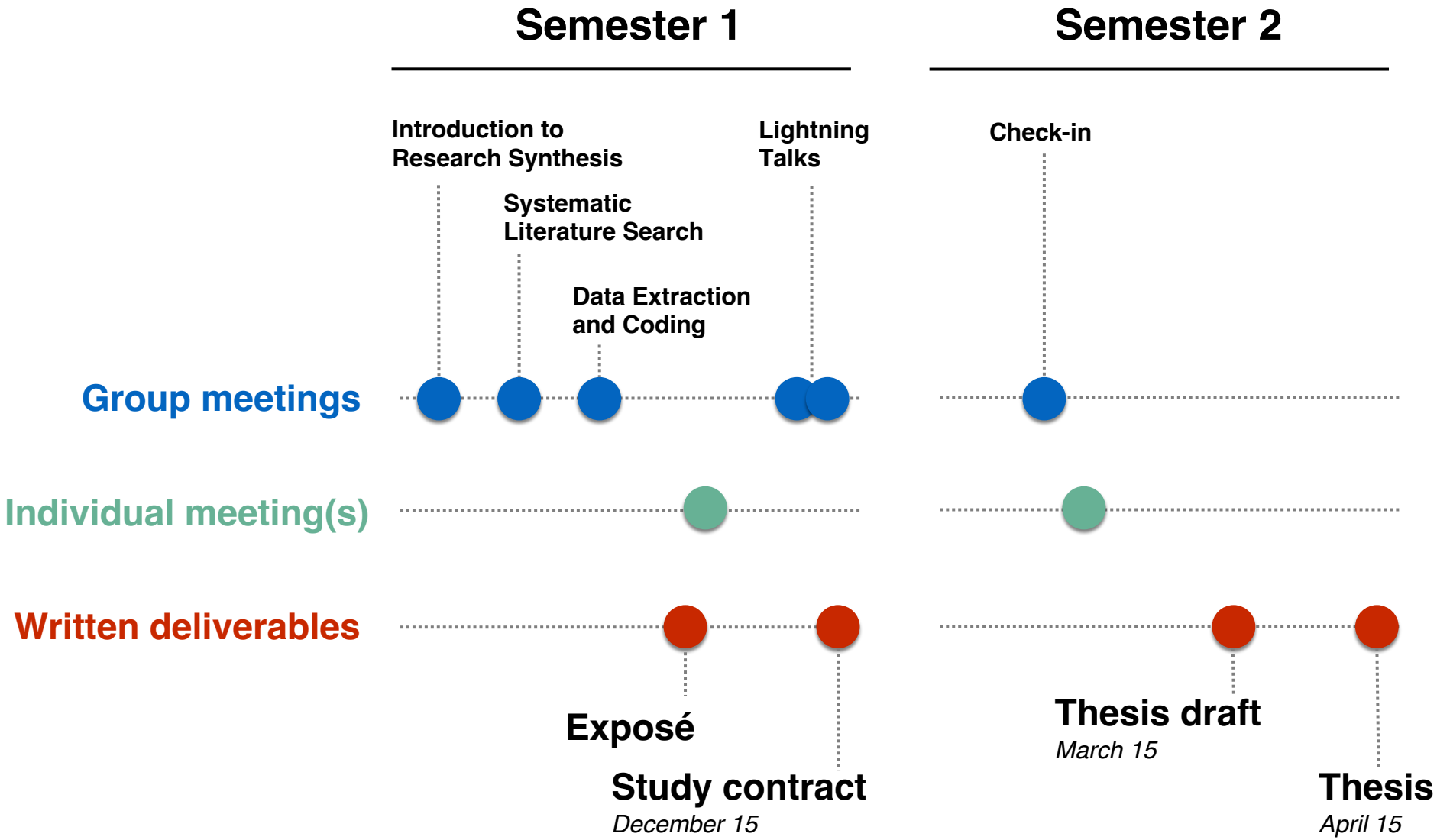
Data Extraction and Coding

Loreen Tisdall, Maximilian Theisen, Rui Mata

October 22nd 2024

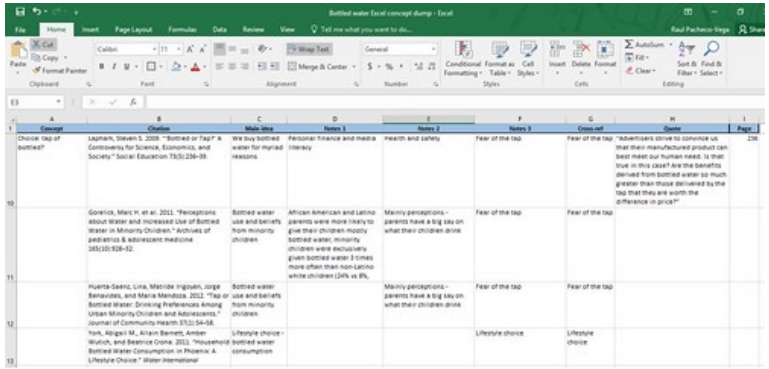
Timeline

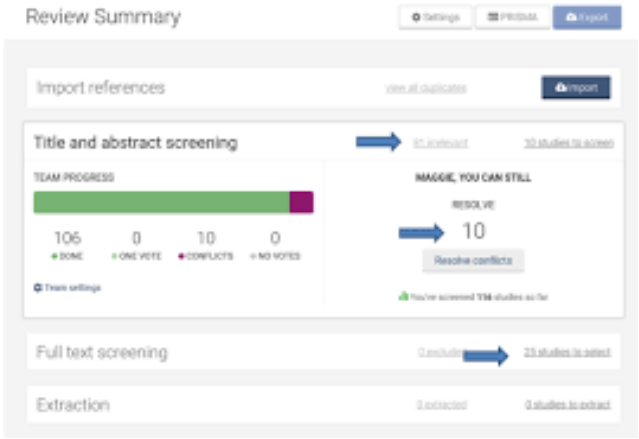

Semester 1
Semester 2
Introduction to Research Synthesis
Lightning Talks
Check-in
Systematic Literature Search
Data Extraction and Coding
Group meetings
Individual meeting(s)
Written deliverables
Exposé
Study contract
December 15
Thesis draft
March 15
Thesis
April 15

Agenda

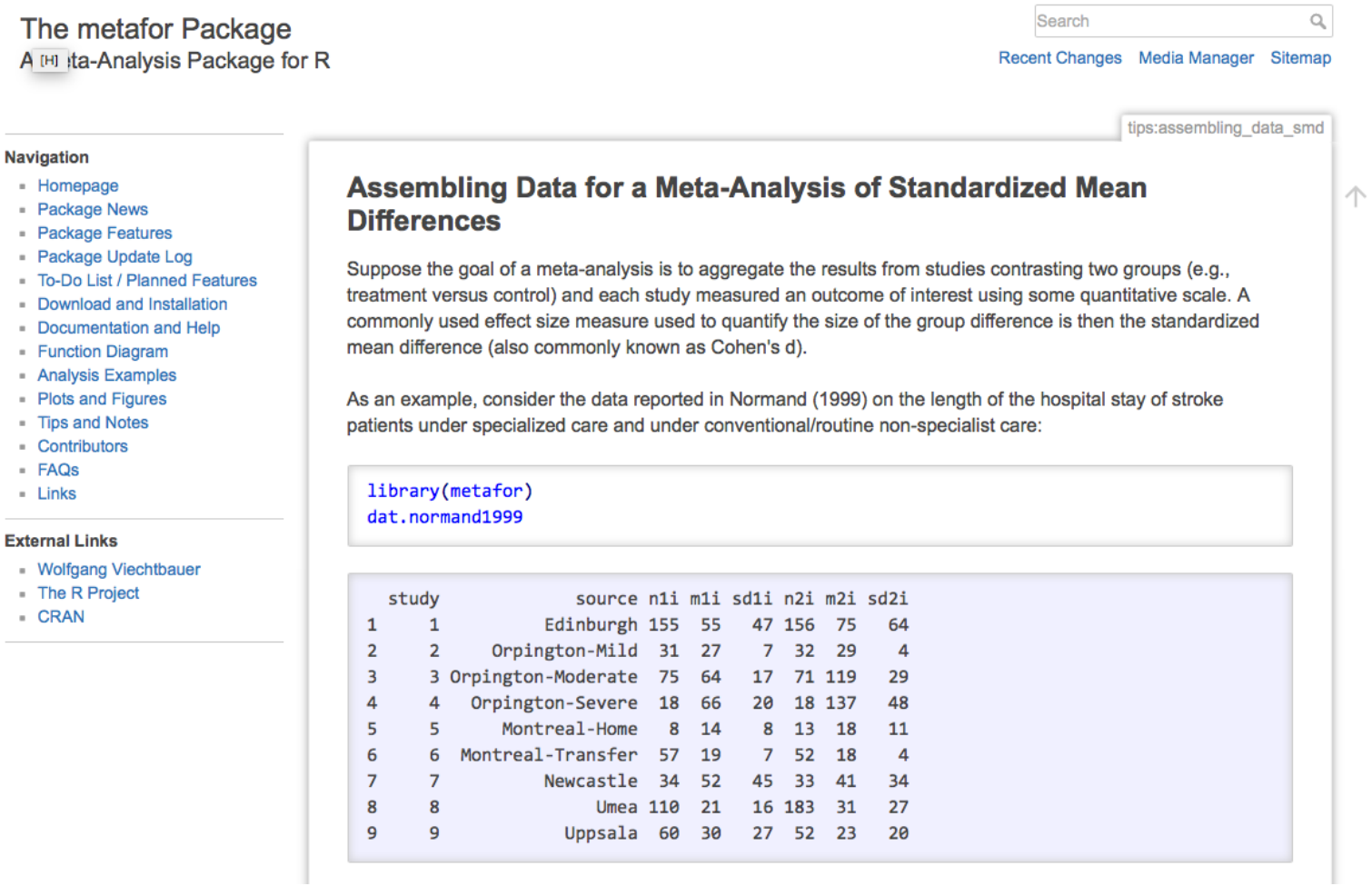
1. Review past session (10 min)
2. Data extraction and coding (30 min)
3. Next steps (5 min)
4. Q&A

Data Extraction and Coding





Meta-analysis: the *metafor* package for *R*



The metafor Package

A [H] ta-Analysis Package for R

Search

Recent Changes Media Manager Sitemap

tips:assembling_data_smd

Navigation

- Homepage
- Package News
- Package Features
- Package Update Log
- To-Do List / Planned Features
- Download and Installation
- Documentation and Help
- Function Diagram
- Analysis Examples
- Plots and Figures
- Tips and Notes
- Contributors
- FAQs
- Links

External Links

- Wolfgang Viechtbauer
- The R Project
- CRAN

Assembling Data for a Meta-Analysis of Standardized Mean Differences

Suppose the goal of a meta-analysis is to aggregate the results from studies contrasting two groups (e.g., treatment versus control) and each study measured an outcome of interest using some quantitative scale. A commonly used effect size measure used to quantify the size of the group difference is then the standardized mean difference (also commonly known as Cohen's d).

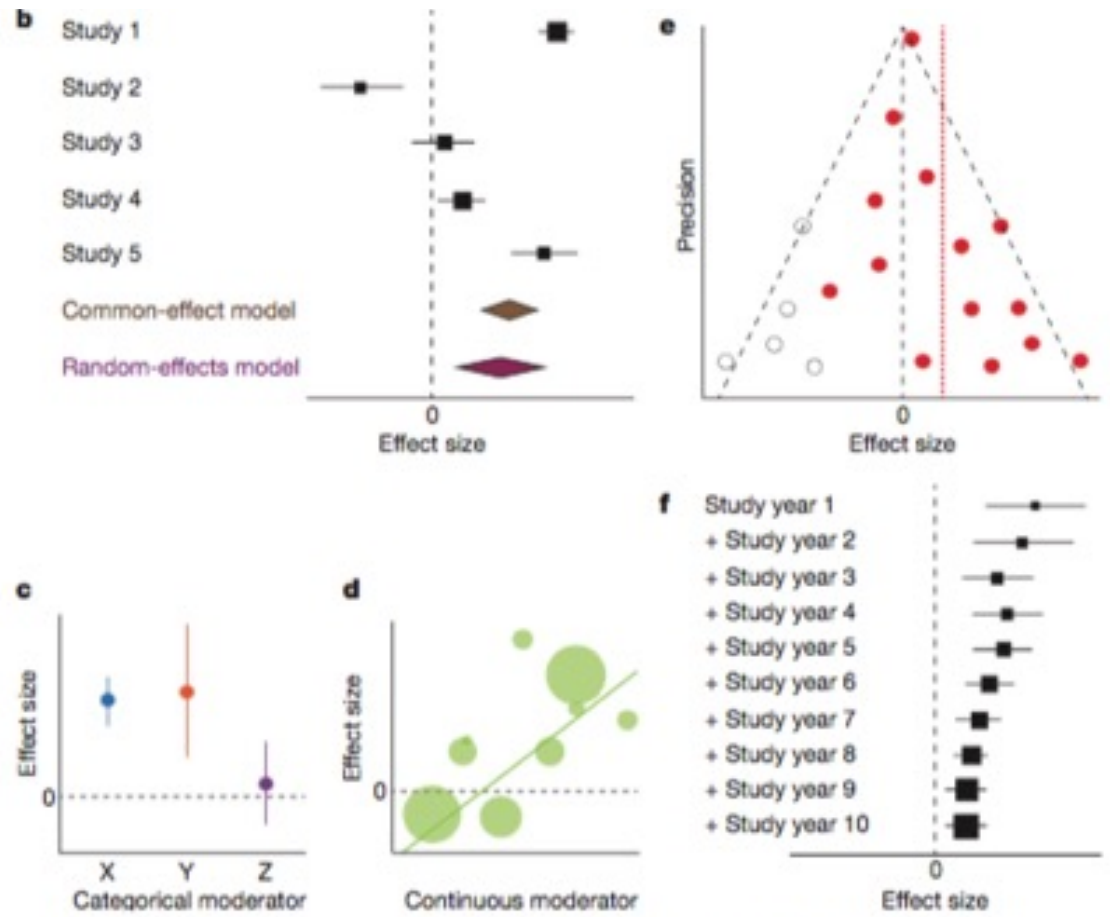
As an example, consider the data reported in Normand (1999) on the length of the hospital stay of stroke patients under specialized care and under conventional/routine non-specialist care:

```
library(metafor)
dat.normand1999
```

```
  study             source n1i m1i sd1i n2i m2i sd2i
1     1          Edinburgh 155  55   47 156  75   64
2     2     Orpington-Mild  31  27    7  32  29    4
3     3 Orpington-Moderate  75  64   17  71 119   29
4     4   Orpington-Severe  18  66   20  18 137   48
5     5      Montreal-Home   8  14    8  13  18   11
6     6  Montreal-Transfer  57  19    7  52  18    4
7     7          Newcastle  34  52   45  33  41   34
8     8               Umea 110  21   16 183  31   27
9     9            Uppsala  60  30   27  52  23   20
```

http://www.metafor-project.org/

Data Extraction and Coding



Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*, 175. http://doi.org/10.1038/nature25753

Data Extraction and Coding

	Flooding	Hurricane	Wildfire	Heatwave	Drought	Climate change	Multiple hazards	Total
Trust in measures	8	3	3	0	0	0	0	14
Trust in government	5	0	0	1	0	6	1	13
Experience	21	14	1	1	2	1	4	44
Place attachment	3	1	5	0	0	0	1	10
Knowledge	1	2	2	2	2	4	0	13
Risk perception	27	11	9	2	4	9	3	65
Climate change belief	1	0	0	0	1	3	0	5
Responsibility	4	3	3	0	0	4	0	14
Injunctive norms	1	0	4	0	1	1	0	7
Self-efficacy	4	1	3	0	1	2	0	11
Outcome efficacy	8	1	6	2	2	1	0	20
Negative affect	9	2	2	0	1	8	0	22
Descriptive norms	1	3	0	0	1	0	0	5

Fig. 2 | Types of climate-related hazards examined. Number of studies observed for each combination of climate-related hazard and motivational factor. Green cells indicate four or more observed studies. Yellow cells indicate one to three observed studies. Red cells indicate no observed studies.

van Valkengoed, A. M., & Steg, L. (2019). Meta-analyses of factors motivating climate change adaptation behaviour. *Nature Climate Change, 9*(2), 158–163. https://doi.org/10.1038/s41558-018-0371-y

Data Extraction and Coding

Dimension	Characteristic								
Setting (100)	Conventional (68)					Digital (32)			
Choice architecture tool (304)	Structuring the choice task (117)					Describe choice options (187)			
Category (290)	Default (60)	Simpli-fication (12)	Social ref. (49)	Change effort (41)	Disclosure (18)	Warnings/ graphics (55)	Precommit-ment (6)	Reminders (34)	Implement. Intent. (8)
Application context (96)	Health (38)	Environment (19)	Finances (12)			Energy (10)	Policy Making (10)		
Clusters of outcomes (317)	Energy consumption (18)	Healthy products chosen or purchase (60)	Sales of environmentally friendly products (43)			Amount donated (20)	Other (176)		
Data collection (291)	Online experiment (46)	Lab experiment (47)	Field experiment (150)			Experiment (other) (6)	Survey (experiment) (42)		
Significance (308)	Insignificant effect (118)					Significant effect (190)			
Magnitude (273)	Low (<10%) (78)		Medium (10%-30%) (81)			High (>30%) (112)			

Fig. 2. Morphological box of empirical nudging studies including counting.

Hummel, D., & Maedche, A. (2019). How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. *Journal of Behavioral and Experimental Economics, 80*, 47–58. http://doi.org/10.1016/j.socec.2019.03.005

Data Extraction and Coding

Table 3

Absolute Frequencies of Data Points and Percentage of Variables by Effect Size (Ordered by the Combined Frequency of Medium and Large Effects)

	Absolute frequency of data points			% of variables			
	Students[a]	Effect sizes	Variables	No effect	Small effect	Medium effect	Large effect
Overall	1,920,239	3,330	105	12	36	36	15
Instruction variables	208,711	1,595	42	5	26	45	24
Social interaction	26,860	123	5	0	0	40	60
Stimulating meaningful learning	49,272	229	9	0	22	56	22
Assessment	41,493	316	8	0	25	50	25
Presentation	46,157	354	9	0	33	33	33
Technology	29,022	401	6	17	33	50	0
Extracurricular training programs	15,907	172	5	20	40	40	0
Student variables	1,711,528	1,735	63	18	43	30	10
Intelligence and prior achievement	74,711	95	4	0	0	50	50
Strategies	133,757	343	18	11	28	50	11
Motivation	137,880	390	12	17	42	25	17
Personality	1,093,174	694	16	31	44	25	0
Context	272,006	213	13	15	77	8	0

Note. No effect = $|d| < .11$; small effect = $.11 \leq |d| < .35$; medium effect = $.35 \leq |d| < .66$; large effect = $|d| \geq .66$.

[a] Estimated by replacing missing values of a meta-analysis by the median value of all meta-analyses.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Visualization in reviews is infrequent …

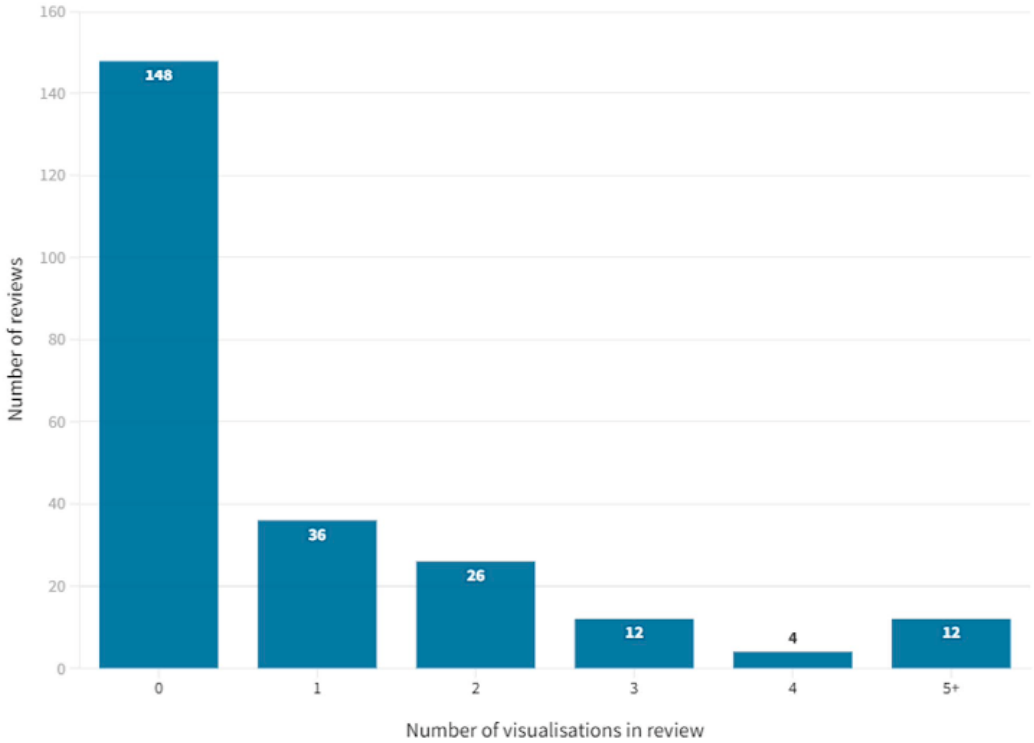
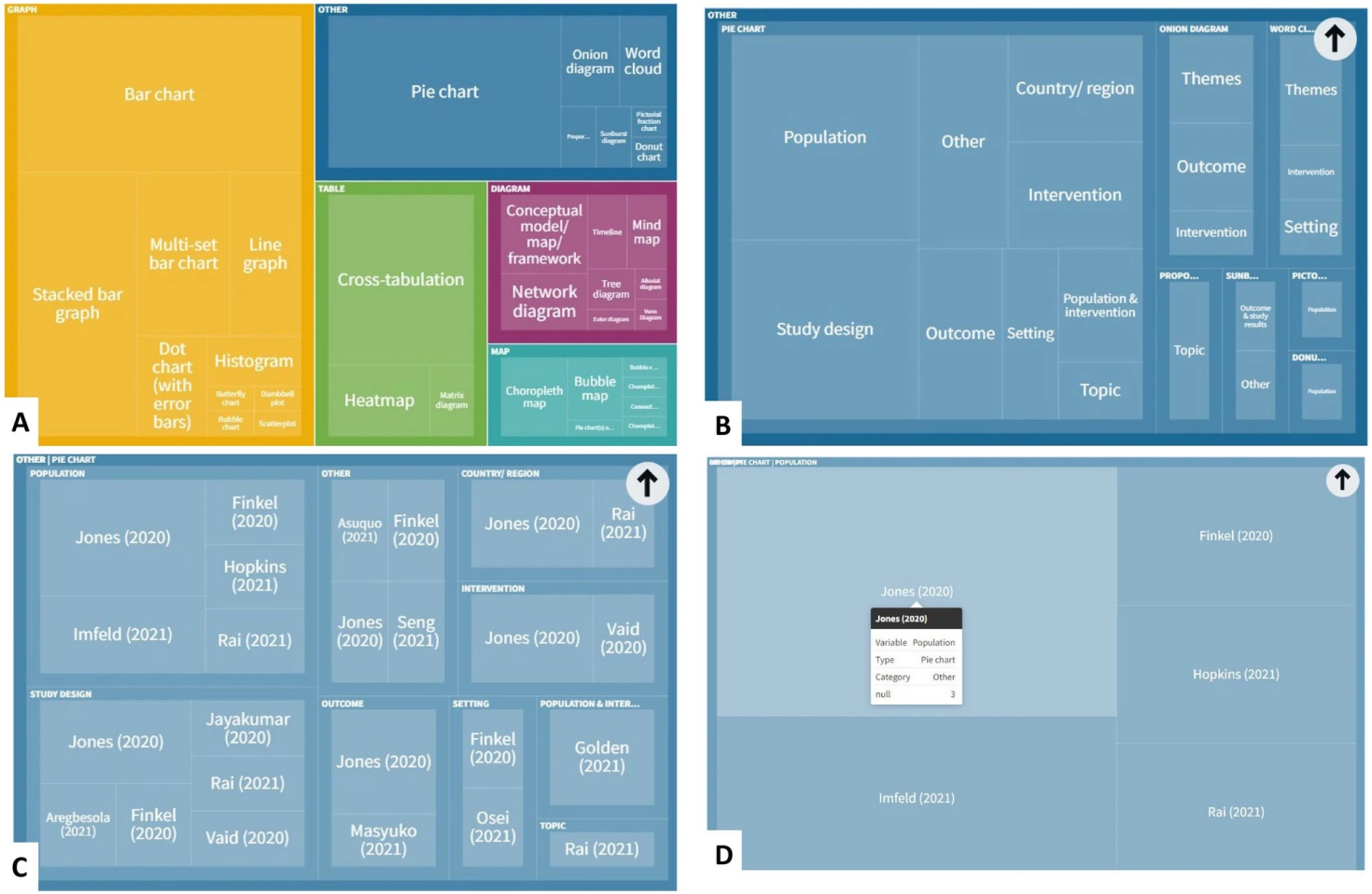


Fig. 2 Number of data visualisations per review

South, E., & Rodgers, M. (2023). Data visualisation in scoping reviews and evidence maps on health topics: A cross-sectional analysis. *Systematic Reviews, 12*(1), 142. https://doi.org/10.1186/s13643-023-02309-y

Visualization in reviews is infrequent … but key!



South, E., & Rodgers, M. (2023). Data visualisation in scoping reviews and evidence maps on health topics: A cross-sectional analysis. *Systematic Reviews, 12*(1), 142. https://doi.org/10.1186/s13643-023-02309-y

Visualization in reviews is infrequent … but key!

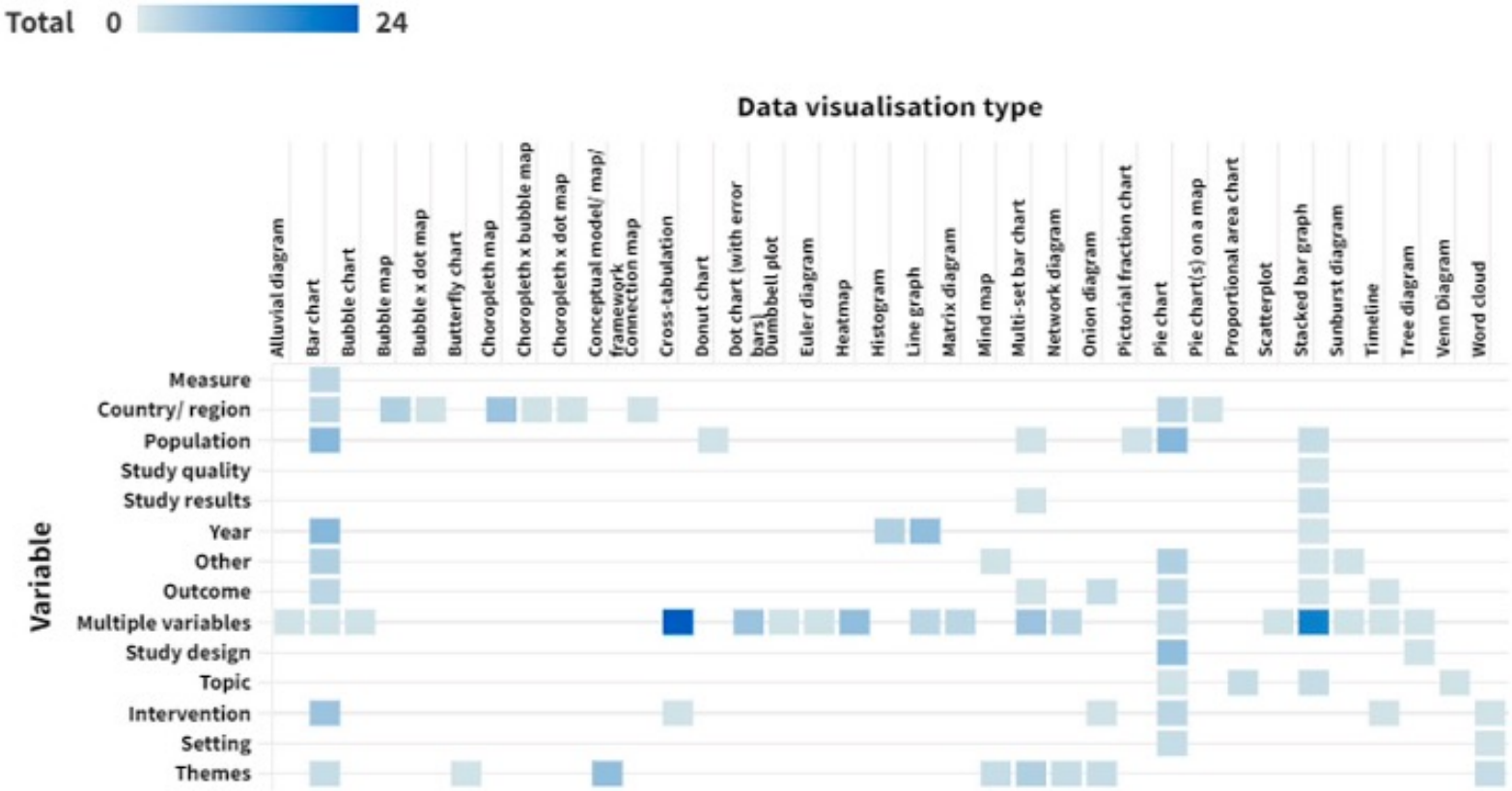


Fig. 5 Variables presented by each data visualisation type. Darker cells indicate a larger number of reviews. An interactive version of this heat map is available online: https://public.flourish.studio/visualisation/10632665/. Users can hover over each cell to see the number of data visualisations for that combination of data visualisation type and variable. The unit of this heat map is the individual data visualisation, so multiple data visualisations within a single scoping review are represented in this map. Created with flourish.studio (https://flourish.studio)

South, E., & Rodgers, M. (2023). Data visualisation in scoping reviews and evidence maps on health topics: A cross-sectional analysis. *Systematic Reviews, 12*(1), 142. https://doi.org/10.1186/s13643-023-02309-y

Product Goal: Exposé

- Title (descriptive and short, max 12 words, use PRISMA)
- Your name, email, student number
- Sections: 1) background knowledge citing key literature (as text) 2) gap, 3) proposed methods, 4) expected results (what is the key table or figure?), 5) references, 6) appendices (if any, e.g., detailed search strategy, flow diagram, current coding scheme)
- Formatting: ca. 2 pages, double spaced, Times New Roman, 12 pt, APA style referencing
- Language: English or German
- Word document: LastnameFirstname_Expose.docx

if you use AI…

Follow the guidelines for citing AI

Documentation table (example)

KI Tool	Use	Affected Sections
DeepL	• Translate quotations from German	Entire document
Chat GPT (version 3.5)	• Conceptualization • Spell checking	Entire document
Copilot	• Code suggestion	Methods and results

UNIBAS guidelines

Be aware of the risks of AI

Confabulation	LLM-based applications (like ChatGPT) are prone to producing incorrect, but plausible facts, a phenomenon known as confabulation or hallucination.
Bias	LLM-based applications are trained on a vast amount of text, and then receive additional training from humans to create guardrails on their output. These processes may introduce biases in the text, which can range from gender and racial biases to biases against particular viewpoints, approaches, or political affiliations.
Privacy	When data is entered into an LLM-based application, it can be used for future training by the organizations developing it. While ChatGPT offers a privacy mode that claims to not use input for future training, the current state of privacy remains unclear for some models and the legal implications are uncertain.
Instructional	LLM-based applications may commit errors and there remains a substantial risk that students will use these tools as a crutch, undermining learning.

Mollick, E., & Mollick, L. (2023). Assigning AI: Seven Approaches for Students, with Prompts. arXiv preprint arXiv:2306.10052.

You are accountable (not AI)

AI Authorship

Large Language Models (LLMs), such as ChatGPT, do not currently satisfy our authorship criteria. Notably an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs. Use of an LLM should be properly documented in the Methods section (and if a Methods section is not available, in a suitable alternative part) of the manuscript.

https://www.nature.com/nature-portfolio/editorial-policies/ai

Assignment of students to supervisors

Image created with AI (ChatGPT 4o), October 2024

To do list for individual meetings

- ☐ Conduct a literature search based on your research question using one or more databases
- ☐ Rewrite your exposé including details about extraction and coding and email it to your instructor at least 2 days before the individual meeting
- ☐ Prepare any supporting materials you think could be helpful for the individual meeting (e.g., timetable for writing the thesis) and compile a list of open questions