International Society of the Learning Sciences

Navigating Multidimensional Data Structures: Insights from Data Experts and Implications for Pedagogy

Natalya St. Clair, Lynn Stephens, Daniel Damelin
nstclair@concord.org, lstephens@concord,org, ddamelin@concord.org
The Concord Consortium

Abstract: Understanding and reasoning with multidimensional data is a critical skill for students in various disciplines. This study explores how data experts navigate and analyze unfamiliar multidimensional datasets. Through our interviews with nine data experts, we identified three main approaches: (1) manipulating flat tables, (2) creating relational databases, and (3) using computational commands. These findings challenge our initial assumption that making hierarchy would be a common expert data move. Rather than revealing a "typical" strategy, these interviews yielded a range of approaches, with most experts describing more than one approach and how they would decide between them. These insights will inform the design of pedagogical techniques and tools to support students' reasoning with multidimensional data.

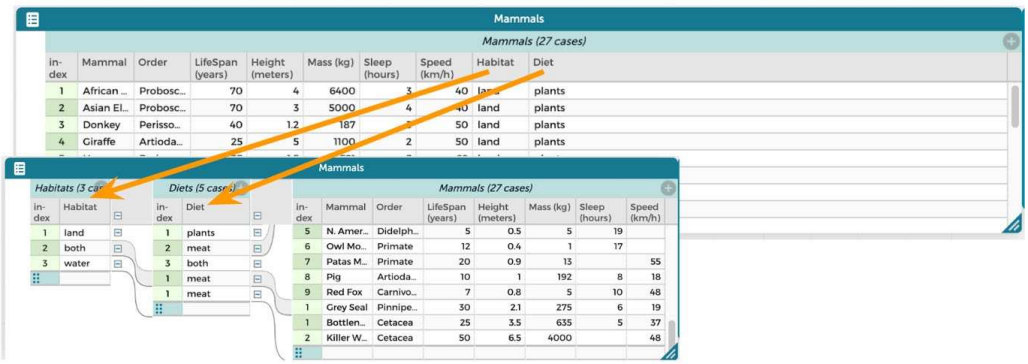
Introduction

Supporting Reasoning with Multidimensional Datasets, a 3-year NSF project, aims to identify design principles to guide technology developers, curriculum developers, and researchers in creating environments that are conducive to promoting data fluency for all learners. The goal is to investigate how learners can best be supported to represent, interact with, and make sense of multidimensional data as they seek to understand and reason with those data. Citizens need to know how to interpret and use data to make informed decisions (Finzer, 2013; Kazak et al., 2021; Wise, 2020). Even simple datasets can involve *multiple dimensions*, meaning they can be organized or grouped by more than two attributes. Erickson (2022) argues that arranging multidimensional sets in hierarchical structures allows students to compare groups, and that visualizing these data organizations can have pedagogical benefits. The Common Online Data Analysis Platform (CODAP) (Finzer, 2014), a widely used open-source pedagogical software, allows users to drag and drop data to create a hierarchically structured view of multidimensional data (Figure 1). Our design-based research project seeks to develop new features that can support CODAP's existing representations of multidimensional data.

We began with an exploratory study of data experts. Understanding how experts in data science approach unfamiliar multidimensional data can help better position us to recognize useful aspects of student intuitive practices when we see them, even if those novice practices are unexpected. These experts included data scientists and professional analysts from business and academic settings. Such experts often have substantial domain knowledge and deep levels of data expertise coupled with computer programming skills (Finzer, 2013). Thus, our research questions for this exploratory study are: 1) What approaches did data experts use to (re)structure multidimensional data? 2) What tools, representations, and prior experiences contributed to those approaches? We conclude by considering questions for pedagogical design raised by these results.

Figure 1

Screenshots Showing Mammal Data and Their Categories in CODAP.



Note: In the bottom screenshot, categorical attributes are configured into a hierarchically structured view.

Related work

We begin by discussing two areas of research that motivate our work with multidimensional datasets: students' understanding of hierarchical data structures and studies of data work that are most similar to ours in terms of goals and methods. We then discuss the characteristics of multidimensional datasets.

Pedagogical tools for scaffolding hierarchy

The term "data move," as defined by Erickson et al. (2019), refers to an action that changes the contents, structure, or values within a dataset. Making hierarchy is one such data move, and it holds significance for the authors because it allows analysts to work with the structure of a dataset intentionally, rather than working with data that has a pre-imposed structure (pg. 12). Konold et al. (2017) have shown that young people have an intuitive understanding of nested tables, which suggests that hierarchical data structures may be more natural and easier to interpret than flat representations. In a single case study, Haldar et al. (2018) observed that when students were given appropriate scaffolding, they were able to use the tools in CODAP with multiple linked visual representations to graph and organize data hierarchically. This suggests that with the right educational frameworks, learners can effectively translate their intuitive understanding of data into more complete structures often used in data analysis. Erickson (2022) conjectures that students who use hierarchy in software like CODAP will perform better when they are brought to a more abstract programming data analysis environment. However, our classroom observations of students using CODAP suggest there is still much to learn about how to support students in working with hierarchical data structures.

Researchers have investigated the development of tools for working with different types of data, which may provide insights into how to scaffold students who are struggling with these new concepts. Chang and Myers (2016) showed in a lab study of adult spreadsheet users and programmers that visualizing hierarchical data using tables in a way that allowed them to organize nested tabular structures helped users comprehend the data more quickly, especially for non-experts. Bartram et al.'s (2022) qualitative study investigated the effectiveness of tables in managing data and found that spreadsheets are useful to data workers (people with diverse expertise working with data) for data reading and ad hoc analysis. These studies suggest that nested and hierarchical tabular structures may help users understand multidimensional data structures.

Formats for representing multidimensional datasets

Our exploratory study focuses on multidimensional data structures and the methods data experts employ to restructure them. This paper uses the following vocabulary.

Flat table: The most basic type of data table is typically structured like a spreadsheet, with rows and columns forming a two-dimensional grid, and the cells contain corresponding values (Bartram et al., 2022; Broman & Woo, 2018).

Layout of datasets: Datasets are often structured in a "long" format, where each row represents a unique item, or *case* (Konold et al., 2017), or a "wide" format where some values are represented within column labels (Figure 2). For analysts, structuring the data in a long "tidy data" (Wickham, 2014) format is the preferred storage method before undertaking analysis according to Broman & Woo (2018), as this facilitates easier data manipulation. The format may become "wider" when the analysis is performed, for instance, to facilitate the identification of trends across data points of a repeated measure (as across a series of tests).

Granularity: In a single table, the *granularity* (Bartram et al., 2022) refers to detail within a dataset; for instance, a "fine-grained" view might look at data by month, whereas a "coarse-grained" view might aggregate data by year (page 691).

Relational databases: These structures organize information into multiple linked tables using common columns to connect them. Several of our experts referred to the linked table(s) with the finest granularity, typically containing data about measurement events, as "*fact table(s)*." These serve as foundations of the databases, holding

Figure 2

"Wide" Data Fformat (a) vs. "Long" or "Tall" Ttable format (b).

Wide Data Format (a)

Student	Quiz 1	Quiz 2	Test
A	98	95	95
B	88	82	90

Long Data Format (b)

Student	Assignment	Score
A	Quiz 1	98
A	Quiz 2	95
A	Test	95
B	Quiz 1	88
B	Quiz 2	82
B	Test	90

the core data points that are analyzed and interpreted. The fact table is often linked to other tables, known as *dimension tables*, which provide additional context and attributes to the measurement events.

Hierarchical databases: These tree-like data structures arrange data in levels with parent-child relationships. Each level can be *multidimensional*, including multiple independent variables. In this kind of structure, a dataset that might otherwise lend itself to a relational structure can instead be displayed within a single table. CODAP is designed to facilitate the creation and manipulation of hierarchical tables (Figure 1).

Theoretical perspective

Our perspective draws from the foundational principles of science education research, in which the idea of identifying and building on students' partial knowledge has deep roots. Hammer (1995) suggested we look at what students already know and can do, identifying seeds of scientific practice to cultivate, but cautions that the beginnings of mature practice in one student may be very different from beginnings in another, just as mature practice varies greatly among professionals. This perspective is complemented by the work of Núñez-Oviedo and Clement (2019), who described a responsive teaching method where a teacher decided on the fly which student ideas could be built on and modified to steer students toward a target conceptual model. The teacher did not know what ideas to expect from the students, but because she had a clearly delineated target, she was able to recognize an idea that superficially bore little resemblance to the target model. This approach relies heavily on teachers' vision of an end goal, allowing them to shape and utilize student contributions that may not initially align with conventional expectations.

The complexity of aligning a learner's conceptual framework with the structural demands of a multidimensional dataset poses significant challenges, even when using advanced digital tools like CODAP that enhance data visualization. The core of this challenge is the learner's mental representation of the data, which underlies any approaches to the organization and reorganization of a dataset. While Erickson et al. (2019) provide insights into six core data moves, our project aims to delve deeper into experts' mental representations of data. Understanding these should help us recognize student ideas that can be usefully built on.

Methods

The goal of this exploratory study is to understand how data experts organize and make sense of multidimensional data, particularly how they try to structure and make sense of an unfamiliar dataset. We used semi-structured think-aloud interviews (Yin, 2009) to explore how they thought about and interacted with such data. Taking an approach reminiscent to that of Pfannkuch et al. (2016) in their study of practitioners of probability modeling, we interviewed experts engaged in different sectors of the field of data analysis to investigate their thinking about data structure.

Participants

Participants were recruited through targeted emails and suggestions from expert consultants. A diversity of participants was sought in terms of sex, race, level of experience, and sector. Prior to the main study, two pilot interviews were conducted with data experts to inform the development of the interview protocols. Ten experts were interviewed. Nine transcripts proved useful to address our research questions. Table 1 summarizes these nine experts, all based in Canada and the United States. Participants represented a variety of sectors and organization sizes, from small business founders/CEOs to large nonprofit and for-profit organizations.

Table 1
Interview Study Participants

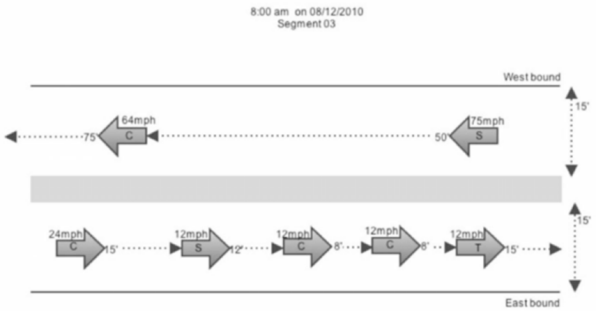
ID	Role	Sector	Org. Size	Analytic Tools
P1	Researcher	nonprofit	1001-5000	Excel, CODAP
P2	Director	global real estate	500-1000	Stata, Python, C++
P3	Data Manager	software services	1001-5000	SQL
P4	CEO	software services	2-10	PANDAS, Python
P5	Data Analyst	software services	1001-5000	Tableau, Power BI
P6	Professor of Data Science	education	1001-5000	R, Stata, proprietary
P7	Data Science Manager	software services	10000+	R, Power BI, Python
P8	Expert Data Scientist	energy; utility services	10000+	Python, PySpark
P9	Applied Science Manager	software services	10000+	Java, C++

Procedure

Participants engaged remotely via Zoom for a duration of 60 to 90 minutes. Participants were invited to bring a dataset of their own if they wished, to demonstrate how they typically work with data. After explaining the motivation behind the project and potential outcomes, we conducted a semi-structured interview involving three datasets we provided: one in the form of diagrammatic case cards (Figure 3), another with data in a flat "wide" table format (Figure 4), and a third with data in a flat "long" table format but accompanied by questions that could not be easily answered without data restructuring (Figure 5). The third dataset (Figure 5) was presented in the CODAP environment, and we asked participants to walk us through achieving a task goal. We assisted them in using the software as they considered how they would approach the problem.

Figure 3

One of Two Snapshots of Traffic along Two Different Road Segments



** Image credit:(Konold et al., 2017), reprinted with permission. Participants were asked to imagine that they would be receiving many of these and asked "Given data like this, how would you think about organizing it?"*

Figure 4

*Recreation of "Wide" Table Format Used for a Public Data Sample**

Date/Time	Time of Day	Location A 2017	Location A 2016	Location A 2015	Location B 2017	Location B 2016	Location B 2015
m/d - Hour 1	night	*data*	*data*	*data*	*data*	*data*	*data*
m/d - Hour 2	night	*data*	*data*	*data*	*data*	*data*	*data*
m/d - Hour 3	morning	*data*	*data*	*data*	*data*	*data*	*data*
m/d - Hour 4	morning	*data*	*data*	*data*	*data*	*data*	*data*

** Data table adapted from Kazak et al. (2021). The data are not reproduced here but were air quality measurements given to three decimal places. Participants were asked what they thought of this data organization.*

Figure 5

*"Long" Table Format Used for a "BARTy" Instructional Module Data Sample.**

BART data

BART data (1774 cases)

index	hour	date	riders	startAt	endAt	startReg	endReg
1	8	Wed Ap...	3	12Str	12Str	Oakland	Oakland
2	8	Wed Ap...	26	12Str	16Str	Oakland	City
3	8	Wed Ap...	2	12Str	19Str	Oakland	Oakland
4	8	Wed Ap...	12	12Str	24Str	Oakland	City
5	8	Wed Ap...	5	12Str	Ashby	Oakland	East Bay
6	8	Wed Ap...	15	12Str	BalbPk	Oakland	City
7	8	Wed Ap...	6	12Str	BayFar	Oakland	East Bay

** Full data available at http://short.concord.org/lwe.*
Participants were asked where they might suggest express routes.

If participants were silent, we asked them what they were thinking or what questions they were asking themselves. If they began describing what they might do with a dataset, we asked them to draw and/or describe their mental imagery. If the participants had brought a dataset, we invited them to describe it and how they would normally work with it. We also asked them to walk us through any software visualizations they would normally use for making sense of multidimensional data. Throughout the interviews, we probed participants for explanations and to confirm our understanding. We concluded by asking about the tools they use in their work and how they use them to make sense of multidimensional data.

Data collection and analysis

The Zoom interviews were audio/video recorded. Audio recordings were transcribed. Notes were also taken during each session. Three researchers were present in all interviews, with one acting as the note-taker. The transcribed interviews were subjected to an exploratory analysis (Yin, 2009). Using a constant comparative method (Merriam & Tisdell, 2016), two researchers annotated the transcripts and identified categories of data structures and themes related to data structure. These were discussed with a third researcher and refined. Emerging themes were agreed upon through discussion.

Results

1. The approaches the experts used to navigate and analyze unfamiliar multidimensional datasets can be grouped into three categories: manipulating a flat table, creating relational databases, and using computational commands for ad hoc analysis.
2. Data experts' professional backgrounds and the software they use appeared to play a key role in how they worked with and visualized the multidimensional data structures. For example, experts who said they would use relational databases tended to be those who use programming languages to structure data in their professional work.

The descriptions from this analysis are not meant to prescribe how *students* should structure datasets but to shed light on the diverse and nuanced ways in which experts work with data. Rather than revolving around the six core data moves, the experts appeared to use a broader spectrum of methods and considerations.

Manipulating a flat table

Participants' data structures often involved a level of data organization within a flat table. In this approach, the flat tables served as a base structure for organizing data. In the case card dataset (Figure 3), the imagery participants described when constructing their datasets included one or more of the following:

- A single flat table where elements could be extracted into subsets to create hierarchy (P1, P3, P6),
- Multiple tables that could be organized into relational databases (P2, P9),
- A large table that could be organized into relational databases if too big (P4, P5, P7, P8).

All of the participants initiated their data structuring process by thinking about how to organize their data in a tabular format. For example, referring to the case card dataset, P7 explained:

> I'd start to structure some type of fact table. You know, just trying to think first. [I'd ask] what would I use as my rows? What would I use as my primary elements? Perhaps the vehicles themselves would be the rows and their direction and speed would be the initial columns. Maybe do some dimension tables if I thought this data was going to get bigger and more unwieldy to do as a single table, but obviously initially, just start to think about putting it in some type of tabular format to make better sense of it.

Others said they would structure the data as flat tables with multiple levels of granularity. P1 said:

> The mental imagery was very much just a flat data table. Um, where every element that I could pull out of this image was captured. [...] and it wouldn't be until I sort of went through that analytical process that I might start actually privileging certain variables and saying, 'This seems to be a really important thing.'

After describing a flat table, both P1 and P7 (among others) immediately began talking about granularity.

Long and wide table formats

The layout of flat tables in a "long" or "wide" format was a common theme in structuring the data. P5 described the "wide" dataset whose structure is recreated in Figure 4:

> I could certainly imagine the first thing I might want to do to this would be to get it into a taller format. [...] I would definitely then start to think about trying to get this into a format that would be easier for me to do where we would have six times the number of rows because there would be a [location A] versus [location B] and year for each one of these observations.

P3 and P8 brought up the challenge of preparing longer and wider tables for data administration or data analysis, saying that data administrators preferred longer formats whereas analysts and software such as Tableau sometimes preferred wider formats. Deciding when best to use "long" or "wide" formats and how to structure each was a frequent theme (P4, P5, P6, P8).

Creating relational databases

The expert participants tended to use relational databases to represent hierarchical aspects of data structure. This helped reduce data redundancy and improved the data organization (P2, P7, P9). P9 said:

> The hierarchical nature of the data is something I definitely agree with, or at least that's how I interpret it. And I think in my mind that I tend to separate the logical structuring of the data from how it's represented as a format, in the sense that, for relational, it'd probably be the most efficient way to represent it, because there's a lot of duplication here with the snapshot.

For the purpose of making hierarchy, separate datasets that represent different levels of organization could then be nested within the larger data structure (P2). P9 described the relationship between relational databases and hierarchy this way:

> So the hierarchical piece, in my mind, would be thinking about *what* the data is, and then the *how* part. I don't know if this is too much in the weeds, but in the industry, I suppose there are some holy debates that go on about how the best way to structure data should be. [...] In my experience, the document-level structuring is really good if we want to do fast aggregations, so I think you're trading off higher speed, but at the cost of more storage. Whereas the relational databases are better in my mind for complex queries that say, 'Hey, we want to look at the relationship between snapshots and time of day, and vehicles, and other sorts of things.' I think both analyses would be accomplishable there. But, I usually prefer simplicity.

Using computational commands for ad hoc analysis

Some experts said they would opt to use ad hoc commands, which they create using functions and operations. For instance, P8 described the process of transforming data as beginning with a flat structure and then using left join operations (a type of command that combines rows from two or more tables) to filter data based on specific constraints. P2 noted that while the visualizations in CODAP were "lovely," thinking about it in terms of code is how he would look at it if the dataset became larger, as with two million or more rows. Merging data and using joins is a data move (Erickson et al., 2019), and the idea of using joins to help with restructuring the data for making hierarchy allowed experts to set up the data for doing the analysis (P2, P8).

While the experts reported a wide use of spreadsheet tools for structuring datasets, none used *only* spreadsheets for analysis. Many of them attempted to describe multidimensional data structure in terms not only of tables and rows but also by referring to the tools they would use for analysis. For instance, P2 and P4 referred to using Python and SQL queries to structure datasets. Others said they would start working in spreadsheet software like Excel and then move to an analysis tool like Python or Tableau (eight participants).

Using a flat table but talking about hierarchy

Another common task we observed experts engage in when analyzing datasets in a flat table was to reorder columns as they described a hierarchy. When columns are structured appropriately, analysis tools such as Tableau or PANDAS can automatically aggregate them to create a hierarchy or explode them to create new columns (P4, P5, P10). However, participants did not always create a hierarchical structure when they talked about hierarchy. For instance, in the third dataset (Figure 5), participants frequently used CODAP for rearranging columns and described nesting, but left the actual table flat (six participants). P1 spoke of "privileging columns" by dragging them left.

Influence of profession

Consistent with Pfannkuch et al. (2016), the backgrounds of the data experts we interviewed appeared to influence their approaches to structuring the datasets. P3, a data manager, tended to reason the most about the layout of the datasets, ensuring that they were organized in a way that was efficient and easy to access. The data analysts we interviewed, on the other hand, tended to start with long formats and shift to wider formats as needed to facilitate analysis (P2, P5, P6, P9). Two participants (P1, P6) had a background in education and spoke explicitly about how they would present these datasets to students. All participants emphasized the importance of thinking about the structure of the datasets for the stakeholders, noting that their responses might change depending on the person they were presenting to.

Discussion

The data experts we interviewed responded to multidimensional data in rich ways, and the structures they used when working with those data were varied. Broadly speaking, the participants described structuring the datasets using relational databases or flat data tables. There was also a subset of participants who started with flat datasets but said they would move to relational if there was a large number of cases. For these experts, their tools and their professional backgrounds related closely to the strategies they employed.

Implications for pedagogy

Our investigation into data experts' approaches to making sense of multidimensional datasets raises several questions concerning pedagogy. The experts we talked to believed in the flexibility of data organization. They didn't just see datasets as having a fixed structure; instead, they approached them as if they could structure and restructure the data as needed. This ability to modify data structure, a feature they found in all their tools, helped them analyze and draw conclusions from the data. However, most of the tools commonly available to students don't support this kind of flexibility in organizing data. From these interviews, we are convinced that students not only need access to tools that allow for such manipulation but also should be taught to see data as something they can organize and reorganize.

This, however, raises another question. The experts frequently used *relational databases* to structure multidimensional data. However, for pedagogical purposes, we must consider whether to develop data analysis tools that support relational databases or whether flat tables with hierarchical options are sufficient for educational settings. As some experts noted (P8, P9), there is a fine line in deciding whether to structure multidimensional datasets as relational or flat, depending on such factors as the size of the dataset and what type of software will be used for analysis. Because relational databases were such a strong focus for the experts when describing how to deal with large datasets, we continue to wonder about the effects on novices of the size and characteristics of datasets and suggest that datasets for instructional purposes need to be chosen carefully.

If, as we suspect, working with relational databases is unlikely a realistic option for novice high school learners, *nested data structures* may be more helpful for students in understanding multidimensional data. This is consistent with Chang and Myers (2016), who argue that visualizing hierarchical data in a way that allows them to have nested tabular structures can help adult users comprehend the data more quickly.

These results also raise a question about whether high school students should be taught about "long" versus "wide" table formats and if, as P8 suggested, they should be taught how to transform one into the other. The experts we interviewed tended to think of data as tabular, but it is important to consider whether students will as well. Some students may find it easier to understand data or how to structure it for analysis when it is presented in other forms such as case/data cards. Our project plans to investigate student interactions with a variety of data representations.

Limitations and future work

Our research was exploratory, focusing on understanding expert data analysis practices. Our study was limited to nine participants; further studies could expand to a larger size of participants. Another limitation lies in the geographic scope of our participants, who were primarily from North America. It's worth recognizing that data analysis practices can vary significantly across different regions and cultures. Future studies could include participants from a greater diversity of locations to explore how individuals around the world interact with datasets. This expansion could reveal insights into the global diversity of data analysis practices and potentially lead to the development of more region-specific pedagogical approaches.

Conclusion

We examined approaches, tools, and affordances data experts use to restructure multidimensional data. We have identified findings related to their data organization strategies and several factors that appeared to influence those choices. Results are informing our ongoing development of CODAP plugins for classroom use that are designed to help students develop the ability to manipulate and make sense of multidimensional datasets. For instance, given the emphasis several of the experts placed on "wide" vs. "long" flat tables, we plan to examine whether and how to approach this idea with students. Understanding more about how experts approach unfamiliar datasets outside their normal professional experience is helping us learn to better support students in the data manipulation and visualization they will need to tackle the complexities of unfamiliar data they will encounter in school and in life.

References

Bartram, L., Correll, M., & Tory, M. (2022). Untidy Data: The Unreasonable Effectiveness of Tables. *IEEE Transactions on Visualization and Computer Graphics*, *28*(1), 686–696.

Broman, K. W., & Woo, K. H. (2018). Data Organization in Spreadsheets. *The American Statistician*, *72*(1), 2–10. https://doi.org/10.1080/00031305.2017.1375989

Chang, K. S.-P., & Myers, B. A. (2016). Using and Exploring Hierarchical Data in Spreadsheets. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2497–2507. https://doi.org/10.1145/2858036.2858430

Erickson, T. (2022, December 1). Grouping and the Power of Hierarchical Data. *Bridging the Gap: Empowering and Educating Today's Learners in Statistics. Proceedings of the Eleventh International Conference on Teaching Statistics*. Bridging the Gap: Empowering and Educating Today's Learners in Statistics. https://doi.org/10.52041/iase.icots11.T2H1

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data Moves. *Technology Innovations in Statistics Education*, *12*(1). https://doi.org/10.5070/T5121038001

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, *7*(2).

Finzer, W. (2014). *Common Online Data Analysis Platform (CODAP)*. Common Online Data Analysis Platform (CODAP). https://codap.concord.org/

Haldar, L. C., Wong, N., Heller, J. I., & Konold, C. (2018). Students Making Sense of Multi-level Data. *Technology Innovations in Statistics Education*, *11*(1).

Hammer, D. (1995). Student inquiry in a physics class discussion. *Cognition and Instruction, 13*(3), 401-430.

Kazak, S., Fujita, T., & Turmo, M. P. (2021). Students' informal statistical inferences through data modeling with a large multivariate dataset. *Mathematical Thinking and Learning*, 1–21. https://doi.org/10.1080/10986065.2021.1922857

Konold, C., Finzer, W., & Kreetong, K. (2017). Modeling as a core component of structuring data. *Statistics Education Research Journal*, *16*(2), 191–212.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). John Wiley & Sons.

Núñez-Oviedo, M. C., & Clement, J.J. (2019). Large scale scientific modeling practices that can organize science instruction at the unit and lesson levels. *Frontiers in Education, 4*, Article 68.

Pfannkuch, M., Budgett, S., Fewster, R., Fitch, M., Pattenwise, S., Wild, C., & Ziedins, I. (2016). Probability Modeling and Thinking: What Can We Learn from Practice? *Statistics Education Research Journal*, *15*(2), 11–37. https://doi.org/10.52041/serj.v15i2.238

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10).

Wise, A. F. (2020). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, *29*(1), 165–181. https://doi.org/10.1080/10508406.2019.1705678

Yin, R. (2009). *Case Study Research: Design and Methods* (Vol. 5). SAGE Publications.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DRL-2201177. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

We thank William Finzer for serving as ongoing catalyst and advisor, constantly challenging us to think about data structure.