Olympiads in Informatics, 2023, Vol. 17, 189–205

DOI: 10.15388/ioi.2023.14

REVIEWS, COMMENTS

Guide to Teaching Data Science: An Interdisciplinary Approach

Orit HAZZAN, Koby MIKE
Faculty of Education in Science and Technology
Technion – Israel Institute of Technology
e-mail: oritha@technion.ac.il, mike@technion.ac.il, kobymike@gmail.com

Introduction

Data science is a new discipline of research that is gaining growing interest in both industry and academia. Data science is converging knowledge, skills and values from computer science, statistics, and various applications domains such as social science, digital humanities, life science and more (see Fig. 1).

As a result of the growing interest in data science, demand is increasing for data science programs for a variety of learners from a variety of disciplines (data science, computer science, statistics, engineering, life science, social science and humanities) and a variety of levels (from school children to academia and industry). While significant efforts are being invested in the development of data science curricula, or in other words, in *what to teach*, only sporadic discussions focus today on the data science pedagogy, that is, on *how to teach.* This is the focus of our recently (March 2023) published book by Springer: *Guide to Teaching Data Science: An Interdisciplinary Approach*, described in this paper.

The guide can be used by all educators in all educational environments and settings: in formal education (from elementary schools through high schools to academia) and informal education, in industry, and in non-profit governmental and third sector organizations. Specifically, the guide can be used as a textbook for Methods of Teaching Data

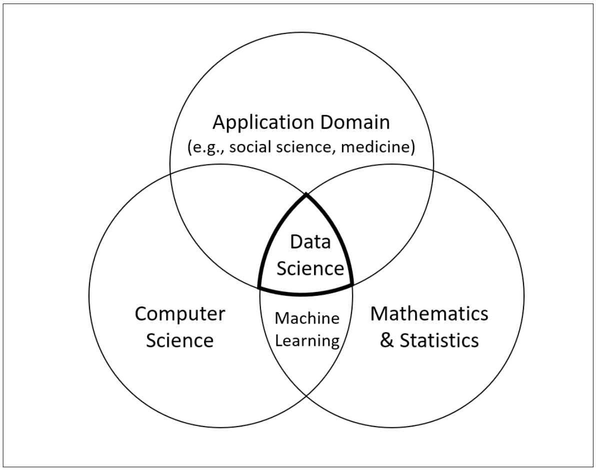

Fig. 1. The authors' version of the data science Venn diagram, as inspired by Conway (2013).

Science courses, in which prospective and in-service teachers learn the pedagogy of data science, which is currently emerging in parallel to the development of the discipline of data science.

To benefit all of its potential user populations, the guide is organized in a way that enables immediate application of its main ideas. This goal is achieved by presenting the rationale behind the inclusion of each topic presented in this guide, its background, development, and importance in the context of data science and data science education, as well as the details of the actual teaching process (including over 200 exercises, work-sheets, topics for discussions, and more).

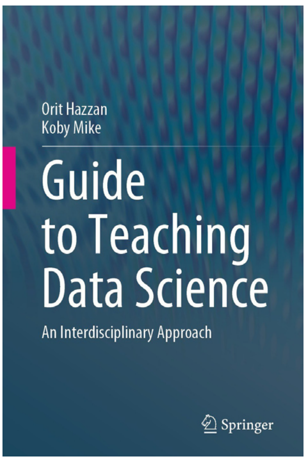

Fig. 2. Guide to Teaching Data Science: An Interdisciplinary Approach – Front Cover.

Description of the Guide parts and chapters

The guide is divided into five parts.

- **Part A – Overview of Data Science and Data Science Education.** In this part, we discuss what data science and data science education are and review the current state of data science education, including curricula and pedagogy.
- **Part B – Opportunities and Challenges of Data Science Education.** This part of the guide elaborates on the educational challenges of data science, from a variety of perspectives (including learners, teachers, and policy makers among other), addressing also the multi-faceted and interdisciplinary nature of data science.
- **Part C – Teaching Professional Aspects of Data Science.** In this part, we examine several topics related to the human aspects of data science, such as data science skills and social issues, in general, and ethics, in particular. We highlight the message that data science skills and other non-technical issues should be given special attention in any data science program regardless of its framework or level.
- **Part D – Machine Learning Education.** We dedicate one part of this guide to machine learning education for two main reasons. First, machine learning is one of the central steps in the data science workflow and an important and central emerging approach for data modeling. Second, machine learning is heavily based on mathematics and computer science content, and therefore poses unique pedagogical challenges that we address in this part of the guide. Specifically, we discuss teaching machine learning algorithms using a white box approach, teaching core concepts of machine learning algorithms that are commonly taught in introductory data science courses, and specific teaching methods suitable for teaching machine learning.
- **Part E – Frameworks for Teaching Data Science.** This part of the guide presents several frameworks for teaching data science to professionals whose core activities are management, education, or research, and who need data science knowledge to foster their professional development and improve their professionalism.

In what follows we present the book chapters along with 1 or 2 illustrative exercises from each chapter. From the variety of over 200 exercises included in the guide, we chose to present in this paper exercises that fit the readership of the *Olympiads in Informatics* journal.

Chapter 1. Introduction – What Is This Guide About?

In the introduction, we present the motivation for writing this guide, followed by the pedagogical principles we applied in it, its structure and how it can be used by educators who teach data science in different educational frameworks. We also present several main kinds of learning environments that are appropriate for teaching and learning data science: Textual programing environments for data science' such as, Jupyter Notebook[1] and Google Colab[2] and visual programing environments for data science, such as, Orange Data Mining[3], KNIME[4], and Weka[5].

Illustrative exercise

Reflection

Reflect on what you have read so far:

(a) What pedagogical ideas were introduced in this chapter?
(b) Can you speculate how *you* will use this guide when you teach data science (now or in the future)?

Part A – Overview of Data Science and Data Science Education

This part includes the following chapters:

Chapter 2: What is Data Science?
Chapter 3: Data Science Thinking
Chapter 4: The Birth of a New Discipline: Data Science Education

Chapter 2. What is Data Science?

Data science integrates knowledge and skills form several disciplines, namely computer science, mathematics, statistics, and an application domain. One way to present such a relationship is using a Venn diagram, which is a diagram that shows logical relationships between different sets (See Fig. 1).

Although many attempts have been made to define data science, such a definition has not yet been reached. One reason for the difficulty to reach a single, consensus definition for data science is its multifaceted nature: it can be described as a science, as a research method, as a discipline, as a workflow, or as a profession. One single definition just cannot capture this diverse essence of data science. In this chapter, we first review the background for the development of data science. Then we present data science from several perspectives: data science as a science, data science as a research method, data science as a discipline, data science as a workflow, and data science as a profession (Mike and Hazzan, 2023). We conclude by highlighting three main characteristics of data science: interdisciplinarity, learner diversity, and its research-oriented nature.

[1] https://jupyter.org/
[2] https://colab.research.google.com/
[3] https://orange.biolab.si/
[4] https://www.knime.com/
[5] https://www.cs.waikato.ac.nz/ml/index.html

Illustrative exercise

> **Pedagogical implications of the multi-faceted analysis of data science**
>
> What pedagogical implications can you draw from the analysis of data science as a science? as a research method? as a discipline? as a workflow? as a profession? What mutual relationships exist between these implications?

Chapter 3. Data Science Thinking

This chapter highlights the cognitive aspect of data science. It presents a variety of modes of thinking, which are associated with the different components of data science, and describes the contribution of each one to data thinking – the mode of thinking required of data scientists (not only professional ones). Indeed, data science thinking integrates the thinking modes associated with the various disciplines that make up data science. Specifically, computer science contributes computational thinking, statistics contributes statistical thinking, mathematics adds different ways in which data science concepts can be conceived, and each application domain brings with it its thinking skills, core principles, and ethical considerations. Finally, we present data thinking. The definition of data science inspires the message that processes of solving real-life problems using data science methods should not be based only on algorithms and data, but also on the application domain knowledge (Mike *et al.*, 2022).

Illustrative exercise

> **Additional modes of thinking required for data science**
>
> Different publications on data science skills mention different thinking skills as being required in order to deal meaningfully with data science. These include, among others, analytical thinking, critical thinking, and data literacy.
>
> Explore these thinking skills (and others you may find) as well as the interconnections between them and the various thinking skills presented in this chapter.

Chapter 4. The Birth of a New Discipline: Data Science Education

Data science is a young field of research and its associated educational knowledge – data science education – is even younger. As of the time of writing this book, data science education has not yet gained recognition as a distinct field and is mainly discussed in the context of the education of the disciplines that make up data science, i.e., computer science education, statistics education, mathematics education, and the educational fields of the applications domains, such as medical education, business analytics education, or learning analytics. In this chapter, we present the story of the birth of the field of data science education by describing its short history. We focus on the main efforts invested in the design of an undergraduate data science curriculum, and on the main initiatives aimed at tailoring a data science curriculum for school pupils. We also suggest several meta-analysis exercises that examine these efforts.

Illustrative exercise

Didactic transposition in data science

Didactic Transposition is a concept that refers to the process of adopting knowledge used by practitioners for teaching purposes (Chevallard, 1989). The term was first coined in the context of mathematics education, in which it refers to the process by which formal mathematics is adapted to fit school teaching and learning. For example, the introduction of a proof using two columns, one for the statement and the other for reasoning, represents *"a didactic transposition from abstract knowledge about mathematical proofs"* (Kang and Kilpatrick, 1992, p. 3).

(a) Suggest several examples of didactic transpositions of formal mathematics to school mathematics. Reflect: What guidelines did you follow?
(b) Read the paper by Hazzan et al (2010) in which the authors demonstrate didactic transpositions of software development methods to educational frameworks. What are the paper's main messages?
(c) Suggest possible didactic transpositions of data science concepts from the academia to high school and elementary school.

Part B – Opportunities and Challenges of Data Science Education

This part includes the following chapters:

Chapter 5: Opportunities in Data Science Education
Chapter 6: The Interdisciplinarity Challenge
Chapter 7: The Variety of Data Science Learners
Chapter 8: Data Science as a Research Method
Chapter 9: The Pedagogical Chasm in Data Science Education

Chapter 5. Opportunities in Data Science Education

Data science education opens up multiple new educational opportunities. In this chapter, we elaborate on six such opportunities: teaching STEM in a real-world context, teaching STEM with real-world data, bridging gender gaps in STEM education, teaching 21st century skills, interdisciplinary pedagogy, and professional development for teachers. We conclude with an interdisciplinary perspective on the opportunities of data science education.

Illustrative exercise

Teaching the STEM subjects in a real-world context

Review the different topics you teach as part of your disciplinary teaching.

Select one of these topics and determine whether or not you currently teach it in the context of real world. If you teach it in a real-world context, choose another topic. Repeat this process until you find a topic that you do not teach in a real-world context.

Describe how you currently teach this topic and design a new teaching process for it, in a real-world context. Compare the two teaching processes. What are your conclusions? Suggest some general guidelines for teaching different subject matters in a real-world context.

Chapter 6. The Interdisciplinarity Challenge

In this chapter, we elaborate on the challenges posed by the interdisciplinary structure of data science. First, we describe the unique and complex interdisciplinary structure of data science. Then, we present the challenge of balancing computer science and statistics in data science education, and the challenge of actually integrating the application domain knowledge into data science study programs, courses, and student projects.

Illustrative exercise

Data science PCK

Imagine you are a data science teacher. Describe your teaching environment, according to your choice: characterize the students, define the study program, describe the physical learning environment, etc.

What pedagogical-content knowledge PCK (Shulman, 1986) would you need in order to teach this class? Describe scenarios in which this PCK might be expressed in your teaching.

Chapter 7. The Variety of Data Science Learners

Since data science is considered to be an important 21st century skill, it should be acquired by everyone - children as well as adults – on a suitable level, to a suitable breadth, and to a suitable depth. And so, after reviewing the importance of data science knowledge for everyone, this chapter reviews the teaching of data science to different populations: K-12 pupils in general and high school computer science pupils in particular, undergraduate students, graduate students, researchers, data science educators, practitioners in the industry, policy makers, users, and the general public. For each population, we discuss the main purpose of teaching it data science, main concepts that the said population should learn and (in some cases) learning environments and exercises that fit it.

Illustrative exercises

The AI + Ethics Curriculum for Middle School initiative

The "AI + Ethics Curriculum for Middle School" initiative, presented at https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/, focuses on artificial intelligence. It seeks to develop an open-source curriculum for middle school students that is made up of a series of lessons and activities.

Explore the activities proposed by this initiative.

In your opinion, what were the pedagogical guidelines applied in the development of these activities? Can these guidelines be applied in the development of learning material that focuses on other data science topics?

Chapter 8. Data Science as a Research Method

In this chapter, we focus on the challenges that emerge from the fact that data science is also a research method. First, we describe the essence of the research process that data science inspires. Then, we present examples of cognitive, organizational, and technological skills which are important for coping with the challenge of data science as a research method, and highlight pedagogical methods for coping with it. In the conclusion of this chapter, we review, from an interdisciplinary perspective, the skills required to perform data science research.

Illustrative exercise

The challenge of leaning the application domain

As can be seen in Chapter 7, a variety of populations nowadays study data science. How can the challenge of familiarity with the application domain be overcome when a specific population studies data science research methods but lacks the required application domain knowledge?

Chapter 9. The Pedagogical Chasm in Data Science Education

As an interdisciplinary discipline, data science poses many challenges for teachers. This chapter presents the story of one of them, specifically of the adoption of a new data science curriculum developed in Israel for high school computer science pupils, by high school computer science teachers. We analyze the adoption process using the diffusion of innovation and the crossing the chasm theories. Accordingly, we first present the diffusion of innovation theory and the crossing the chasm theory. Then, we present the data science for high school curriculum case study. Data collected from teachers who learned to teach the program reveals that when a new curriculum is adopted, a *pedagogical chasm* might exist (i.e., a pedagogical challenge that reduces the motivation of most teachers to adopt the curriculum) that slows down the adoption process of the innovation. Finally, we discuss the implications of the pedagogical chasm for data science education.

Illustrative exercise

Reflection on your experience with the adoption of innovation

According to the Diffusion of Innovation theory (Rogers, 1962), innovations spread in society by flowing from one of the following five distinct groups of adopters to the next: Innovators, early adopters, early majority, late majority, laggards.

(a) Explore the characteristics of each group.

(b) Reflect on your personal experience with the adoption of innovations. What group did you belong to in each case? Was it the same group? Were they different groups? What can you learn about your personality as an adaptor of innovation?

Part C – Teaching Professional Aspects of Data Science

This part includes the following chapters:

Chapter 10: The Data Science Workflow
Chapter 11: Professional Skills and Soft Skills in Data Science
Chapter 12: Social and Ethical Issues of Data Science

Chapter 10. The Data Science Workflow

The examination of data science as a workflow is yet another facet of data science. In this chapter we elaborate on the data science workflow from an educational perspective. First, we present several approaches to the data science workflow, following which we elaborate on the pedagogical aspects of the different phases of the workflow: data collection, data preparation, exploratory data analysis, modeling, and communication and action. We conclude with an interdisciplinary perspective on the data science workflow.

Illustrative exercises

Data preparation

Search for a dataset in an application domain you are familiar with. Review the data. Can you find erroneous data? Can you find outliers?

Search for a dataset in an application domain you are not familiar with. Review the data. Can you find erroneous data? Can you find outliers?

(a) Analyze differences (if such exist) between the results in the two cases: the familiar application domain and the unfamiliar application domain.

(b) Reflect on how you performed this exercise: How did you look for erroneous data? How did you look for outliers? Did you use any resources in these processes? If you did, which resources? If not, why not? Can these processes be improved? If yes, how? If not, why?

Chapter 11. Professional Skills and Soft Skills in Data Science

Abstract In this chapter, we highlight skills that are required to deal with data science in a meaningful manner. The chapter describes two kinds of skills: professional skills and soft skills. Professional skills are specific skills that are needed in order to engage in data science, while soft skills are more general skills that acquire unique importance in the context of data science. In each section, we address both cognitive, organizational, and technological skills.

Illustrative exercises

Critical thinking

One of the most important cognitive skills required for dealing meaningfully with data science ideas is critical thinking. Propose a case study of a data science project in which decision-making processes, which were not accompanied with critical thinking processes, led to a chain of undesirable events.

Chapter 12. Social and Ethical Issues of Data Science

The teaching of social issues related to data science should be given special attention regardless of the framework or level at which data science is taught. This assertion is derived from the fact that data science (a) is relevant for many aspects of our lives (such as health, education, social life, and transportation); (b) can be applied in harmful ways (even without explicit intention); and (c) involves ethical considerations derived from the application domain. Of the many possible social topics whose teaching might have been discussed in this chapter, we focus on data science ethics. We also present teaching methods that are especially appropriate for the teaching of social issues of data science.

Illustrative exercise

Famous cases that illustrate the need for an ethical code for data science

(a) Locate resources about the two cases mentioned above that illustrate the need for an ethical code for data science. List the ethical considerations involved in each case.

(b) Find additional cases that illustrate the importance of an ethical code for data science.

Part D – Machine Learning Education

This part comprises the following chapters:

Chapter 13. The Pedagogical Challenge of Machine Learning Education

Machine learning (ML) is the essence of the modeling phase of the data science workflow. In this chapter, we focus on the pedagogical challenges of teaching ML to various populations. We first describe the terms *white box* and *black box* in the context of ML education. Next, we describe the pedagogical challenge with respect to different learner populations including data science major students as well as non-major students. Then, we present three framework remarks for teaching ML (regarding statistical thinking, interdisciplinary projects, and the application domain knowledge), which are important to be kept in mind in ML teaching processes. We conclude this chapter by highlighting the importance of ML education in the context of the application domain.

Illustrative exercise

The concepts of *explainability and interpretability*

Search the web and find 3-5 stories that exemplify the concepts of *explainability* and *interpretability*.

(a) For each story, identify its main actors, the ML algorithm it refers to, the context in which these concepts are discussed, the end of the story, and what conclusions are drawn (if at all).
(b) Add your conclusions from the examination of each story.
(c) Formulate three guidelines for users of ML methods.
(d) Reflect on what you have learned while working on this exercise. What would you do differently if you were asked to repeat it?
(e) What conclusions can you draw for your own usage of ML results?

Chapter 14. Core Concepts of Machine Learning

In this chapter, we focus on the teaching of several core concepts that are common to many machine learning (ML) algorithms (such as hyper-parameter tuning) and, as such, are essential learning goals in themselves, regardless of the ML algorithms. Specifically, we discuss types of ML, ML parameters and hyperparameters, model training, validation, and testing, ML performance indicators, bias and variance, model complexity, overfitting and underfitting, loss function optimization and the gradient descent algorithm, and regularization. We conclude this chapter by emphasizing what ML core concepts should be discussed in the context of the application domain.

Illustrative exercise

True or false

(a) Define the concepts true-positive, true-negative, false-positive, and false-negative. Which of these concepts does the medical diagnosis problem use?

(b) Select a problem from any domain of life, whose formulation includes these concepts. Formulate it in two ways: using frequencies and using probabilities (percentages).

(c) If you are working in a team, the team can discuss the different problems, addressing questions such as: In what context is each formulation clearer? Was it easy to transition between the two formulations? Why?

Chapter 15. Machine Learning Algorithms

In this chapter, we describe the teaching of several machine learning (ML) algorithms that are commonly taught in introduction to ML courses, and analyze them from a pedagogical perspective. The algorithms we discuss are the K-nearest neighbors (KNN), decision trees, Perceptron, linear regression, logistic regression, and neural networks.

The exercises in this chapter are based on explorations whose length and depth are beyond the scope of this paper.

Chapter 16. Teaching Methods for Machine Learning

In this chapter, we review four teaching methods for machine learning: visualization, hands-on tasks, programming tasks, and project-based learning. When relevant, as part of the presentation of these pedagogical tools, we analyze them from the perspective of the process-object duality theory and the reduction of abstraction theory.

The exercises in this chapter are based on explorations whose length and depth are beyond the scope of this paper.

Part E – Frameworks for Teaching Data Science

This part includes the following chapters:

Chapter 17: Data Science for Managers and Policymakers

Chapter 18: Data Science Teacher Preparation: The "Method for Teaching Data Science" Course

Chapter 19: Data Science for Social Science and Digital Humanities Research
Chapter 20: Data Science for Research on Human Aspects of Science and Engineering

Chapter 17. Data Science for Managers and Policymakers

In this chapter, we describe a workshop for policy makers that focuses on the integration of data science into education systems for policy, governance, and operational purposes. The messages conveyed in this chapter can be applied in other systems and organizations in all sectors – governmental (the first sector), for-profit organizations (the second sector), and non-profit organizations (the third sector). We conclude with an interdisciplinary perspective on data science for managers and policymakers.

Illustrative exercise

Data culture

Explore the term *data culture*. Use at least three resources that address this concept.

(a) Give five examples of organizations that promote a healthy data culture.

(b) List five characteristics of organizations that promote a healthy data culture.

(c) Describe five practices that employees in originations that promote healthy data cultures should master.

(d) Describe five scenarios involving managers in organizations that promote a healthy data culture that illustrate the relevance of data science for their decision-making processes.
For each scenario, specify the data science knowledge the managers should have and suggest how and what they can learn from this specific data science content.

(e) Explore the concept of exponential organizations (Ismail, 2014). Exponential Organizations: Why new organizations are ten times better, faster, and cheaper than yours (and what to do about it), Diversion Books). How do exponential organizations promote data culture?

Chapter 18. Data Science Teacher Preparation: The "Method for Teaching Data Science" Course

In this chapter, we present a detailed description of the Method for Teaching Data Science (MTDS) course that we designed and taught to prospective computer science teachers at our institution, the Technion – Israel Institute of Technology. Since our goal in this chapter is to encourage the implementation and teaching of the MTDS course in different frameworks, we provide the readership with as many details as possible about the course, including the course environment, the course design, the learning targets and structure of the course, the grading policy and assignments, teaching principles we employed in the course, and a detailed description of two of the course lessons. Full, detailed descriptions of all 13 course lessons are available on our Data Science Education website[6].

Illustrative exercise

> **Topics to be included in a Method of Teaching Data Science course**
>
> Before reading the description of the course lessons, suggest topics that you would include in a Methods of Teaching Data Science course.

Chapter 19. Data Science for Social Science and Digital Humanities Research

In this chapter and in Chapter 20, we describe two data science teaching frameworks for researchers: this chapter addresses researchers in social science and digital humanities; Chapter 20 addresses researchers in science and engineering. Following a discussion of the relevancy of data science for social science and digital humanities researchers, we describe a data science bootcamp designed for researchers in those areas. Then, we present the curriculum of a year-long specialization program in data science for graduate psychology students that was developed based on this bootcamp. Finally, we discuss the data science teaching frameworks for researchers in social science and digital humanities from motivational perspectives and conclude by illuminating the importance of an interdisciplinary approach in designing data science curricula for application domain specialists.

Illustrative exercise

> **Data science applications that require knowledge in social sciences and digital humanities**
>
> (a) Search the web for data science applications whose development required knowledge in social sciences. What do these applications have in common? In what ways are they different?
>
> (b) Search the web for data science applications whose development required knowledge in digital humanities? What do these applications have in common? In what ways are they different?
>
> (c) Are there similarities between the applications that require knowledge in the social sciences and the applications that require knowledge in the digital humanities? If yes – what are the similarities? If not – explain the differences between these two.

[6] https://orithazzan.net.technion.ac.il/data-science-education/

Chapter 20. Data Science for Research on Human Aspects of Science and Engineering

In this chapter and in Chapter 19, we describe two data science teaching frameworks for researchers: Chapter 19 addresses researchers in social science and digital humanities; this chapter addresses science and engineering researchers and discusses how to teach data science methods to science and engineering graduate students to assist them in conducting research on human aspects of science and engineering. In most cases, these target populations, unlike the community of social scientists (discussed in Chapter 19), have the required background in computer science, mathematics, and statistics, and need to be exposed to the human aspects of science and engineering which, in many cases, are not included in scientific and engineering study programs. We start with the presentation of possible human-related science and engineering topics for investigation. Then, we describe a workshop for science and engineering graduate students that can be facilitated in a hybrid format, combining synchronous (online or face to face) and asynchronous meetings. We conclude with an interdisciplinary perspective of data science for research on human aspects of science and engineering.

Illustrative exercise

Data-driven research

(a) Suggest five research topics in *scientific* disciplines that may be initiated by data that is gathered incidentally.

(b) Suggest five research topics in *engineering* disciplines that may be initiated by data that is gathered incidentally.

(c) For each topic, suggest a human-related topic that you would find interesting to investigate.

(d) Select two topics from the human-reacted scientific disciplines and two topics from the human-related engineering disciplines and describe how you would research them.

Epilogue

In the epilogue of the *Guide to Teaching Data Science: An Interdisciplinary Approach*, we view it from a holistic perceptive, reflecting on its big ideas and their interconnections, highlighting the following facts:

- The guide is multifaceted and addresses teaching methods, skills, learners, perspectives, habits of mind, and data science topics – from programming and statistics, through problem solving processes to organizational skills.
- The guide reflects the richness of the discipline of data science, its relatedness to many aspects of our life, and its centrality and potential contribution to the education of future generations in the 21st century.

- The richness of the discipline of data science is reflected in the interconnections between the different chapters of the guide, as they are specified throughout the guide.

Illustrative exercise

Final reflection task

Reflect:

(a) What do you like about data science? What do you dislike about it?

(b) What do you like about data science education? What do you dislike about it?

(c) If you had to formulate one main idea of data science education, what would it be?

(d) What will *your* main new contribution to data science education be?

Conclusion

In this paper we present the content of the *Guide to Teaching Data Science: An Interdisciplinary Approach.* We hope that it reflects the richness of data science and of data science education as emerging disciplines. Supplementary pedagogical material is available in our website at `https://orithazzan.net.technion.ac.il/data-science-education/`. We will be happy to continue the dialogue with the readership of the *Olympiads in Informatics* journal. Specifically, we welcome questions and suggestions for pedagogical tools and teaching methods as well as suggestions for collaboration for the promotion of data science education in the interdisciplinary spirit suggested in the guide.

References

Chevallard, Y. (1989). On didactic transposition theory: Some introductory notes. In: *Proceedings of the International Symposium on Selected Domains of Research and Development in Mathematics Education*. 51–62.

Conway, D. (2013). The data science venn diagram. *Datist.* `http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram`

Hazzan, O., Dubinsky, Y., Meerbaum-Salant, O. (2010). Didactic transposition in computer science education. *ACM Inroads*, 1(4), 33–37.

Ismail, S. (2014). *Exponential Organizations: Why New Organizations are Ten Times Better, Faster, and Cheaper than Yours (and what to do about it)*. Diversion Books.

Kang, W., Kilpatrick, J. (1992). Didactic transposition in mathematics textbooks. *For the Learning of Mathematics*, 12(1), 2–7.

Mike, K., Hazzan, O. (2023). What is Data Science? *Communications of the ACM*, 66(2), 12–13.

Mike, K., Ragonis, N., Rosenberg-Kima, R.B., Hazzan, O. (2022). Computational thinking in the era of data science. *Communications of the ACM*, 65(8), 31–33.

Rogers, E.M. (1962). *Diffusion of Innovations*. Free Press.

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.

O. Hazzan is a faculty member at the Technion's Department of Education in Science and Technology since October 2000. Her research focuses on computer science, software engineering and data science education. Within this framework she researches cognitive and social processes on the individual, the team and the organization levels, in all kinds of organizations. She has published about 130 papers in professional refereed journals and conference proceedings, and eight books. In 2007-2010 she chaired the High School Computer Science Curriculum Committee assigned by the Israeli Ministry of Education. In 2011-2015 Hazzan was the faculty Dean. From 2017 to 2019, Hazzan served the Technion Dean of Undergraduate Studies. Additional details can be found in her personal homepage:
`https://orithazzan.net.technion.ac.il/`

K. Mike is a Ph.D. graduate from the Technion's Department of Education in Science and Technology under the supervision of Professor Orit Hazzan. He continued his a post-doc research on data science education at the Bar-Ilan University, and retains B.Sc. and an M.Sc. in Electrical Engineering from Tel Aviv University. After two decades of professional career is the Israeli hi-tech industry, he returned to academia for his doctoral studies on data science education. As part of his research, Koby developed and taught several data science programs for high school students, high school computer science teachers, and graduate students and researchers in social sciences and digital humanities.