From SQL Struggles to Visual Insights: Revealing Transformations for Data Science Education

Tobias Antensteiner
tobias.antensteiner@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Abstract

Despite data's pervasive role in modern society and the growing emphasis on K-12 data literacy, students struggle with core data transformations. SQL—curricular centerpiece despite known limitations—motivates our design of an educational relational programming language that natively streamlines visualization pipelines and thereby offers visual feedback on data transformations. Through this approach, we aim to foster more accurate mental models of how data transformations relate to visualizations, enabling learners to better make sense of data. We will evaluate this work using design-based research methods.

CCS Concepts

• Social and professional topics → K-12 education; **• Human-centered computing → Visualization systems and tools**; **• Software and its engineering → Domain specific languages**.

Keywords

relational programming language design, data science education, data visualization literacy, notional machines, design-based research

ACM Reference Format:
Tobias Antensteiner. 2026. From SQL Struggles to Visual Insights: Revealing Transformations for Data Science Education. In *Proceedings of the 31st ACM Conference on Innovation and Technology in Computer Science Education V. 2 (ITiCSE 2026), July 10–15, 2026, Madrid, Spain.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3803401.3811954

1 Context and Motivation

Data has become central to modern life, yet many people still struggle to interpret and reason with it, especially in cognitively demanding contexts [24, 28]. This has made data literacy an increasingly important educational goal, encompassing the ability to read, analyze, and argue with data [7, 23].

As data science education expands, research has increasingly focused on K-12 learning opportunities that help students interpret and create data visualizations [4, 9]. Data visualization literacy—the capacity to understand and produce graphical representations of data [6]—is central to making sense of complex information and developing critical, data-driven thinking [10].

However, the field remains underdeveloped, and students often struggle with core data transformations like joining, grouping, and aggregation–central to data analysis and visualization [17–19]. Much of this work has centered on SQL[1], which remains an important tool in data science practice and education [16, 18].

Against this background—and motivated by recent critiques of SQL's limitations [3, 22, 27]—we aim to reconsider relational programming language design from both theoretical and pedagogical perspectives [8, 11, 21].[2]

2 Background

Data visualization literacy—essential for interpreting and producing graphical representations of data [6]—has been defined through conceptual frameworks like Munzner's What–Why–How [21]. These frameworks position visualization as central to data literacy, enabling people to explore, reason about, and communicate insights through data storytelling [1, 20]. Yet teaching visualization remains difficult, particularly in K-12 contexts where learners must integrate data concepts, computational tools, and visual reasoning [4, 9, 10]. Within the broader field of data science education, these challenges are well-recognized [12, 14, 24].

Programming provides a promising pathway for developing data and data visualization literacy. Accessible, inquiry-based environments help students manipulate, analyze, and visualize data through code [7]. Such practices bridge statistical reasoning with computation and support deeper conceptual understanding, allowing learners to construct their own data analyses and create graphical representations [15, 25, 26].

Recent educational tools increasingly integrate visual feedback on intermediate computational steps, leveraging notional machines to help learners understand how data transformations lead to graphical representations and supporting the development of accurate mental models of data pipelines [13, 17]. At the same time, research suggests opportunities for reconsidering programming language design to better align data transformations with visualization workflows and educational needs [3, 8, 11].

3 Research Goals and Methods

Given the difficulty students face in reasoning about data pipelines due to the implicit connection between data transformations and

ACM ISBN 979-8-4007-2633-0/2026/07
https://doi.org/10.1145/3803401.3811954

[1] Although "data scientists often use multiple tools and create custom ones for specific purposes" [14], Python, R, and SQL remain among the most widely used programming languages in computational data science [7]—with SQL serving as the primary, universally supported language for structured data processing [27].

[2] This extended abstract summarizes our ongoing work. We make full details, materials, and prototypes available on `osf.io/3kwth` [2].

visual outcomes, our goal is to develop a theoretically grounded educational relational programming language that simplifies core relational programming features and integrates visualization pipelines from both theoretical and pedagogical perspectives [8, 11, 21].

In doing so, we also seek to explore the pedagogical integration of visual feedback on data transformations into existing learning environments [13]. While design-based research methods are intended [5]—including theoretical assessment based on notional machines [17], expert interviews, and iterative classroom pilots with pupils and first-year university students—the precise methodological framework remains under consideration.

4 Contributions

Driven by the belief that, in a data-centric society, it is essential for all individuals to become literate in working with, understanding, and interpreting data [6, 7, 23], we

(1) promote research and practice in data science and data visualization education by developing a theoretically grounded educational relational programming language [8, 11, 21];
(2) propose a pedagogical approach that integrates visual feedback on data transformations to support learners' mental models [9, 13, 17]; and
(3) provide empirical insights into novice learners' understanding of data transformations and visualization pipelines via qualitative studies [5].

References

[1] Rohan Alexander. 2023. *Telling Stories with Data: With Applications in R* (1 ed.). CRC Press, Boca Raton, USA. doi:10.1201/9781003229407

[2] Tobias Antensteiner. 2026. *Revealing Transformations for Data Science Education*. doi:10.17605/OSF.IO/3KWTH

[3] Molham Aref, Paolo Guagliardo, George Kastrinis, Leonid Libkin, Victor Marsault, Wim Martens, Mary McGrath, Filip Murlak, Nathaniel Nystrom, Liat Peterfreund, Allison Rogers, Cristina Sirangelo, Domagoj Vrgoč, David Zhao, and Abdul Zreika. 2025. Rel: A Programming Language for Relational Data. In *Companion of the 2025 International Conference on Management of Data*. ACM, Berlin, Germany, 283–296. doi:10.1145/3722212.3724450

[4] Benjamin Bach, Mandy Keck, Fateme Rajabiyazdi, Tatiana Losev, Isabel Meirelles, Jason Dykes, Robert S. Laramee, Mashael AlKadi, Christina Stoiber, Samuel Huron, Charles Perin, Luiz Morais, Wolfgang Aigner, Doris Kosminsky, Magdalena Boucher, Søren Knudsen, Areti Manataki, Jan Aerts, Uta Hinrichs, Jonathan C. Roberts, and Sheelagh Carpendale. 2024. Challenges and Opportunities in Data Visualization Education: A Call to Action. *IEEE Transactions on Visualization and Computer Graphics* 30, 1 (2024), 649–660. doi:10.1109/TVCG.2023.3327378

[5] Arthur Bakker and Dolly van Eerde. 2015. An Introduction to Design-Based Research with an Example From Statistics Education. In *Approaches to Qualitative Research in Mathematics Education*, Angelika Bikner-Ahsbahs, Christine Knipping, and Norma Presmeg (Eds.). Springer Netherlands, Dordrecht, Netherlands, 429–466. doi:10.1007/978-94-017-9181-6_16

[6] Katy Börner, Andreas Bueckle, and Michael Ginda. 2019. Data Visualization Literacy: Definitions, Conceptual Frameworks, Exercises, and Assessments. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1857–1864. doi:10.1073/pnas.1807180116

[7] Ruijia Cheng, Aayushi Dangol, Frances Marie Tabio Ello, Lingyu Wang, and Sayamindu Dasgupta. 2023. Concepts, Practices, and Perspectives for Developing Computational Data Literacy: Insights from Workshops with a New Data Programming System. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. ACM, Chicago, USA, 100–111. doi:10.1145/3585088.3589364

[8] Jens Dittrich. 2025. A Functional Data Model and Query Language Is All You Need. doi:10.48550/arXiv.2507.20671

[9] Cassia Fernandez, Paulo Blikstein, and Roseli De Deus Lopes Lopes. 2024. Design Failures in Data Visualization Programming Activities. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*. ACM, Delft, Netherlands, 574–586. doi:10.1145/3628516.3663348

[10] Cassia Fernandez, João Adriano Freitas, Paulo Blikstein, and Roseli De Deus Lopes. 2025. The Design Space of Visualization Tools for Data Science Education: Literature Review and Framework for Future Designs. *International Journal of Child-Computer Interaction* 43 (2025), 1–18. doi:10.1016/j.ijcci.2024.100698

[11] Wolfgang Gatterbauer and Miguel Diandre. 2026. Database Research Needs an Abstract Relational Query Language. In *Proceedings of the 16th Conference on Innovative Data Systems Research*. CIDR, Chaminade, USA, 1–16.

[12] Orit Hazzan and Koby Mike. 2023. *Guide to Teaching Data Science: An Interdisciplinary Approach*. Springer International Publishing, Cham. doi:10.1007/978-3-031-24758-3

[13] Sam Lau, Sean Kross, Eugene Wu, and Philip J. Guo. 2023. Teaching Data Science by Visualizing Data Table Transformations: Pandas Tutor for Python, Tidy Data Tutor for R, and SQL Tutor. In *Proceedings of the 2nd International Workshop on Data Systems Education: Bridging Education Practice with Education Research*. ACM, Seattle, USA, 50–55. doi:10.1145/3596673.3596972

[14] Victor R. Lee. 2025. *Advancing Data Science Education in K-12: Foundations, Research, and Innovations* (1 ed.). Routledge, New York, USA. doi:10.4324/9781003385264

[15] Aleksi Lukkarinen and Juha Sorva. 2016. Classifying the Tools of Contextualized Programming Education and Forms of Media Computation. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*. ACM, Koli, Finland, 51–60. doi:10.1145/2999541.2999551

[16] Daphne Miedema. 2023. Toward a Fundamental Understanding of SQL Education. In *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume 2*. ACM, Chicago, USA, 64–68. doi:10.1145/3568812.3603454

[17] Daphne Miedema, Martin Goodfellow, Chandrika Satyavolu, Georgiana Haldeman, Leonard Busuttil, Laura Farinetti, Giovanna Guerrini, Yuhan Pan, Sujeeth Goud Ramagoni, Raja Sooriamurthi, and Xiaoying Tu. 2025. Explanations of Data Systems Concepts in CS Education: An Updated View on Notional Machines. In *Proceedings of the 2025 Working Group Reports on Innovation and Technology in Computer Science Education*. ACM, Nijmegen, Netherlands, 108–135. doi:10.1145/3760545.3783971

[18] Daphne Miedema, Michael Liut, George H. L. Fletcher, and Efthimia Aivaloglou. 2023. "There Is No Ambiguity on What to Return": Investigating the Prevalence of SQL Misconceptions. In *Proceedings of the 23rd Koli Calling International Conference on Computing Education Research*. ACM, Koli, Finland, 1–12. doi:10.1145/3631802.3631821

[19] Daphne Miedema, Toni Taipalus, Vangel V. Ajanovski, Abdussalam Alawini, Martin Goodfellow, Michael Liut, Svetlana Peltsverger, and Tiffany Young. 2025. Data Systems Education: Curriculum Recommendations, Course Syllabi, and Industry Needs. In *2024 Working Group Reports on Innovation and Technology in Computer Science Education*. ACM, Milan, Italy, 95–123. doi:10.1145/3689187.3709609

[20] Mikaela Elizabeth Milesi, Paola Mejia-Domenzain, Laura Brandl, Vanessa Echeverria, Yueqiao Jin, Dragan Gasevic, Yi-Shan Tsai, Tanja Käser, and Roberto Martinez-Maldonado. 2025. "Piecing Data Connections Together Like a Puzzle": Effects of Increasing Task Complexity on the Effectiveness of Data Storytelling Enhanced Visualisations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama, Japan, 1–18. doi:10.1145/3706598.3714270

[21] Tamara Munzner. 2014. *Visualization Analysis and Design* (1st ed.). A K Peters/CRC Press, New York, USA. doi:10.1201/b17511

[22] Thomas Neumann and Viktor Leis. 2024. A Critique of Modern SQL and a Proposal Towards a Simple and Expressive Query Language. In *Proceedings of the 14th Conference on Innovative Data Systems Research*. CIDR, Chaminade, USA, 1–7.

[23] Viktoriya Olari and Ralf Romeike. 2021. Addressing AI and Data Literacy in Teacher Education: A Review of Existing Educational Frameworks. In *The 16th Workshop in Primary and Secondary Computing Education*. ACM, Virtual Event Germany, 1–2. doi:10.1145/3481312.3481351

[24] Alex Young Pedersen and Francesco Caviglia. 2019. Data Literacy as a Compound Competence. In *Digital Science*, Tatiana Antipova and Alvaro Rocha (Eds.). Vol. 850. Springer International Publishing, Cham, Switzerland, 166–173. doi:10.1007/978-3-030-02351-5_21

[25] Susanne Podworny, Sven Hüsing, and Carsten Schulte. 2022. A Place for a Data Science Project in School: Between Statistics and Epistemic Programming. *Statistics Education Research Journal* 21, 2 (2022), 1–6. doi:10.52041/serj.v21i2.46

[26] Sadia Sharmin. 2022. Creativity in CS1: A Literature Review. *ACM Transactions on Computing Education* 22, 2 (2022), 1–26. doi:10.1145/3459995

[27] Jeff Shute, Shannon Bales, Matthew Brown, Jean-Daniel Browne, Brandon Dolphin, Romit Kudtarkar, Andrey Litvinov, Jingchi Ma, John Morcos, Michael Shen, David Wilhite, Xi Wu, and Lulan Yu. 2024. SQL Has Problems. We Can Fix Them: Pipe Syntax In SQL. *Proceedings of the VLDB Endowment* 17, 12 (2024), 4051–4063. doi:10.14778/3685800.3685826

[28] Justice Walker and Emmanuel Schanzer. 2025. Twenty Constructionist Things to Do With Data Science. *Constructionism Conference Proceedings* 8 (2025), 137–150. doi:10.21240/constr/2025/95.X