International Society of the Learning Sciences

Dataland: An Informed, Situated, and Critical Approach to Data Literacy

Lingyu Wang, University of North Carolina at Chapel Hill, <lyw@unc.edu>
Sayamindu Dasgupta, University of Washington, <sdg1@uw.edu>[1]

Abstract: In this demo, drawing from Kafai and Proctor's (2021) framework, we propose an *informed, situated, and critical* approach to data literacy, and present *Dataland*, a block-based language and computational notebook system we are developing for young learners to create engaging and interactive data narratives. Addressing the three framings of computational thinking–cognitive, situated, and critical–*Dataland* aims to help learners gain basic knowledge, invite them to bring in personal experiences, and encourage them to critically reflect on data. *Dataland* is a combination of three key parts: easy-to-use block-based programming, narrative support inspired by environments like Jupyter Notebook and blogging tools, and interactive projects that engage learners through puzzle-solving and puzzle-designing. *Dataland* is critical and unique in its ability to help learners both get familiar with data and reflect on their social, cultural, and epistemological implications in a fun and flexible setting.

Introduction

The past few years have seen steadily increasing interests in engaging young people, specifically middle and high school students, in data literacies (Lee & Wilkerson, 2018) . While such endeavors used to focus on teaching solely technical concepts and skills, more recent works have shifted from teaching individualized *cognitive* elements towards forming *situated* knowledges (identity, community, practice, and participation) and reflecting on *critical* contexts (broader social and cultural formations) (Kafai & Proctor, 2021). As computation and data become increasingly involved in social life and public discussions, scholars and the general public alike are reexamining how data gets created, narrated, and mobilized, and existing tools have been revamped and extended, encouraging learners to incorporate personal, social, and pedagogical considerations into data analytic works (Dourisch & Cruz, 2018; Feinberg, 2017; Wilkerson et al., 2021).

In this we see an opportunity to develop *Dataland*, an educational system for data programming and storytelling. In its core design, the system incorporates all three key framings that Kafai and Proctor (2021) argue for--cognitive framing, situated framing, and critical framing. We have thus developed an informative, approachable, and flexible system that invites and encourages young learners to 1) understand how data-centered tools and statistical knowledge work, 2) bring such understanding into conversation with their lived experiences and personal interests, and 3) reflect on the relations between their work and the broader society: racial and gender profiling, cultural appropriation, social engineering, criminalization, colonization, etc. In all, we develop *Dataland* to approach data literacy in an *informed, situated, and critical* way.

To show how *Dataland* works, in our proposed demo, we intend to present its core programming grammar, prototype interface, and sample data stories. Best described as a codeblock-and-notebook system, *Dataland* is designed to be an organic combination of three key parts: easy-to-use block-based programming for data analyses and visualization, narrative writing environment inspired by computational notebooks and contemporary blogging tools, and interactive components that engage learners through puzzle-solving and potentially puzzle-designing. As a result, *Dataland* is unique in its ability to help students both get familiar with data and reflect on their social, cultural, and epistemological implications in a fun and flexible setting.

Background

Kafai and Proctor (2021) define three major framings–cognitive, situated, and critical–of computational thinking in this way: the *cognitive* framing sees learning as the acquisition of knowledge and skills, orienting the learner towards career requirements; the *situated* framing sees learning as identity formation through participation in disciplinary practices, so the emphasis is to encourage creative output and social engagement; finally, the *critical* framing sees learning as developing an understanding of how realities are shaped, under which framing the major goal will be to create strategies against marginalization and oppression. For Kafai and Proctor, all three framings are key to educational efforts in computational thinking, and in our project, we similarly strive to approach data literacy education in an *informed, situated, and critical* way.

[1] Dasgupta was at the University of North Carolina at Chapel Hill when this work was submitted for review. Since then, he has moved to the University of Washington.

Dataland, gaining its namesake from the early constructionist notion of "Mathland" (Papert, 1980) aims to acquaint young learners of *informed, situated, and critical* data literacy in a hands-on experience. Working with Papert's argument that the best way to learn and know a language is by being immersed in its natural environment, we also argue that learning data literacy by real-world use, instead of learning by grammar and vocabulary, is the natural environment for the kind of learning we hope to approach.

In *Dataland*, the informed (cognitive) aspect of data literacy is approached through the design of a low-floor, wide-wall visual coding environment. According to Papert (1980), living in a computer-based learning environment where the language of mathematics is the natural medium of communication, learners will soon pick up fluency in mathematics just like people living in France picking up the French language. Similarly, immersing learners in an approachable, exploratory environment where they can play with data as a construction material allows learners to naturally engage with the fundamental knowledge and skills underlying data analytic tools. In this way, *Dataland* allows learners to casually and immersively engage with abstract ideas that pertain to data analysis, fulfilling the first cognitive framing.

To step beyond the basic knowledge and skills, *Dataland* also addresses the situational aspect of data literacy. To do this, we design and develop *Dataland*'s sample data story with the concept of empathetic perspective taking in mind. One example along this line is the feminist-inspired *DoggyVision* environment Kelly et al. (2020) that allows family participants to explore what a dog may see through its eyes. This idea of perspective-taking as an everyday exercise in empathy speaks a lot to the value of affect-situated, emotional, and embodied responses-which is usually left out of the abstract, scientific hierarchy of knowledge. Being able to take the perspective and empathize is a key step towards situated knowledge.

Finally, to address the critical framing, *Dataland*'s narrative functions are designed with epistemological pluralism in mind. Employed by Turkle and Papert (1990) discussing the feminist approach to computer science education, epistemological pluralism argues for a re-evaluation of different styles of scientific thought and different ways to approach a computational object. Turkle and Papert especially argue for concrete, non-hierarchical, painterly and "soft" ways of using programming codes and user interface components, instead of seeing them only in the conventional, supposedly "hard" logics of engineering. By incorporating narrative elements into data analysis, *Dataland* actively engages with epistemological pluralism and invites learners to reflect on the nature of data.

Building upon these previous works, *Dataland* aims to help learners gain the basic cognitive understanding, while stepping towards situated and critical ways of knowing. Kafai and Proctor (2021) imagined that different communities will come up with different articulations of computational knowledge, approaching the knowledge as an integral part of broader social reality, and adapting the knowledge to local needs. To achieve this goal, we propose, learners must immerse themselves in the environment, recognize their personal experiences, and relate to each other in a reflexive, critical way. In *Dataland*, immersion, empathy, and reflection are made possible by the presence of narrative and interactive components, which we describe next.

System Design

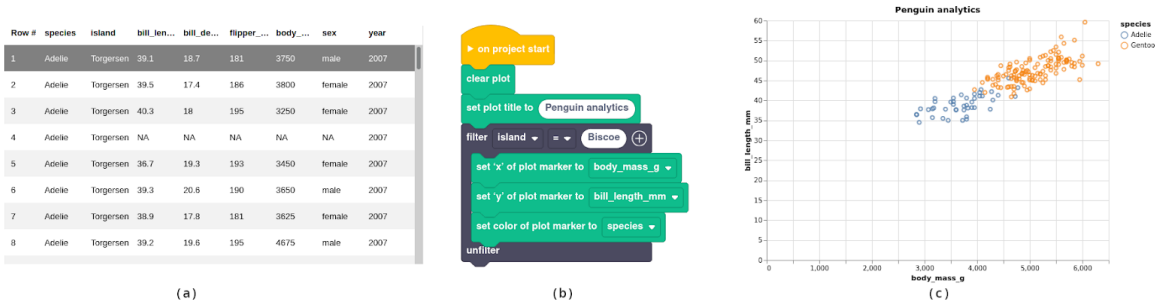
The idea of engaging young people in analyzing and visualizing data in ways that connect to their personal interests has been explored in learning and design research before. For example, the Quantified Self movement—a global trend where sensors on mobile and wearable technologies are used to collect data on one's own everyday activities—has been seen as a rich environment for learning with and about data (Lee, 2013). The Common Online Data Analysis Platform (CODAP) (2014) system allows students of grades 6–12 to access and explore data from a wide variety of sources. Dasgupta and Hill (2017) designed Scratch Community Blocks, a system built on top of the Scratch programming environment and online community (Resnick et al., 2009) that allowed Scratch community members to access, analyze, and visualize their own data from the Scratch online community.

Language design

Data operations in *Dataland* are centered around tabular data that can be imported with a button click. Currently Comma-Separated Values (CSV) files can be imported. In a way similar to Scratch, the programming language in *Dataland* uses visual blocks to represent programming primitives in the code editor. Programs in *Dataland* are constructed by dragging and snapping together blocks which have specific shapes to allow for only certain possibilities of composition—this minimizes the possibility of syntax errors. The block-based *Dataland* programming language supports, in addition to standard programming constructs (e.g., conditional statements, arithmetic operators), capabilities for selecting, filtering, and aggregating data, as well as capabilities for generating visualizations. In contrast to Scratch Community Blocks, whose sole focus was data from the Scratch online community, *Dataland* provides a more general grammar for accessing and working with data.

We draw from design principles used for Scratch, especially those articulated by Resnick and Silverman (2005) in the design of the data selection, filtering and aggregation capabilities of the *Dataland* programming system. For visualization capabilities, we draw from Wilkinson (2005) and related work by Satyanarayan et al. (2017) which systematically examines the elements of statistical graphs and establishes a high-level syntax to produce a wide variety of graphs from a small set of well-chosen basic specifications. In Figure 1 we show the programming language in action, with the resultant visualization output.

Figure 1: *Dataland's* programming interface. The data table (a) shows the Palmer Islands penguins dataset (Horst et al., 2020). The code (b) filters the data and produces a scatter plot shown in (c).



Narrative data: Notebook interface and a sample data story

Beyond the block-based coding interface, we have been exploring ways to make data narratable and relatable for learners throughout our research. Our goal is to emotionally engage learners in a personal, embodied way. This emotional dimension, we hypothesize, can be approached through storytelling, whether fictional or non-fictional, and doing this will be key to opening up towards situated and critical knowledge. This is suggested by the fact that a popular genre of projects in Scratch are games and interactive stories, and we also draw inspirations from various mini games on Code.org, the SQL Murder Mystery game made by Knight Lab at Northwestern University (Park & He, 2018), as well as computational notebooks like the popular Jupyter Notebok, and blogging tools like WordPress (WordPress.org, 2021). It is also worth noting that researchers are also designing and testing a new Stories Builder plugin for CODAP, which allows users to build some level of structured narratives (Wilkerson et al., 2021).

In our design process, we developed the notebook interface and the sample data story iteratively and side-by-side. In the sample data story, we tell the fictional story of a penguin working with human researchers to find their lost cousin, and for this project we crafted a semi-fictional dataset based on the Palmer Islands Penguin research dataset (Horst et al., 2020). In the prototype interface, this story is organized in chapters and sections as if in a novel or a blog, with interactive data-based puzzles at the end of each section. Solving these puzzles with data blocks will move the narrative forward, thus providing an engaging experience for learners. This structure also allows the story to be divided into multiple web pages, potentially providing an extra level of interactive storytelling in further developments.

Moreover, when building the story itself, we aim to encourage learners to critically examine the original Palmer Penguin dataset. What does it mean to define species and sex? What are the stakes of scientific data gathering? What knowledge is normalized, flattened, or entirely left out of the process? Because the story is told from the first-person perspective of a penguin, we are able to use their lines to ask such questions against human characters, confronting and challenging learners to rethink the implications of the original dataset.

Demo

For the virtual demo, we will use the video-conferencing system used by the conference to answer questions and interact with participants. We will include a 1–2-page PDF virtual "handout" for participants of our demo and a link to the PDF will be shared through the chat feature of the video-conferencing system. The handout will include web URLs to access the working system, as well as a brief description of the project and our contact information. We recommend using the Firefox or Chrome browsers to access the system. We intend to demonstrate two aspects of the system, the code editor, and the stories interface. For the code editor, we will provide a few sample projects that are intended to highlight the capabilities of the data-focused part of the programming system (e.g., visualizations, data aggregation/grouping, data filtering). Participants will be also free to create their own projects or modify existing ones. For the notebook interface, we will invite participants to engage with the Penguins data

story and complete the prompts. We hope that these interactions will catalyze conversations between us and the participants in our demo.

Future Developments

We are currently planning young learner workshops to evaluate *Dataland*, which will provide our project with more on-the-ground empirical inputs, and we hope to use such inputs towards fully fleshing out *Dataland* as a tool for informed, situated, and critical data literacy.

References

The Concord Consortium. (2014). Common Online Data Analysis Platform (CODAP).

Dasgupta, S., & Hill, B. M. (2017). Scratch Community Blocks: Supporting Children as Data Scientists. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 3620–3631. https://doi.org/10.1145/3025453.3025847

Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. Big Data & Society, 5(2), 2053951718784083. https://doi.org/10.1177/2053951718784083

Feinberg, M. (2017). A Design Perspective on Data. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2952–2963. https://doi.org/10.1145/3025453.3025837

Horst, A. M., Hill, A. P., & Gorman, K. B. (2020). palmerpenguins: Palmer Archipelago (Antarctica) penguin data. https://doi.org/10.5281/zenodo.3960218

Kafai, Y. B., & Proctor, C. (2021). A Revaluation of Computational Thinking in K–12 Education: Moving Toward Computational Literacies. Educational Researcher, 0013189X211057904. https://doi.org/10.3102/0013189X211057904

Kelly, A., Chang, C., Hill, C., West, M., Yoder, M., Polman, J., Kane, S., Eisenberg, M., & Shapiro, R. B. (2020). "Our Dog Probably Thinks Christmas is Really Boring": Re-mediating Science Education for Feminist-Inspired Inquiry. https://doi.org/10.22318/icls2020.935

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Jupyter development team. (2016). Jupyter Notebooks—A publishing format for reproducible computational workflows. In F. Loizides & B. Scmidt (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87–90). IOS Press. https://eprints.soton.ac.uk/403913/

Lee, V. R. (2013). The Quantified Self (QS) Movement and Some Emerging Opportunities for the Educational Technology Field. Educational Technology, 53(6), 39–42.

Lee, V., & Wilkerson, M. (2018). Data Use by Middle and Secondary Students in the Digital Age: A Status Report and Future Prospects. Instructional Technology and Learning Sciences Faculty Publications, 1–43.

Papert, S. (1980). Mindstorms: Children, Computers, and Powerful Ideas. Basic Books.

Park, J., & He, C. (2018). SQL Murder Mystery. Northwestern University Knight Lab. https://github.com/NUKnightLab/sql-mysteries

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Millner, A., Rosenbaum, E., Silver, J., Silverman, B., & Kafai, Y. (2009). Scratch: Programming for all. Communications of the ACM, 52(11), 60–67. https://doi.org/10.1145/1592761.1592779

Resnick, M., & Silverman, B. (2005). Some Reflections on Designing Construction Kits for Kids. Proceedings of the 2005 Conference on Interaction Design and Children, 117–122. https://doi.org/10.1145/1109540.1109556

Satyanarayan, A., Moritz, D., Wongsuphasawat, K., & Heer, J. (2017). Vega-Lite: A Grammar of Interactive Graphics. IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis). http://idl.cs.washington.edu/papers/vega-lite

Turkle, S., & Papert, S. (1990). Epistemological Pluralism: Styles and Voices within the Computer Culture. Signs, 16(1), 128–157.

Wilkerson, M., Finzer, W., Erickson, T., & Hernandez, D. (2021). Reflective Data Storytelling for Youth: The CODAP Story Builder. Interaction Design and Children, 503–507. https://doi.org/10.1145/3459990.3465177

Wilkinson, L. (2005). The Grammar of Graphics. Springer New York.

Acknowledgments

Financial support for this work came from the US National Science Foundation (award #1948113), as well as the Graduate School and the School of Information and Library Science at UNC Chapel Hill.