Remixing as a key practice for coding and data storytelling

6 authors, including:

Rabia Yalcinkaya
North Carolina State University
3 PUBLICATIONS 9 CITATIONS

Hamid Sanei
North Carolina State University
6 PUBLICATIONS 2 CITATIONS

CHANGZHAO WANG
University of Southern California
13 PUBLICATIONS 70 CITATIONS

Li Zhu
University of Iowa
1 PUBLICATION 1 CITATION

International Society of the Learning Sciences

Remixing as a Key Practice for Coding and Data Storytelling

Rabia Yalcinkaya, North Carolina State University, ryalcin@ncsu.edu
Hamid Sanei, North Carolina State University, hrsanei@ncsu.edu
Changzhao Wang, University of Miami, cxw662@miami.edu
Li Zhu, University of Miami, lxz567@miami.edu
Jennifer Kahn, University of Miami, jkahnthorne@miami.edu
Shiyan Jiang, North Carolina State University, sjiang24@ncsu.edu

Abstract: In this study, we investigated high school youth's computational data literacy practices, focusing on remixing in building interactive data visualizations for telling stories about climate change. Our interaction analysis revealed remixing processes demonstrating how students navigated the code, data, visualization, and story. We illustrate and discuss one of these processes, wayfinding the code, in which youth learned the code structure and the relationship between the code and data visualization through collaboratively locating codes for remixing. These findings shed light on learning designs for fostering data wrangling, building dynamic data visualizations, and data storytelling about socioscientific issues.

Introduction

In the era of the data revolution and datafication (Pangrazio & Sefton-Green, 2020), opportunities to learn how to wrangle data, build dynamic data visualizations, and tell compelling stories about important socioscientific issues are rare for youth. We report on an initial analysis of data collected from a design-based research project for high school youth in a virtual (synchronous) summer program in which we guided students to program to build a dynamic data visualization and develop accompanying data stories about climate change. We examine remixing in coding environments as a key *computational data literacy* practice. Computational data literacy encompasses both computing practices for developing data visualizations as well as modeling and storytelling with data through public media formats. Looking at student cases, we identify key aspects of students' remixing activity that describe how they attended to the code, data, visualization, and story.

Constructionist perspective and remixing in coding for data storytelling

Our study follows the perspective of constructionism (Papert, 1980). Viewing learning as building knowledge structures, constructionists believe that the most effective learning happens when the learner participates in the process of making a product in a social context (Papert & Harel, 1991). In our study, coding and data storytelling provided such a learning context. Students created a data visualization and a story, here referring to a short narrative text that describes a trend in the data visualization (Segel & Heer, 2010), through the remixing practice of modifying existing programming codes. In our interactive learning environment, students could discuss their problems with peers and instructors when necessary. In the process, they developed knowledge about coding, data visualization, storytelling, and the broader socioscientific background of the data.

Coding is considered as an essential computational thinking skill for all learners, including K-12 students (Wing, 2006). Expanding applications of coding aim to foster learners' critical thinking, analyzing, problem-solving, and data literacy skills (Kalelioglu, 2015). Researchers have tried teaching these skills through various approaches (e.g., through story composition; Kafai & Burke, 2014), coding platforms (e.g., Scratch), and programming languages (e.g., JavaScript, Python, R; Kafai & Peppler, 2011).

In particular, remixing has emerged as a promising approach to developing coding knowledge, computational thinking, and literacy skills (Dasgupta et al., 2016). Remixing is a familiar practice in various digital fields, including literacy, media, arts, and computer programming (Jocson, 2013). In computer science, remixing involves repurposing or changing the code to create new products from previously created work (Dasgupta et al., 2016). Remixing supports learning to program as well as the development of interactive and digital storytelling skills by promoting users' narrative expressions, particularly within collaborative environments like Scratch (Fields et al., 2014). However, less is known about remixing for visualizing and storytelling with socioscientific data, and research on data storytelling for youth has generally not incorporated coding as part of the learning design (e.g., Kahn, 2020; Wilkerson & Laina, 2018). To contribute to an understanding of remixing for impactful data storytelling, we investigated the following research question*: How do students engage in remixing as a computational data literacy practice to build dynamic data visualizations and tell stories about important socioscientific issues?*

Methods

Data VIZion was a free 5-day (4 hours/day) virtual (synchronous, over Zoom) summer camp designed to introduce diverse high school youth to programming to tell stories with data visualizations about climate change. Five teens (2 male, 3 female, ages 14-17, identifying as White, Asian, Latinx, and African-American) participated from high schools across two states in the Southeastern US, recruited through the research team's networks. We screen-recorded all sessions. We also conducted a pre-post survey and individual exit interviews.

We guided students to remix code to build a data visualization and a story with an instructor-created climate change dataset in ObservableHQ, a web-based open data visualization platform that uses the D3.js (Data-Driven Documents JavaScript) library as its main programming language. ObservableHQ is popular with data representation designers and data journalists and is friendly to novice programmers. Since it is a framework of Scalable Vector Graphics wrappers, users can flexibly visualize data in traditional (e.g., bar charts), non-traditional (e.g., beeswarm charts), and interactive (e.g., tooltips, dropdown list) forms and can take existing coded projects and modify them based on their needs. Our dataset included census tract level CO_2 emissions based on car traffic estimates from the Database of Road Transportation Emissions (Gately et al., 2019) for 2000, 2010, 2017 and select indicators from ESRI's US census database (median household income, population density, a diversity index, population counts below poverty level, with a bachelor degree, taking public transportation, taking other transportation types, average commute time) for the counties where students lived. Students chose variables to visualize, investigated conjectures about relationships between variables, and composed a story about emissions where they live for their final projects.

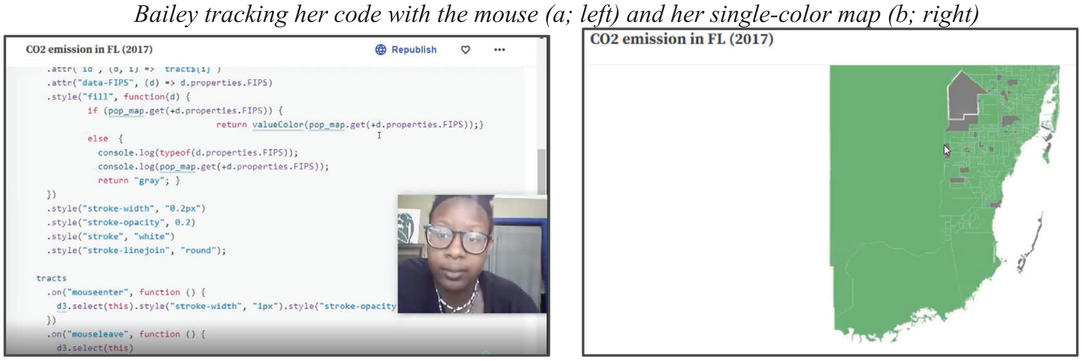
In order to build case studies for each of the five students, we developed content logs of each student's screen recordings, noting episodes around the development of computational data literacy practices. For the current analysis, we looked for moments in their screen capture records in which they engaged in the remixing activities, including their efforts to explore and understand code from existing data visualizations and expand on or customize the code to create their data visualizations. We then micro-analyzed selected remixing episodes using interaction analysis methods (Jordan & Henderson, 1995), focusing on multimodal interactions among participants and instructors and engagement with the ObservableHQ platform. We paid specific attention to references in talk and physical attention (gaze when visible, mouse clicks) to the code, the visualization, and story text within ObservableHQ. e developed conceptual categories to describe students' remixing activities, including ways they modified code, renamed datasets, and variables, and made changes such as color, scale, and range. We present one of those activities, wayfinding, with an illustration from one student case, Bailey (pseudonym), drawn from program Day 3. Bailey's case is an example of how remixing challenged students and how collaboration among participants and with researchers facilitated resolutions of those challenges.

Findings: Wayfinding the code

Wayfinding refers to students locating specific lines of the code to edit and change the data visualization (for other uses of this term, see Hall, 1999 and other studies of spatial problem-solving). Often students solicited instructor help to locate a code line in the code block, such as when they encountered an error or when they needed to find the targeted piece of code to remix it. When instructors stepped in to assist, students typically shared their screens, and all students in the Zoom room could follow their activity. In one case below, Bailey and Raney engaged in joint wayfinding the code (Figure 1a) when working on a CO_2 emissions map (Figure 1b).

Figure 1

Bailey tracking her code with the mouse (a; left) and her single-color map (b; right)



Bailey, who was developing a data story about transportation type (public or private) contributing to CO_2 emissions, had trouble wayfinding the code to replace the "COE_2017" variable (CO_2 emissions in 2017) with the "took_public_trans" variable (population commuting by public transportation). Transcript conventions used: *(Observer notes, significant gesture)*; [overlapping speech; = latched talk; … a pause longer than two seconds.

> **1 Bailey:** *(shares her screen and scrolls through the data array where values are listed below the code)* So...I am in the data section and for my data story it is supposed to be like emissions from public transportation versus, like private or other forms of transportation. Uhh… and I see like the data for public transportation and then for like other forms of transportation *(Bailey points out the variables in the dataset with her mouse)* so...like to...that's my, that's gonna be my variables. So...Do I like get rid of bachelor_degree, hh_below? *(softens her voice)*.
>
> **2 Researcher:** Uh, so you don't get rid of it. Instead, you call it. For example you have a box, a box with many variables and you want to pick the one you want. For example, I picked the COE_2017. You only change the COE_2017 and pick a different one. So, you are not changing the dataset, you are calling different parts of the dataset to use *(Bailey leans into her screen, then scrolls to her code block).*

Bailey thought she should delete variables (bachelor degree, household income) irrelevant to her story from the uploaded data, as displayed in the data array. In Turn 2, the researcher said to keep the data as is and instead use the code to call the name of desired variables and pull specific data from the uploaded dataset. Following this guidance, Bailey found the specific code-line, renamed the variable name to "Public Transportation," and ran the code, which produced a single-color map instead of a choropleth map (see Figure 1b). She expected that the map would show different colors (and different rates of public transit use) for different parts of her county. The researcher tried to troubleshoot when another student (Raney) interjected. [Some turns of talk omitted.]

> **9 Raney:** *(Bailey scrolling in the code)* Actually I tried to change the color green but I did not, I found something like in here, showing me, umm…*valueColor*, if I think so, and I saw, yeah, *valueColor* but in *valueColor* it did not show that, it was like, uh...it was like *pop_map.get*, [and then in parenthesis]
>
> **10 Bailey:** (*nods; tracks code with her mouse, stops on pop_map.get*) [Oh, yeah]
>
> **11 Raney:** It is like *+d.properties.FIPS* but it doesn't show exactly the color of it. And when I look at the *fill* it is like *function(d)* but it doesn't say like it is green. I searched the code block very well, I didn't find the green exactly. I don't know why though.

Raney said she got a single-color map too and tried to fix it (Turn 9) by changing the code (*valueColor* in Turn 9), but this did not produce the desired visualization (multiple shades of green; Turn 11). The researcher then shared the reason for the single-color and the solution (transcript not shown): the code template had a data range that was too small to capture the variation. Students needed to change the domain range (values for "Public Transportation") of the scale function so that light green represented smaller values and darker green represented bigger values. This shared trouble led students to dig deeper in the code block (e.g., Raney reading the code to identify where to change color) while applying recently learned coding knowledge (e.g., Raney remembering that "fill" is a function for changing color). In Bailey's case, following this exchange with Raney and additional step-by-step guidance from the researcher, Bailey changed her range (her minimum value to 1) and observed some color variation on her map when she ran the code. Students' collaborative wayfinding indicated their developing understanding of the code structure, coding language, and how changing code (remixing) impacts visualizations and consequent data stories. Furthermore, students voluntarily (without prompting) shared their experiences to help each other through wayfinding and resolve encountered errors.

Discussion and conclusion

Remixing with script-based programming (in this study, D3.js) was both rewarding and challenging for students. The processes of students' wayfinding the code showed that they learned the code structure while identifying specific lines to remix for building interactive data visualizations for their stories, though they did not write the whole code. Our microanalysis indicates that remixing holds promise for fostering computational data literacy. In terms of challenges, we should be prepared to support remixing processes as students navigate the code, data (dataset), data visualization (representation), and stories. In our study, students were confused about transformations of the data frame (e.g., calling a specific variable in the data frame in Bailey's case). Therefore, we suggest a fruitful future direction for the field is to develop tools that could help students see immediate transformations, such as highlighting changes to a data frame while students are writing codes.

Additionally, as challenging mathematical concepts (Li et al., 2010), the concepts of domain and range are critical for understanding scale functions in coding data visualizations, which are used to transform data values into visual variables such as color (e.g., a light color representing a smaller value). Understanding these mathematical concepts can better support students in resolving trouble when wayfinding the code (i.e., when reasoning about the single-color map in Bailey's and Raney's cases). This echoes the call of integrating math education to foster computational data literacy (Khan & Mason, 2021; Vahey et al., 2012).

In conclusion, this study contributes to our understanding of students' learning computational data literacy practices. Our description of a remixing process, wayfinding the code, illustrates how students navigated and managed various representational media (i.e., code, data, visualization, and story). While students encountered challenges in remixing, overall, students readily engaged in (and enjoyed, as reported in final interviews and surveys) the exploration of coding, data visualization, and story simultaneously. Future analyses will investigate how students drew on their backgrounds and social identities to reason about the impact of human activities on climate when creating stories using a dataset containing socioscientific variables. We call for more work in designing and studying learning environments for cultivating computational data literacy.

References

Dasgupta, S., Hale, W., Monroy-Hernández, A., & Hill, B. M. (2016). Remixing as a pathway to computational thinking. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. https://doi.org/10.1145/2818048.2819984

Fields, D. A., Kafai, Y. B., Strommer, A., Wolf, E., & Seiner, B. (2014, August). *Interactive storytelling for promoting creative expression in media and coding in youth online collaboratives in Scratch*. Presented at the conference of Constructionism 2014: Constructionism and Creativity, Vienna, Austria.

Gately, C., Hutyra, L. R., & Wing, I. S. (2019). DARTE Annual On-road CO2 Emissions on a 1-km Grid, Conterminous USA, V2, 1980-2017. *ORNL DAAC*.

Hall, R. (1999). The organization and development of discursive practices for "having a theory". D*iscourse Processes*, 27(2), 187–218.

Jocson, K. M. (2013). Remix revisited: Critical solidarity in youth media arts. *E-Learning and Digital Media*, *10*(1), 68–82. https://doi.org/10.2304/elea.2013.10.1.68

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences, 4*(1), 39–103.

Kahn, J. (2020). Learning at the intersection of self and society: The family geobiography as a context for data science education. *Journal of the Learning Sciences*, 29(1), 57-80.

Kafai, Y. B., & Peppler, K. A. (2011). Youth, technology, and DIY. *Review of Research in Education*, *35*(1), 89–119.

Khan, K., & Mason, J. (2021). The M in STEM and issues of data literacy. In M. M. T. Rodrigo, S. Iyer, A. Mitrovic, H. N. H. Cheng, D. Kohen-Vacs, C. Matuk, A. Palalas, R. Rajenran, K. Seta, & J. Wang (Eds.), 29th International Conference on Computers in Education Conference, ICCE 2021 - Proceedings (Vol. 1, pp. 632-637). (29th International Conference on Computers in Education Conference, ICCE 2021 - Proceedings; Vol. 1). Asia-Pacific Society for Computers in Education.

Kalelioğlu, F. (2015). A new way of teaching programming skills to K-12 students: Code.org. *Computers in Human Behavior*, *52*, 200–210. https://doi.org/10.1016/j.chb.2015.05.047

Li, K., Uvah, J., Amin, R., & Okafor, A. (2010). A study of college readiness for college algebra. *Journal of Mathematical Sciences & Mathematics Education*, *5*(1), 52-66.

Pangrazio, L., & Sefton-Green, J. (2020). The social utility of 'data literacy.' *Learning, Media and Technology*, *45*(2), 208–220.

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

Papert, S., & Harel, I. (1991). Constructionism. Norwood, NJ: Ablex Publishing.

Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148.

Vahey, P., Rafanan, K., Patton, C., Swan, K., van 't Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics, 81,* 179–205 (2012). https://doi.org/10.1007/s10649-012-9392-z

Wilkerson, M. H., & Laina, V. (2018). Middle school students' reasoning about data and context through storytelling with repurposed local data. *ZDM*, *50*(7), 1223–1235.

Wing, J. M. (2006). Computational Thinking. *Communications of the ACM, 49*, 33–35.

Acknowledgments

We thank Dr. Alberto Cairo for his support and inspiration for this project.