

# Reflective Data Storytelling for Youth: The CODAP Story Builder

MICHELLE HODA WILKERSON, University of California, Berkeley, USA

WILLIAM FINZER, The Concord Consortium, USA

TIM ERICKSON, eeps media, USA

DAMARIS HERNANDEZ, University of California, Berkeley, USA

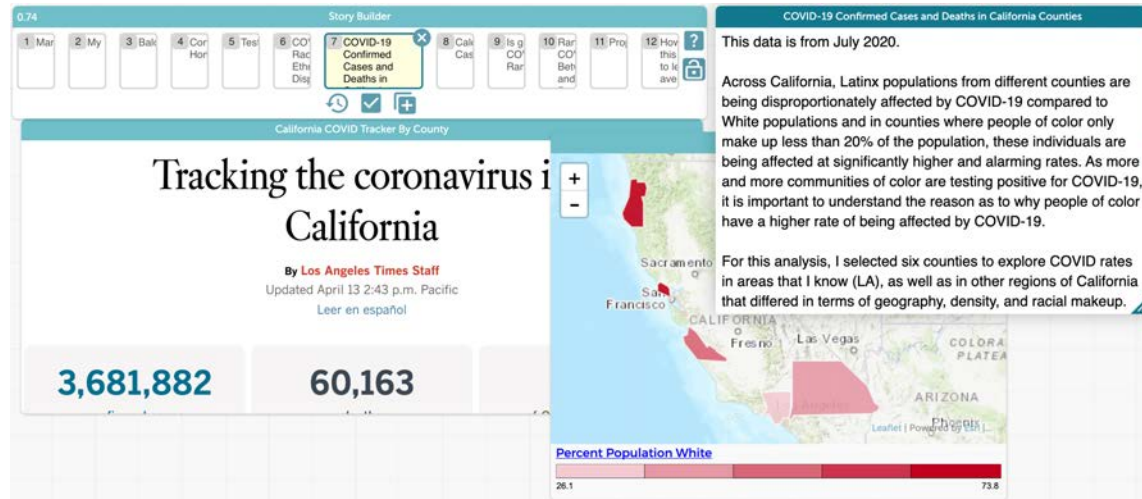


Fig. 1. Story Builder allows students to organize scatter plots, data maps, web pages, images, text, and other media into narratives that describe how their analysis and findings unfold over "moments" in time.

We describe the design of Story Builder, a tool to support adolescents in building interactive multimedia stories that integrate data analysis and visualization with text, images, and other multimodal resources. Story Builder is a plug in for the Common Online Data Analysis Platform (CODAP), a free, online, open-source drag-and-drop interactive data analysis system. Based on early prototyping conducted with several 7th grade classrooms, we sought to design a tool that would allow students to (a) integrate their data investigations with relevant personal and contextual information; (b) record the step-by-step process and rationale of their data analysis; and (c) reflect on and share these contextual and process-oriented elements of data analysis work with others. A major goal of Story Builder is to encourage students to consider and incorporate personal, social, and scientific considerations during data analysis.

CCS Concepts: • **Applied computing** → **Interactive learning environments**; *Education*; • **Human-centered computing** → Visualization systems and tools.

Additional Key Words and Phrases: data visualization, data storytelling, data science education, data analysis, middle school

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2021 Copyright held by the owner/author(s).

Manuscript submitted to ACM

**ACM Reference Format:**

Michelle Hoda Wilkerson, William Finzer, Tim Erickson, and Damaris Hernandez. 2021. Reflective Data Storytelling for Youth: The CODAP Story Builder. In *Interaction Design and Children (IDC '21), June 24–30, 2021, Athens, Greece*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3459990.3465177>

**1 INTRODUCTION**

There is growing interest in integrating data literacy and data science throughout the K-12 curriculum [6, 8, 15]. One practice that is increasingly common in the professional world, and is generating enthusiasm from educators, is *data storytelling* [19]. In data storytelling, the presentation of data and statistics are coupled with other visual and narrative elements to explain what is done to data during the process of analysis, why such analysis is important, and to communicate to a broad audience the results and implications of said analysis.

The *Writing Data Stories* project (NSF IIS-1900606) seeks to introduce students to data storytelling by encouraging them to articulate their personal and social connections to data as they simultaneously engage with the statistical patterns contained therein. In doing so, our project recognizes that a great deal of conceptual and personal work goes into students' initial preparation and exploration of data. For example, students may seek a grounding in the data by filtering and searching a dataset to find specific records that represent themselves or other familiar information [18, 21]. Or, they might want to reduce the complexity of their initial investigation by focusing on only a few key attributes of a large dataset, expanding their investigation later [7]. These decisions are not merely about "preparing" or "cleaning" data, but also represent some of the ways that students build relationships to data and leverage their background knowledge to drive exploration [12, 17, 22]. However, it is difficult to capture these process-oriented aspects of data work in most spreadsheet or exploratory visual data tools. The Story Builder tool allows students to capture and reflect on snapshots or "moments" of their process as they transform, analyze, visualize, interpret, and communicate the results of large-scale scientific datasets. We have been developing and testing the Story Builder tool to support student investigations related to socioscientific topics including health, climate, and local ecosystems.

**2 THEORY AND RELATED WORK**

The *Writing Data Stories* project, and the design of Story Builder specifically, are inspired by literature in both science and statistics education that emphasize the need to introduce students to data as evidentiary texts. The science education literature, for instance, has demonstrated that it is important for students to understand how data can be made useful for the needs that arise in their own course of investigation, rather than for a predetermined goal [16]. While much research has explored how this can be done by having students collect their own data, less has explored how students might approach large, pre-existing datasets as resources that they can re-construct and re-represent for purposes of analysis. However, working with such large existing datasets is central to science practice [5]. When working with externally-sourced data, practitioners must decide whether those data are relevant to their questions, and what types of changes, visualizations, or reorganizations are required to make them useful [13].

This transformation and visualization of datasets through can be thought of as the reading and writing of data in new ways that communicate particular investigations and findings. This treatment of data is consistent with research that has explored the use of provided datasets in statistics education, as a means to develop learners' capacities to engage in what Tukey called exploratory data analysis [20]. Most such work, suggests it is important to emphasize provided data not as a static resource but rather something to learn from [3] and use in argument [2]. It is also consistent with what Duschl refers to as transitional practices of science, encoded and communicated through "data texts" that describe how

evidence is turned into explanations (including through data manipulation; p. 208) [4]. Importantly, this involves having students explicitly reflect on their work with data including how data were generated, their processes of analysis, their predictions, and their interpretations [14].

Importantly, Story Builder is intended to support a more integrative and intentional weaving of students' personal contexts with scientific and mathematical study. This approach is consistent with the constructionist tradition within the child-computer interaction literature, which highlights the importance of perspective and meaning-making in learning with digital tools [1, 11]. We further complement this interest in the *personal* with an interest in the *social*, particularly the historical and political contexts in which socio-scientific issues unfold. Drawing from interdisciplinary theories of literacy development, particularly among marginalized youth [10], one goal of integrating the personal, social, and scientific is to expose and make available for analysis the tensions between formal and informal discourses as a means to interrogate power.

### 3 THE DESIGN OF STORY BUILDER

Story Builder seeks to support students in approaching data as a *text* and a *source of argumentative evidence* by providing an infrastructure for students to (a) integrate their data investigations with relevant motivational and contextual information; (b) record the step-by-step process and rationale of their data analysis; and (c) reflect on and share these contextual and process-oriented elements of data analysis work with others. It is a plug-in for the Common Online Data Analysis Platform (CODAP; accessible at [codap.concord.org](http://codap.concord.org)), a free, online, drag-and-drop data analysis and visualization tool designed for novice data analysts from early adolescence into their early college years. We begin this section by providing a brief description of the CODAP platform, then describing the specific features of the Story Builder tool; a more complete and detailed description of CODAP's features can be found in [9].

#### 3.1 The Common Online Data Analysis Platform (CODAP)

CODAP allows users to load one or more datasets by dragging and dropping any CSV data file onto the screen to create a "CODAP Document". A dataset can be viewed as a table, or or as a stack of "case cards" that describe each record as if it were written on an individual index card (Figure 2).

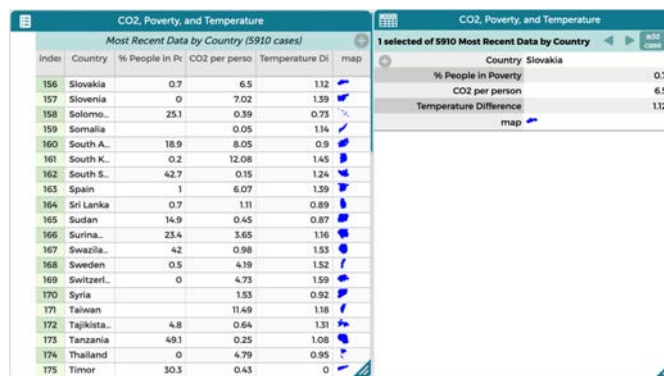


Fig. 2. An example of the table and case card views of data records in the standard CODAP view.

Once a dataset is loaded, young analysts can add one or more scatterplots or maps to the document to visualize and explore their data. When a scatterplot is first added, it displays dots (each representing a record in the dataset)

scattered randomly in the graph space. They can organize the plot into a visualization of their data by dragging and dropping attributes (column names) from the dataset onto the axis they would like to appear. Scatterplots can also include statistical overlays that provide visual indicators of, for instance, the mean, median, or line of best fit for a given set of data, depending on how those data are visualized. If a dataset includes geographic information, students can also create maps and drag-and-drop attributes to colorize the map accordingly. When one or more records in a dataset are selected, the corresponding visual representations of that record are also highlighted to allow for analysis across visualizations (Figure 3).



Fig. 3. An example of the standard CODAP interface, without Story Builder. Records corresponding to the countries of Oceania are highlighted in the table, map, and graph through "hotlinking."

CODAP also allows students to transform their data using any number of well-known data transformation actions such as calculating new variables, grouping data hierarchically, filtering or sorting their dataset, and more [7]. As they work, students can save their "document" as a particular arrangement of windows including the dataset, graphs, maps, and other elements. This functionality is similar to an Excel spreadsheet where the final state of the spreadsheet can be saved and re-opened, but all elements of the document remain interactive and can be further manipulated.

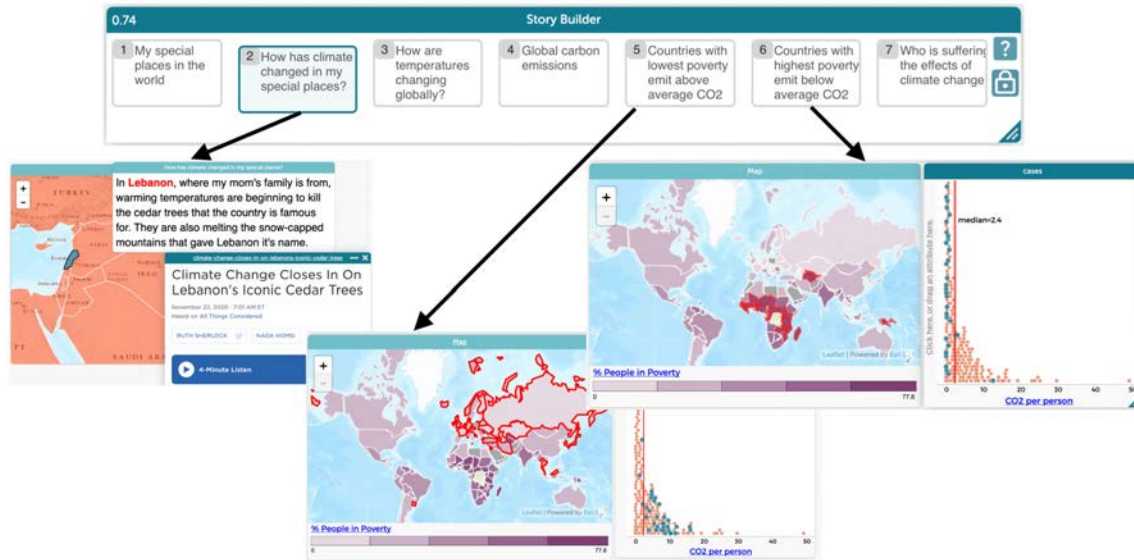
### 3.2 Story Elements

Earlier versions of CODAP without Story Builder allowed for a limited number of other multimodal elements to be added to a document including web pages, images, and text. As we developed Story Builder and consulted with learners, designers, and educators, we have expanded the type and richness of elements that can be added to better support fully multi-modal storytelling. This includes more robust support for adding and resizing images, adding a rich text editor, and introducing the capability to add video and audio recordings as CODAP elements. Each story element appears in a window that can be moved, resized, and even stacked over other windows on the screen.

### 3.3 The Moments Bar

The core innovation behind Story Builder is that it introduces a temporal dimension to a CODAP document so that students' work with data can be captured, edited, and retold as a series of "moments." Moments can be thought of as snapshots of a CODAP document as students move through their analysis process. As students conduct data investigations, their moments may capture and allow them to reflect upon, explain, and share actions such as transforming data; adding, deleting, or moving elements on the screen; creating and revising graphs and maps to discover new relationships; and developing media and text explanations of their motivations, decisions, and interpretations. They

209 preserve the identity of particular windows over time, so that as windows are moved around on the screen, made  
 210 smaller or larger, or moved to the foreground or background, these actions are animated to document the flow and  
 211 continuity of analysis. Students can put moments into a preferred order using a drag and drop "Moments Bar" in order  
 212 to construct a narrative that brings together these motivations, decisions, findings, and interpretations (Figure 4).  
 213



236 Fig. 4. An example of a CODAP Document with the Story Builder plugin activated. The moments bar (top) allows students and their  
 237 audience to move through a data story temporally, while still being able to interact with the visualizations, maps, and media.  
 238

239 Because moments capture a variety of changes ranging from adjusting the location of elements to significant  
 240 transformations of data, there is a great deal of flexibility in how a Data Story unfolds. Once students are satisfied with  
 241 their data story, they can "lock" the moments bar and make their CODAP document available to others as a standard  
 242 CODAP file. Locking a Data Story allows viewers of the document to click through and interact with each moment,  
 243 without editing or overwriting any of the saved work done by the original author. All of the data, visualizations, and  
 244 other media in Story Builder are still interactive after they are captured in a moment, even when the story is locked.  
 245 This allows others who are reading a peer's data story to engage first hand with the data and claims contained therein.  
 246

247  
 248  
 249 **4 ONGOING EVALUATION AND EXAMPLES**

250 The Story Builder design is based on preliminary research we conducted with 7th grade students in several science  
 251 classes in a public school that served high proportions of Latinx and emerging multilingual students in a working class  
 252 suburb of the U. S. state of California during Fall 2019 and Spring 2020.  
 253

254 During that study, we piloted two units in collaboration with the classroom teacher. Importantly, the very foci of  
 255 these units, and the datasets selected for each, were identified by the teacher who had developed her instruction for  
 256 the year around the themes of diabetes and climate change. We collected student written and digital artifacts from all  
 257 consented students, and video recorded instruction from the back of the classroom. We also collected additional audio  
 258  
 259

261 and/or video recordings of daily activities from 2-4 focal groups per classroom (a total of 11 student groups). We also  
262 collected screen recordings of all focal group interactions with the CODAP tool.

263 In the first unit, students explored a dataset consisting of commercial cereals and interrogated the extent to which  
264 the dataset (and common narratives and advertisements about breakfast in the United States) did or did not reflect their  
265 actual breakfast concerns and choices. The datasets included information from over 70 popular cereals' nutrition labels,  
266 and students also analyzed commercials and other marketing materials for several of the featured cereals. Part of the  
267 motivation for this focus was to allow students to bring in knowledge from prior lessons about the impact of the sugar  
268 lobby, economic access, and marketing techniques (such as the manipulation of serving size to make cereals appear  
269 healthier) on consumers' food choices. Students then added new records to the dataset to represent their own home  
270 breakfasts, as well as variables in the dataset that mattered to them (such as taste or price) before analyzing the data to  
271 justify their selection of a healthy and accessible breakfast choice.

272 In the second unit, students explored patterns of climate change. They began this investigation by identifying  
273 personal "special places" anywhere in the world, and exploring how those places may have already been impacted  
274 by climate and extreme weather events through news stories and web search. Examples of students' selected "special  
275 places" included parents' and families' countries of origin, homes of family members, favorite vacation spots, or places  
276 students would like to visit some day. The unit was disrupted by COVID-19 shelter-in-place orders, however, the  
277 intention was that students would then locate their special places within a global dataset, and begin to compare carbon  
278 emissions and climate indicators among the countries where those special places are located within their own student  
279 groups. Through this personal investigation, students would then be supported in eventually expanding to global  
280 comparisons in order to evaluate the inequitable distribution of who is most responsible (via emission of carbon and  
281 other greenhouse gases) and most vulnerable (due to temperature anomalies and poverty) to climate change.

#### 282 4.1 Initial Prototyping of Story Building Activities

283 Our current version of Story Builder is informed by early data storytelling activities on paper, in digital Google  
284 Documents, and using the standard version of CODAP. These activities included interviews with peers about the  
285 data analysis process, open-ended stories about special places and other connections to datasets and the issues they  
286 represent, and daily observations of students as they conducted analyses using CODAP and worked to justify and  
287 record their process. Together, these observations identified what supports would be necessary for students to develop  
288 narrative accounts of their data explorations. In particular, we recognized the need for the following features, which  
289 have been incorporated into the current version of the Story Builder plugin:

- 290 • Rich text that encourages and supported student explanations for conclusions drawn from data;
- 291 • Multi-media elements that encourage personalization of data stories;
- 292 • "Replay-ability" so that students could refine and continue to interact with data and patterns identified during  
293 particular moments during their analysis process.

#### 304 4.2 Model Data Stories

305 We are developing a suite of example Data Stories using the current version of Story Builder to be used as models for  
306 youth. We are seeking to support a diversity of story types (in terms of modality, scope of analysis, and topic) that all  
307 illustrate the blending of personal experience/perspective; social issues; and principled analysis and presentation of  
308 patterns data. For example, the story featured in Figure 1 documents the impact of the COVID-19 pandemic on coauthor  
309

and undergraduate research assistant Damaris Hernandez's family. It uses these experiences to motivate an exploration of the disproportionate impact of COVID on Latinx families in California.

Other example data stories unpack the close analysis of only one dataset, with a focus on process. For instance, the data story featured in Figure 4 begins with an in-depth exploration of the impact of warming temperatures on the country of Lebanon. It then uses different data story moments to compare CO<sup>2</sup> emissions of people in countries with low poverty rates versus those in countries with high poverty rates using the same graphs highlighted differently.

### 4.3 Ongoing User Testing

We are currently beginning to test Story Builder with adolescent youth ages 10-14. We also look forward to using this tool during a series of workshops this summer that explore local datasets related to environmental justice issues including pollution, health, industry, race, and class.

## ACKNOWLEDGMENTS

This work is funded by National Science Foundation Grant No. IIS-1900606. We are grateful to all project partners and members, particularly those who have been actively involved in the testing and refinement of the Story Builder tool.

## REFERENCES

- [1] Edith Ackermann. 1996. Perspective-taking and object construction: Two keys to learning. *Constructionism in practice: Designing, thinking, and learning in a digital world* (1996), 25–35.
- [2] Dani Ben-Zvi. 2006. Scaffolding students' informal inference and argumentation. In *Proceedings of the seventh International conference on Teaching Statistics*. 1–6.
- [3] Paul Cobb and Kay McClain. 2004. Principles of instructional design for supporting the development of students' statistical reasoning. In *The challenge of developing statistical literacy, reasoning and thinking*. Springer, 375–395.
- [4] Richard Duschl. 2008. Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of research in education* 32, 1 (2008), 268–291.
- [5] Richard A Duschl and Richard E Grandy. 2008. *Teaching scientific inquiry: Recommendations for research and implementation*. BRILL.
- [6] Joachim Engel. 2017. Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal* 16, 1 (2017), 44–49.
- [7] Tim Erickson, Michelle Wilkerson, William Finzer, and Frieda Reichsman. 2019. Data moves. *Technology Innovations in Statistics Education* 12, 1 (2019).
- [8] William Finzer. 2013. The data science education dilemma. *Technology Innovations in Statistics Education* 7, 2 (2013).
- [9] William Finzer and Dan Damelin. 2015. Building the CODAP community. *@Concord* 19, 2 (2015), 8–9.
- [10] Kris D Gutiérrez and A Susan Jurow. 2016. Social design experiments: Toward equity by design. *Journal of the Learning Sciences* 25, 4 (2016), 565–598.
- [11] Nathan Holbert, Matthew Berland, and Yasmin B Kafai. 2020. *Designing constructionist futures: The art, theory, and practice of learning designs*. MIT Press.
- [12] Jennifer Kahn. 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 (2020), 57–80.
- [13] Sean Kandel, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank Van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono. 2011. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization* 10, 4 (2011), 271–288.
- [14] Clifford Konold and Alexander Pollatsek. 2002. Data analysis as the search for signals in noisy processes. *Journal for research in mathematics education* 33, 4 (2002), 259–289.
- [15] Victor R Lee and Michelle H Wilkerson. 2018. Data use by middle and secondary students in the digital age: A status report and future prospects. (2018).
- [16] Eve Manz. 2016. Examining evidence construction as the transformation of the material world into community knowledge. *Journal of Research in Science Teaching* 53, 7 (2016), 1113–1140.
- [17] Jessica Roberts and Leilah Lyons. 2020. Examining spontaneous perspective taking and fluid self-to-data relationships in informal open-ended data exploration. *Journal of the Learning Sciences* 29, 1 (2020), 32–56.

- 365 [18] Andee Rubin. [n.d.]. Facebook or instagram? Teens explore data about technology use. *Hands On! Magazine* ([n. d.]). "[https://www.terc.edu/facebook-](https://www.terc.edu/facebook-or-instagram-teens-explore-data-about-technology-use/)  
366 [or-instagram-teens-explore-data-about-technology-use/](https://www.terc.edu/facebook-or-instagram-teens-explore-data-about-technology-use/)"
- 367 [19] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6  
368 (2010), 1139–1148.
- 369 [20] John W Tukey et al. 1977. *Exploratory data analysis*. Vol. 2. Reading, Mass.
- 370 [21] Michelle Hoda Wilkerson and Vasiliki Laina. 2018. Middle school students' reasoning about data and context through storytelling with repurposed  
371 local data. *ZDM* 50, 7 (2018), 1223–1235.
- 372 [22] Michelle Hoda Wilkerson and Joseph L Polman. 2020. Situating data science: Exploring how relationships to data shape learning. *Journal of the*  
373 *Learning Sciences* 29, 1 (2020), 1–10.

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416