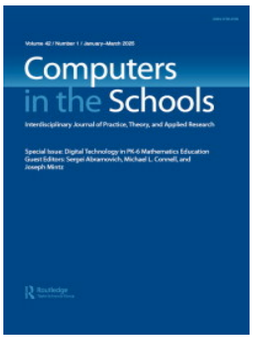

Routledge
Taylor & Francis Group

Computers in the Schools

Interdisciplinary Journal of Practice, Theory, and Applied Research

ISSN: 0738-0569 (Print) 1528-7033 (Online) Journal homepage: www.tandfonline.com/journals/wcis20

Data Moves as a Focusing Lens for Learning to Teach with CODAP

Rick A. Hudson, Gemma F. Mojica, Hollylynne S. Lee & Stephanie Casey

To cite this article: Rick A. Hudson, Gemma F. Mojica, Hollylynne S. Lee & Stephanie Casey (17 Oct 2024): Data Moves as a Focusing Lens for Learning to Teach with CODAP, Computers in the Schools, DOI: 10.1080/07380569.2024.2411705

To link to this article: https://doi.org/10.1080/07380569.2024.2411705

COMPUTERS IN THE SCHOOLS
https://doi.org/10.1080/07380569.2024.2411705

Routledge
Taylor & Francis Group

Data Moves as a Focusing Lens for Learning to Teach with CODAP

Rick A. Hudson[a], Gemma F. Mojica[b], Hollylynne S. Lee[b] and Stephanie Casey[c]

[a]University of Southern Indiana, Evansville, Indiana, USA; [b]NC State University, Raleigh, North Carolina, USA; [c]Eastern Michigan University, Ypsilanti, Michigan, USA

ABSTRACT

Innovative dynamic data tools afford opportunities for K-12 students and teachers to explore multivariate data and create linked data representations. These tools also support engagement in data moves, which are transnumerative actions to process, organize, and visualize data. The current study sought to understand how prospective K-12 mathematics teachers (PMTs) use data moves in the Common Online Data Analysis Platform (CODAP) to create and interpret visualizations and statistical measures to make sense of state-level data about education in the United States. Extending the work of Erickson et al. (2019), a framework is presented to characterize data moves and provide examples of actions within CODAP that illustrate each data move. Based on analysis of thirty screencasts created by PMTs, four examples highlight PMTs' use of data moves to investigate data in CODAP.

KEYWORDS

Data moves; statistics; data science; data visualizations; prospective teachers

Introduction

There has been increased emphasis around the world for providing opportunities to learn statistics and data science concepts within middle and high school mathematics courses (Bargagliotti et al., 2020; Sukol, 2024). Recommendations encourage teachers to provide opportunities for students to learn statistics through actively engaging in real data investigations (Ben-Zvi et al., 2018; Rubin, 2020). Using real, larger datasets requires use of technology tools throughout an investigative process, such as in preparing, collecting, exploring, visualizing, and summarizing data (Gould et al., 2018). Several researchers have shown that learners in secondary schools can successfully use data tools to manage, wrangle, visualize, model, and craft data stories using big datasets (Kahn & Jiang, 2020; Lee & Wilkerson, 2018; Sanei et al., 2023).

CONTACT Rick A. Hudson rhudson@usi.edu Mathematical Sciences Department, University of Southern Indiana, Evansville, IN, USA.

To develop knowledge for teaching statistics to support students' learning with data and technology (Lee & Hollebrands, 2011), teachers need experiences as learners of statistics to investigate large multivariate data (Gould & Çetinkaya-Rundel, 2014; Lee et al., 2014). Our paper expands research in this area to offer a framework and examples of the nuanced ways that future teachers engage with data using a highly visual and dynamic online data tool, CODAP.

Potential of dynamic data tools for changing the teaching of statistics

Wild and Pfannkuch (1999) introduced the idea of transnumeration as a data practice that statisticians engage in to transform data to investigate, summarize and visualize data. Those who work with data in their career use a variety of modern tools to assist them in transnumerative actions necessary to transform data in ways that answer questions and communicate trends and patterns (Lee et al., 2022). Over the past 20 years, there have been a variety of data tools developed and used in educational settings, each with their own interfaces, features, and capabilities. For a recent review of data tools for K-12 data science education see Israel-Fishelson et al. (2023). Although there are multipurpose tools such as spreadsheets and programming languages and environments based on Python and R (both commonly used by data scientists and statisticians), other tools are developed specifically for educational purposes, often referred to as dynamic statistics software or dynamic data tools (e.g., CODAP, TinkerPlots, Fathom), that have low/no code requirements, drag-and-drop interfaces, and possibilities of viewing multiple representations of data that are dynamically linked–meaning that selections and changes in one representation are visibly seen across all representations (Mojica et al., 2019).

Dynamic data tools, such as CODAP, are designed based on research findings of how students learn with data and have the potential to change what teaching and learning statistics and data science can look like in middle and high school settings (Finzer, 2013). These tools support exploratory data analysis with larger datasets, intuitive ways to organize and visualize data, and connections among multiple data representations (Mojica et al., 2019).

CODAP and other dynamic data tools have been incorporated in statistics curricula and teacher education materials (e.g., Lee et al., 2010; 2021; Zieffler & Catalysts for Change, 2023). However, McCulloch et al. (2021) reported that only about half of U.S. mathematics teacher education programs include dynamic statistics software in courses for future teachers – 56% of 213 program respondents. Research has shown there are many benefits for mathematics teachers who have been educated in a technology-intensive teacher education program that includes use of dynamic

statistics software, including the ability to use these tools with large multivariate datasets to engage in statistical investigations themselves (Lee et al., 2014), design statistics investigation lessons (Casey et al., 2021), and look for ways to enhance their statistics lessons using such tools once they are practicing teachers (McCulloch et al., 2018).

Working with data

The data science process involves importing, tidying, transforming, visualizing, modeling and communicating about data (Wickham et al., 2023). Many of the steps in this process correspond to transnumerative actions that are executed using specific actions or commands in different data tools. For example, Wickham (2014) described the specific operations to change the structure of a dataset to result in a "tidy" dataset, where each case is a row and the value for each variable is in a separate column, and actions of modifying tidy datasets, such as filtering, transforming, aggregating, and sorting.

Other scholars have focused on how secondary students can engage in practices of data wrangling using various actions in tools such as R and iNZight with diverse command structures that require different knowledge of the purpose of such commands and specific structure for parameters in a function (Burr et al., 2020). Recognizing the difficulty learners have in understanding different commands in R and Python to act on data frames in order to filter or calculate summaries, Sundin et al. (2020) investigated the use of visual cues to help college-level learners decompose actions on data in order to assist students in understanding the purposes of certain actions on data. Others, such as Jiang and Kahn (2020), have studied how younger learners can engage in data wrangling practices such as filtering a dataset, focusing on a single data point for interpretation (often outliers), and reasoning about data in the aggregate through trends and summaries using highly visual, interactive data interfaces such as Gapminder and Social Explorer. In an earlier study on ways pre-service teachers used TinkerPlots and Fathom, Lee et al. (2014) identified several transnumerative actions that were particularly useful when investigating questions with larger datasets. These included overlaying statistical measures within a graph of a distribution, using existing attributes to compute a new attribute, linking cases across representations, removing cases based on some criteria, and computing proportions for subgroups of data.

Within any data tool, transnumerative actions for processing, organizing and visualizing data are made possible by enacting actions supported by the specific data tool. Erickson et al. (2019) coined different types of transnumerative actions as "data moves" that consist of "an action that

alters a dataset's contents, structure, or values" (p. 3). They proposed six core data moves:

- Filtering – process of removing extraneous cases irrelevant to an investigation or focusing on a subset of data that removes complexity or quantity of cases.
- Grouping – process of dividing a dataset into multiple subsets, or groups, often for the purpose of making comparisons among groups.
- Summarizing – process of producing and displaying a statistical measure.
- Calculating – process of transforming data to produce a new attribute by using formulas to calculate values based on other attributes.
- Merging/Joining – a process used to expand a dataset by combining datasets together, either through merging two datasets about the same phenomena to get a bigger dataset (i.e., more cases) or joining one dataset with another to add new information about each case (combining a dataset of students' heights with another one containing their arm span measurements and gender identity)
- Making a Hierarchy – a specific process to restructure a flat or "tidy" dataset (every case has a unique row and every attribute in a column) into a nested format by forming subsets, often with labels (e.g., states or school names) or categorical attributes such as census region (West, South, Northeast, Midwest).

Erickson et al. (2019) also described how others may consider moves such as Sorting for organizing and visualizing data, Stacking for transforming datasets into tidy rows and columns, and Sampling as a key move to generate new datasets, either from probabilistic models or from selecting cases from a larger dataset. They explicitly anticipated and hoped that "the broader statistics education and data science communities will help us recognize and characterize others [data moves] as appropriate" (p. 15). Indeed, within the context of a study that utilized spreadsheets with secondary students, van Borkulo et al. (2023) noticed that students often used Sorting and considered it a key data move. Our intent is to use the context of our research to dive deeper into the nuanced ways a user could enact specific data moves, and combinations of data moves, to accomplish different goals in their investigation.

Research focus

Although the work by Erickson et al. (2019) created a list of data moves, we believe that teachers and teacher educators may benefit from a framework that helps to connect how specific actions in CODAP correspond

to these data moves. Furthermore, Peters et al. (2024) have suggested that data explorations in teacher education "should be designed for teachers to be computationally nimble and comfortable with data moves using technology such as CODAP" (p. 138). However, the field lacks empirical evidence about how teachers use data moves as they engage in such data investigations. To help fill this gap and call for research, we aimed to examine ways in which pre-service K-12 mathematics teachers (PMTs) utilize data moves made possible through the use of dynamic data tools to help them answer statistical questions. To that end, our specific research question is:

> How do PMTs use data moves in CODAP to create and interpret visualizations and statistical measures to make sense of data?

Methods

Context

As part of a larger study, we collected various sources of data relating to field testing our ESTEEM materials (Lee et al., 2021) from Fall 2017 to Spring 2020, when PMTs used these materials in various courses in their mathematics teacher education programs at different institutions across the U.S. For this paper, we focus on an assignment within these curriculum materials[1] where PMTs created screencasts–digital video recordings of their computer screens along with audio–to document their completion of a statistical investigation in CODAP. The assignment presented three choices of datasets and associated questions. PMTs were asked to verbalize their thinking as they investigated the data and made their recording. PMTs were also directed to illustrate strong statistical thinking (e.g., using an investigative cycle and habits of mind), appropriate statistical language, and advanced and powerful features of CODAP to conduct an in-depth analysis.

One dataset focused on education-related data from each of the 50 U.S. states and the District of Columbia (DC) (hereafter collectively referred to as 'states')[2]. The U.S. State Education dataset consisted of nine attributes, including name, two categorical attributes (census region and census division), and six quantitative attributes (expenditure per student, average teacher salary, number of teachers, number of high school graduates, revenue per student, and student-to-teacher ratio). When accessing the data within CODAP, a PMT initially saw a data table, an empty graph, a

[1] https://place.fi.ncsu.edu/local/catalog/course.php?id=22&ref=3.

[2] https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=18336.

map of the U.S., and a textbox describing the dataset's source. The U.S. State Education dataset was selected as the focus for our analysis because more PMTs selected this choice than any other dataset. PMTs were asked to address the following questions in their data investigation:

- Based on the average teacher salaries for states in the South and the West, where would you prefer to teach and why?
- Explore other quantitative attributes in this dataset and compare the distributions for the South and West. Based on the data you have examined, in which region do states seem to have good teaching conditions in their schools?

The assignment was created using several important design features. First, we developed the task to engage PMTs in a data investigation within a context that is relevant to teachers. Second, the task was designed to utilize CODAP. Third, the task encourages PMTs to compare distributions, which many recognize as a strategy for developing statistical reasoning (Makar & Confrey, 2004). An important feature of the U.S. School Education dataset is that multiple conclusions could be drawn relating to whether the South or the West would be a better region to teach in, depending on what attributes were considered, as well as personal preferences (e.g., location or weather).

Data collection

Sixty screencasts were collected in total. Due to the extensive, time-consuming work of analysis, we randomly selected 30 PMTs' screencasts to analyze for this study. Participants came from two institutions in the U.S., a large university in the Southeast and a mid-sized university in the Midwest. They were enrolled in one of the following courses: (1) teaching secondary mathematics with technology, (2) elementary/middle statistics content, or (3) secondary mathematics methods.

Framing a focus on data moves

Each screencast video was viewed by two members of our research team. After viewing the videos multiple times, researchers recorded detailed descriptions of PMTs' actions in CODAP, as well as their verbalizations. Researchers also documented additional notes about what they were noticing in relation to PMTs' actions in CODAP and their reasoning. Building on the work of Erickson et al. (2019), we used an iterative process to create a framework for identifying data moves used by PMTs and developed definitions of each move (DeCuir-Gunby et al., 2010).

Researchers rewatched screencasts and refined categories of data moves and definitions while also integrating literature on data moves and learners' use of dynamic data tools through this process.Through watching and listening to the nuanced ways PMTs were engaging with data in CODAP, we primarily attended to the reason or purpose for the different data moves they would enact. This led us to have slightly different categories of data moves than had been noted in prior literature. The final iteration of the data moves framework is shown in Table 1. In the following description and throughout this text, we bold and italicize data moves from our framework.

Although Making a Hierarchy was a core data move for Erickson et al. (2019), we saw this as an example of an action for *grouping* that restructured a data table by *using subsets* already in the dataset. We further expanded upon *grouping* to have three types of data moves with different purposes: *using subsets*, *creating subsets*, and *highlighting subsets*. Calculating was another data move identified by Erickson et al. that we reframed as part of a data move we call *expanding datasets* that has three distinct data moves with different purposes: *adding data*, *merging*, and *joining*. One way to add data is through creating a new attribute based on a calculation using existing attributes, which was also noted by Lee et al. (2014). Although Sorting was mentioned by Erickson et al. (2019) and identified as a key data move by van Borkulo et al. (2023), we broadened the idea of Sorting into a data move of *ordering* since that is the purpose of why one might sort data, and CODAP supports ordering cases in a variety of ways. Other new data moves we identified and described in Table 1 will be elaborated through examples in the results and discussion sections.

Data analysis

After the data moves framework was fully developed, two researchers watched all 30 screencasts again to identify the combination of data moves utilized by PMTs. These combinations of data moves lead to the development of major themes that described commonalities among the PMTs' data moves in CODAP, the inferences they made based on the outcomes of these data moves, and the subsequent reasoning about the dataset's context needed to answer the statistical questions of the task. The two researchers discussed and agreed on four prominent themes concerning the analytic approaches and the collection of data moves used by PMTs (Saldaña, 2021). They identified how each of the 30 screencasts supported each theme and took additional researcher notes. Researchers then selected PMTs' screencasts that illustrate the four themes, which describe general approaches of using data moves to

Table 1. Data moves framework.

Data move		Purpose	Examples of actions within CODAP
Grouping	Using Subsets	Actions that reorganize or restructure data using existing categories of a categorical attribute	• Creating a hierarchy in a table that reorganizes data into subsets according to categories in an existing attribute • Separating data in a graph by placing a categorical attribute on an axis • Placing (or removing) a categorical attribute as a legend to a graph or a map (this colors cases by distinct groups) • Placing an attribute to the right axis in a graph to separate data into distinct subsets, each represented on its own axes
	Creating Subsets	Actions that separate data into user-defined subsets	• Grouping into bins when a quantitative attribute is on an axis in a graph • Changing the width or alignment of existing bins to create a different grouping • In a graph, adding a moveable line, adding a value, or adding shaded regions to create subsets of data • Adding a quantitative attribute as a legend attribute to a graph or a map to color cases by a gradient scale which has groupings in it, and necessarily puts the groups in order from least to greatest • Creating a new attribute in a table to group existing cases in categories • Changing a quantitative attribute to treat it as a categorical attribute to create groups
	Highlighting Subsets	Actions to identify and select a group of cases	• Highlighting a collection of cases, through a manual process of selecting cases or dragging a rectangle within a graph or through a table or map. • Highlighting a collection of cases through selecting a predefined element in a graph such as a bar in a bar graph or histogram, the lower whisker in a boxplot, or a segment of a segmented bar graph. • When a categorical attribute is already displayed as a legend to a graph/map or is used to hierarchically organize a table, using the legend (or row in a table) to select a particular categorical group(s) so that members of that group(s) are highlighted in the representation • When a quantitative attribute is already displayed as a legend to a graph/map, selecting a particular portion(s) of the gradient scale so that cases in that portion(s) are highlighted in the representation
Filtering		Actions that reduce a dataset to only include a subset of cases	• Setting aside cases in a table, either individually or as a collection of cases that belong to a certain subset • Hiding cases in a graph (e.g., certain groups, cases, outliers) • Deleting a case, either from a table or a graph • Choosing a random sample from a dataset to obtain a smaller dataset

Data move		Purpose	Examples of actions within CODAP
Ordering		Actions that sort data into a particular order	• Organizing cases by sorting by an attribute in a table (ascending, descending) • Putting a quantitative attribute on an axis to sort (organize) cases from low to high (e.g., dotplot) • Treating a categorical attribute as a numeric attribute to order cases • On a graph axis, dragging categorical attribute names to reorder them on the graph scale (e.g., with Census Region on an axis, moving South and West next to each other) • Creating a scatterplot to order two quantitative attributes and visualize their coordinated location in the plane
Summarizing		Actions that use computation to describe a characteristic of a dataset	• In a hierarchical structure in a table, adding an attribute to compute a measure for groups • Adding visual overlays in a graph of statistical measures (e.g., mean, median, standard deviation, MAD, linear model, squares of residuals) • Adding counts or percents to a grouping of cases in a graph • Showing a boxplot with or without outliers (using a five-number summary)
Linking		Actions to select case(s) in a representation to identify corresponding case(s) in another representation	• Identifying cases in a graph, table, or map by selecting cases in a graph • Identifying cases in a graph, table, or map by selecting cases in a map • Identifying cases in a graph, table, or map by selecting cases in a table
Inspecting	Obtaining Information	Actions of hovering or clicking on an object to gain information	• Hovering on an attribute name in any representation to obtain a definition or formula • Accessing a tooltip in a graph or map by hovering or clicking on a case or geographic region • Hovering on a measure in a graph or part of a graph to obtain its value (e.g., median, Q1 in a boxplot, count/percent on a bar graph) • Hovering on the legend for a quantitative attribute displayed as a legend attribute of a graph/map to obtain the interval of values represented by that color
	Locating	Actions using a geospatial representation to situate a case or subset	• Clicking on a map to locate and situate cases based on their location (e.g., where is that state?; it looks like there are lots of cases in NC)
Expanding Datasets	Adding Data	Actions to add data values	• Entering data values in a case card or data table, often to clean data or add a missing value • Adding a new record to the dataset in a table or case cards • Adding a new attribute and entering values for that attribute manually • Using existing attributes to calculate a new attribute (e.g., defining students per teacher as number of students enrolled / number of teachers in school)
	Merging	Actions that expand a dataset by adding multiple new cases	• Adding another dataset with the same attributes as a current dataset and merging the datasets together (e.g., merging a dataset that contains territories as cases with same attributes to a dataset of U.S. states)
	Joining	Actions that append a dataset	• Getting more information about existing cases from a different dataset by appending a dataset using an identifier/label to align cases and to add attributes to existing cases

investigate data in CODAP. We elaborate on these four themes in the next section.

Results

In this section we report on the four major themes that include approaches identified, along with data moves PMTs utilized, to reason about data using CODAP. The examples of four PSTs, with the pseudonyms Laila, Monica, Eleanor, and Samantha, were selected to illustrate significant characteristics of their approaches; so, although each PMT used a variety of data moves, we focus our findings on the data moves that were most salient for each PMT. At the time of data collection, Laila, Monica, and Eleanor were preservice elementary teachers seeking a minor in mathematics teaching and taking an elementary/middle statistics content course (in different semesters), and Samantha was a preservice secondary mathematics teacher enrolled in a course on teaching secondary mathematics with technology.

We provide an example to illustrate each of the four approaches: (1) comparing groups by *using subsets* in graphs, *ordering* and *summarizing*; (2) *linking* across representations, *filtering* and *obtaining*; (3) using maps and *locating*, and, (4) *using subsets* in a hierarchical structure and *highlighting subsets*. In the descriptions that follow, we use underscores and italics to represent names of attributes in the dataset.

Comparing groups by using subsets in graphs, ordering, and summarizing

A common way that PMTs used data moves to compare groups in their investigations was to utilize a combination of *using subsets* in graphs, *ordering* cases within the graph, and *summarizing* cases using measures of center and spread. For example, Laila began her investigation by *using subsets* to add a categorical attribute to the horizontal axis to *group* cases by *Census_Region*. She then created stacked dotplots for *Average_Salary,* an *ordering* data move. She noted the South showed a cluster of points near the bottom of the scale. To investigate further, she engaged in a series of *summarizing* data moves, including adding means, boxplots (shown in Figure 1), mean absolute deviations (MADs), and numeric counts (e.g., 17 for South) to the graph. After making each data move, she analyzed the graph using the measure or representation shown. She noted that the West had slightly better average salaries than the South, so based on the means, she would probably want to teach in the West. When analyzing the boxplots, she identified some points in both regions that she thought were outliers. She said she used an equation to calculate whether these values were indeed outliers. (CODAP allows users to show outliers when a boxplot is displayed, but she

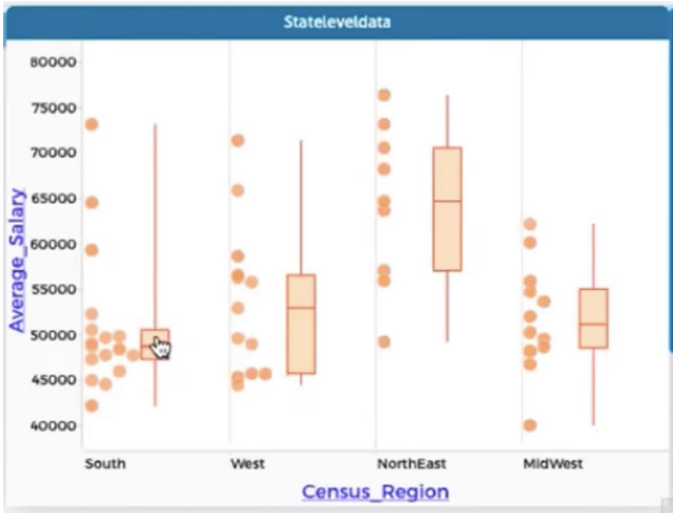

Figure 1. Laila's stacked dotplots with overlaid Boxplots of average teacher salaries.

did not appear to be aware of this.) When MADs were displayed, Laila called points in the South "more accurate, because [the West] is more spread out." Regarding counts, she read 17 and 13 as the counts in the South and West, respectively, and she stated, "Because they're not the exact same amount of points, we can't say for certain which region is better, because they're not accurate in number, cause they're not the same." Laila demonstrated some problematic reasoning about the data, namely that she seemed to equate variation with being less accurate and that subsets needed to have the same number of values in order to make comparisons.

Laila next used an *ordering* move by replacing *Average_Salary* on the vertical axis with *Expenditure_per_Student.* This move resulted in reordering the data values in each subset in order from least to greatest for *Expenditure_per_Student*. For this attribute, her first *summarizing* move was to display the standard deviations, as shown in Figure 2. Through *obtaining information,* she reported that the South had a slightly lower standard deviation than the West. She again referred to the South's points as "more accurate to the average." She removed the standard deviations from the graph and *summarized* by adding a mean. Although she reported that you would want to choose the region with the higher level of expenditures per student, she also admitted there were not large differences between the two regions.

Laila continued to make multiple *ordering* and *summarizing* data moves with different attributes. She investigated means and boxplots for *Revenue_per_Student* and *Students_per_Teacher* (see Figure 3) to compare the South and West. She then shifted her attention to look at all four regions using *Average_Salary*, *Students_per_Teacher*, and *Revenue_per_Student.* She concluded that teaching in the Northeast was her preference across all four regions.

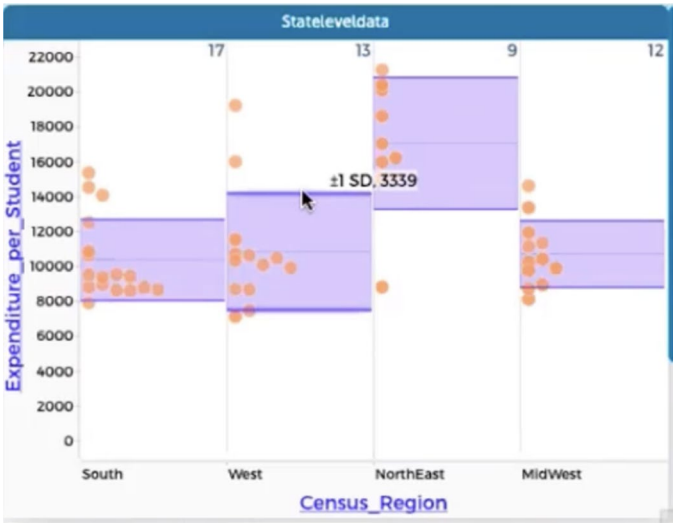

Figure 2. Laila's stacked dotplots with standard deviations of expenditures per student.

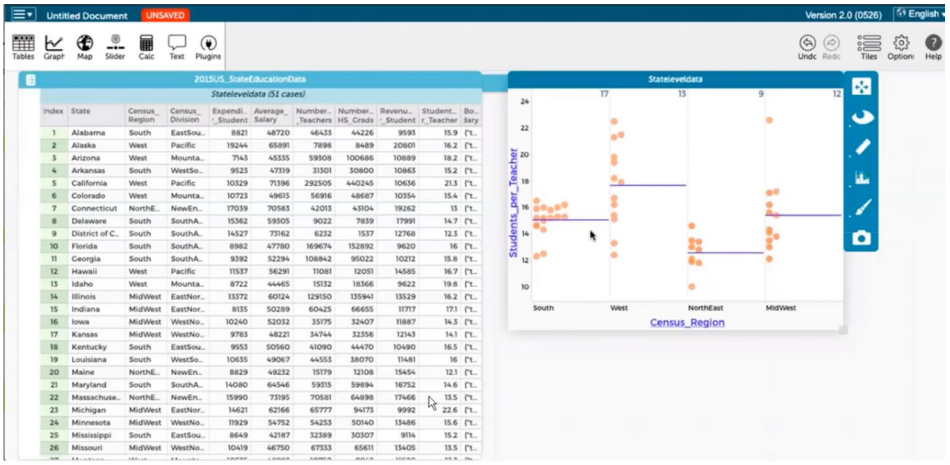

Figure 3. Laila's workspace in CODAP while investigating *Students_per_Teacher*.

Laila only ever had a single graph open in her CODAP workspace, as shown in Figure 3). She maintained *Census_Region* in the horizontal axis and repeatedly replaced attribute names along the vertical axis, an *ordering* move. Although Laila's repetitious actions allowed her to make comparisons across groups, her reasoning was limited to inferences that could be made by observing a single graph.

Linking across representations, filtering and inspecting

Monica is an example of a PMT who linked across representations, using *filtering* and *inspecting* data moves. She began by dragging *Average_Salary*

from the case table to the horizontal axis of an unordered graph to create a dotplot, an *ordering* data move. She then employed *using subsets* to add a categorical attribute (*Census_Region*) to the vertical axis to create stacked dotplots, allowing her to visually compare the distributions of each region on the same graph. She then used a *filtering* move, first *highlighting subsets* of the Northeast and Midwest cases in the graph, and then hiding the selected cases since they were "irrelevant" to her investigation. Finally, she used a *summarizing* data move to plot means of the average salaries in both regions. Through *inspecting*, hovering with her cursor, she *obtained information* about the mean values and stated that the West had a higher average salary as compared to the South (see Figure 4). She concluded that if only the mean average salaries in the South and West were considered, she might prefer to teach in the West but that "there are a lot of other variables to go along with that."

Monica turned her focus to exploring the number of teachers by region. Utilizing a *using subsets* move, she drug the quantitative attribute *Number_of_Teachers* into the center of the graph in Figure 4 to create a legend as shown in Figure 5. This *using subsets* move allowed Monica to *group* cases into numerical intervals shown by color, from least (lightest color gradient) to greatest (darkest color gradient). With three attributes on the graph, she *obtained information* about several states (see rectangle embedded in Figure 5) and reasoned the West offers teachers more money, because there is a higher demand for teachers and the cost of living in the West is higher than the South, concluding there are more teachers in the South than the West which could be the reason western teachers are getting paid more.

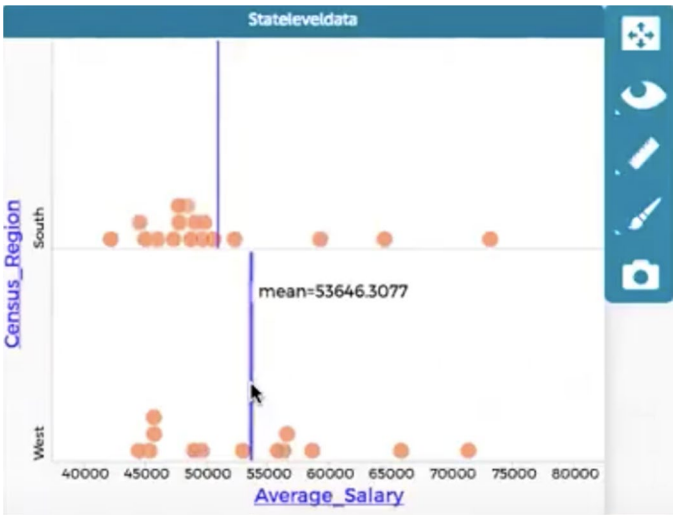

Figure 4. Monica's graph comparing means of average teacher salaries.

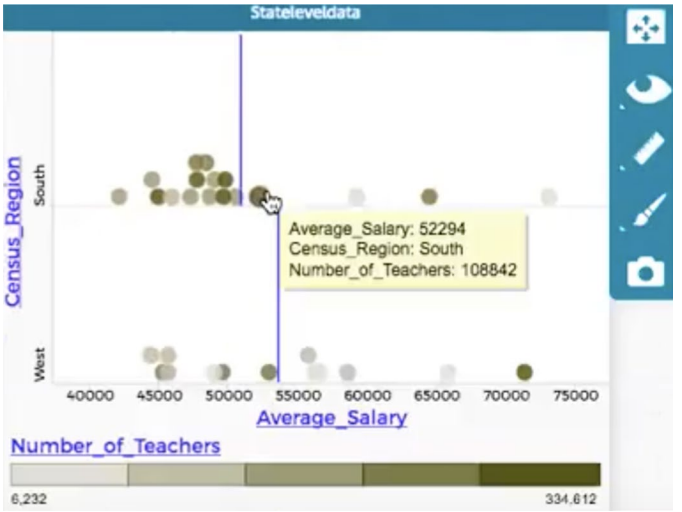

Figure 5. Monica's graph displaying average teacher salaries and the number of teachers by region.

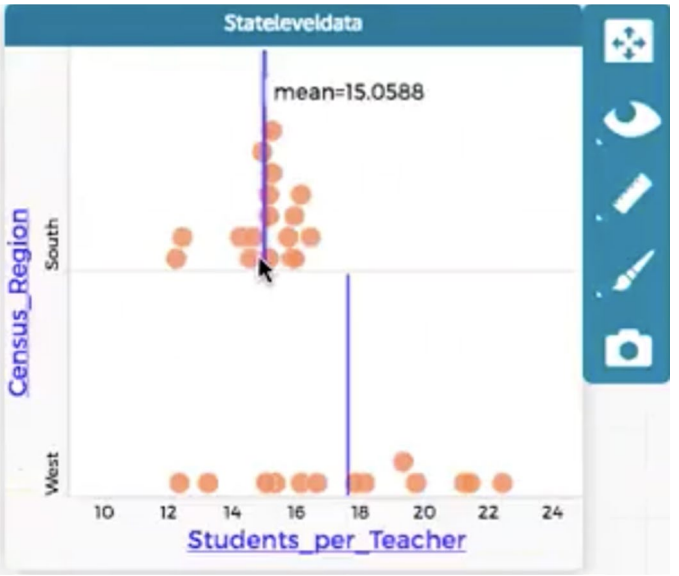

Figure 6. Monica's graph comparing student-to-teacher ratio in the South and West.

Recognizing that other attributes may be important to consider, Monica used similar data moves to examine the ratio of *Students_Per_Teacher* by (see Figure 6): (1) creating a new graph, *ordering* the ratios on the horizontal axis; (2) *using subsets* by dragging *Census_Region* to the vertical axis; (3) *filtering* irrelevant cases by *highlighting* and hiding cases in the Northeast and Midwest; and, (4) *summarizing* to plot the means of the ratio of *Students_Per_Teacher*. As shown in Figure 6, she *obtained information* about the average teacher salaries in the South by hovering over the plotted means.

Unlike Laila's approach of using one graph throughout her investigation, Monica's main goal throughout her investigation seemed to be to consider multiple attributes to inform her decisions. Like many PMTs in our study, she compared means of average teacher salaries of the South and West. Many PMTs also went on to explore other attributes one at a time, comparing particular attributes by region. Monica differed in her approach to many PMTs in our study in that she reasoned across more than two graphs simultaneously.

The significance of Monica's approach is that she utilized CODAP's capabilities to *link* multiple representations. This *linking* move characterized the majority of her actions, illustrating her purposeful data move to connect cases across representations and multiple attributes supporting her multivariate reasoning. In addition to *linking*, she frequently used *inspecting* to *obtain information*. While examining Figures 5 and 6 side by side (see Figure 7), Monica concluded that teachers in the West get paid more; they "have more students in their classrooms, that is a lot more work, a lot more planning. And in the South you may be getting paid less, but you are also dealing with less students in your classroom." Here she used stacked doptlots of *Average_Salary* and *Students_Per_Teacher*, along with the means of each attribute per region and the gradient colors of *Number_of_Teachers* to reason about multiple attributes, which would be challenging to do without CODAP.

Monica continued to utilize CODAP to provide more information related to *Number_HS_Grads*, where again she engaged in the data moves of *ordering*, *using subsets, highlighting subsets, filtering* and *summarizing* to produce stacked dotplots of *Number_HS_Grads* disaggregated for South and West (shown in the lower right corner of Figure 8). As she articulated her thinking, where she continuously *linked* multiple representations and *obtained information*, Monica reasoned using all three graphs and the case

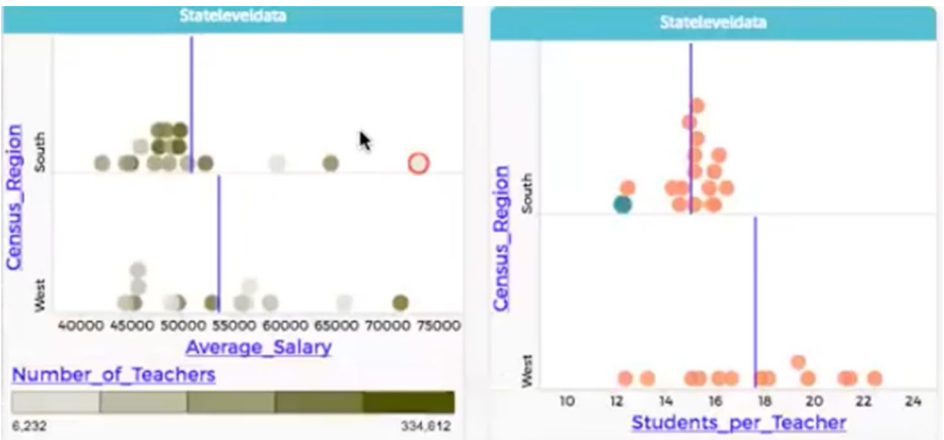

Figure 7. Monica's linking across multiple representations.

table (see Figure 8). While connecting information about California across all representations (i.e., graphs, case table, and map), she explained,

> that "we have this one right here that has the highest of the West for graduates" [hovered over California, clicked on case in *Number_HS_Grads* graph]. "And you would also be getting paid the highest in the West in your salary" [hovered over California, clicked in *Average_Salary* graph]. "But there is also a lot of teachers … and you would be required to have around 21.3 students per teacher" [hovered over California, clicked in *Students_per_Teacher* graph]. "So even though it looks like you are getting paid a lot more, and you will be having a lot of graduated students, you would be having a higher student-per-teacher ratio".

Monica continued *linking* across multiple representations and *obtaining information* by clicking on individual states in one dotplot and observing where that state fell in the distributions displayed in the dotplots for the other two attributes. She described her reasoning about these states using each of the three attributes, first focusing on the state with the highest average salary in the South (D.C.) and then cases that were near the average salary for each region (i.e., Kentucky and Washington). While examining D.C., she reasoned that it had one of the lowest numbers of high school graduates in the South but the highest average salary in the South. She further noted that D.C. has one of the smallest student-to-teacher ratios, and questioned whether that might reflect the small number of high school graduates, and mentioned there are many variables that might influence this.

She reasoned that if you only consider average salary, one might want to teach in the West. She pointed out that the student-to-teacher ratio was higher in the West, which she deemed less desirable. She also indicated the importance of accounting for where there are fewer high

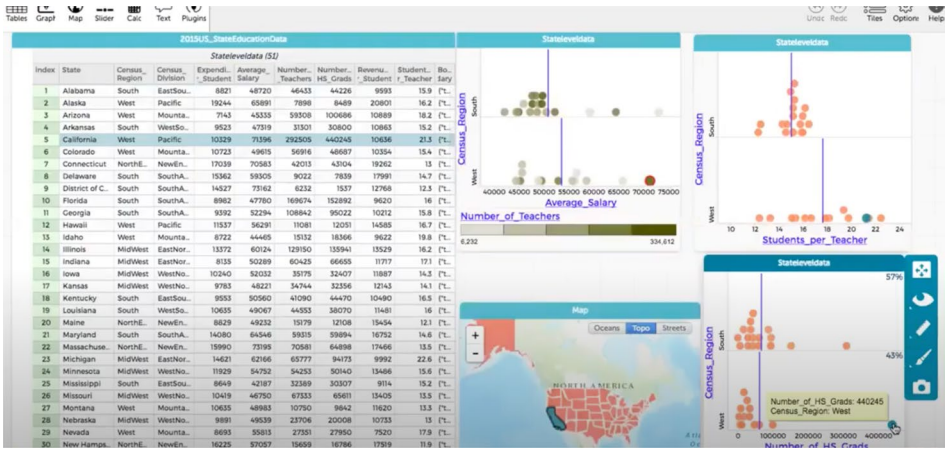

Figure 8. Monica's workspace linking *Average_Salary*, *Number_of_Teachers* by Region, and *Students_Per_Teacher*.

school graduates. She concluded her investigation by emphasizing that from looking across all three graphs she would personally choose the South.

Using maps and locating

Unlike Laila and Monica's approaches that relied heavily on creating graphs, Eleanor used two maps and *locating*. She *used subsets* to overlay *Census_Region* on Map 1 (see Figure 9), taking advantage of CODAP's ability to use color to visualize distinct groups (e.g., states in the South are colored pink). She also *created subsets* by placing the quantitative attribute *Average_Salary* on her other map (Map 2), creating a legend which *grouped* cases in several bins from least (lightest color) to greatest (darkest color). She used *locating* while saying "I can see the South has the lower end of the salaries [used mouse to encircle southern states], and the West has the higher end of the salaries [moving her mouse along the western coast of the U.S.]. So based on average salaries, I would prefer to teach in the West because they have more variety of higher pay than the South."

Several states' census regions (i.e., D.C., Delaware, and Maryland) were identified as South, but these states are geographically and culturally connected to the Northeast. These states had very high average teacher salaries and were outliers in the South. PMTs who used stacked dotplots and *inspected* to *obtain information* about individual states within regions with the highest average teacher salaries were often surprised or grappled with whether these states were really in the South. Eleanor did not struggle with this. Eleanor likely used Map 2 to visually identify states where the darkest colors were located while at the same time using Map 1 to help her keep track of where the darkest states were located within regions.

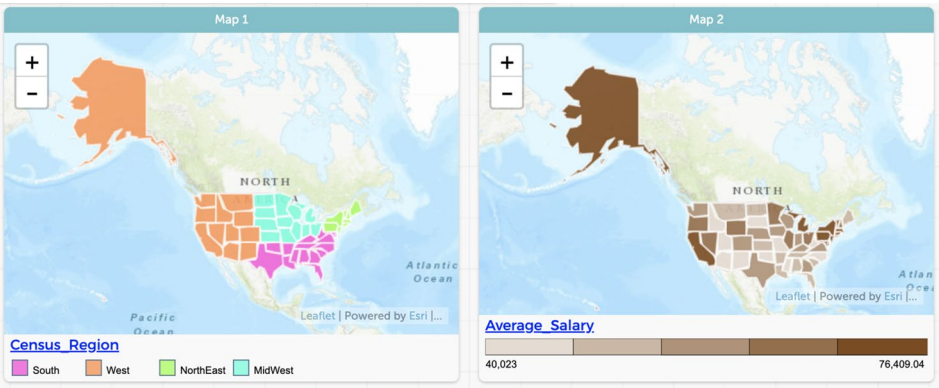

Figure 9. Eleanor's maps showing states colored by region (Map 1) and average teacher salaries (Map 2).

Because D.C., Delaware and Maryland are small and were somewhat challenging to locate without *locating* the individual states, Eleanor reasoned more broadly about the aggregate. Although the evidence she provided using this unique capability of CODAP may not have been the strongest argument about choosing a region based on salary, we argue that keeping track of the states' region in a map supported her reasoning about the aggregate of average teacher salaries in the South and West.

Using groups in a hierarchical structure and highlighting groups to link

In contrast to the other approaches we have shared, Samantha utilized a unique feature of CODAP allowing the user to organize a case table using a hierarchy, where she used *grouping* and *highlighting subsets*. Samantha framed her screencast as if she were talking to her hypothetical students as an audience. She began her exploration examining *Census_Region* as an attribute for a map, similar to Eleanor's Map 1, shown in Figure 9. Next, Samantha attended to her data table by clicking on *Census_Region* and dragging it to the far left side of the data table. In CODAP, this *using subsets* action changes the structure of the data table by creating a hierarchical structure. The result is shown in Figure 10, where the states are *grouped* by region. Since her statistical question was limited to the South and West regions, Samantha collapsed the Northeast and Midwest groups, so the data table only showed the two focus regions. In contrast to Laila's and Monica's approaches of calculating means in graphs, Samantha initially chose to create a new attribute, which she titled *Mean_Salary,* and created a formula to *summarize* the means of the states' average salaries in each

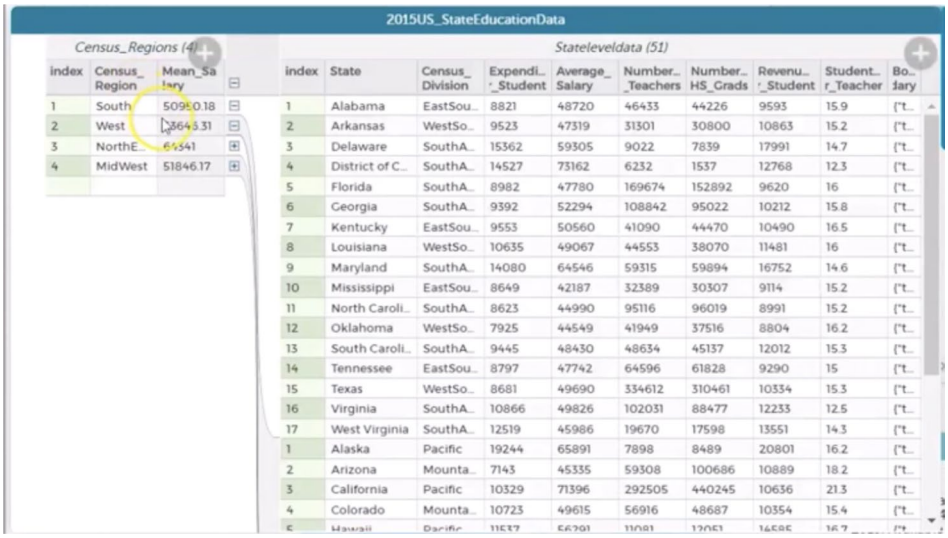

Figure 10. Samantha's data table with hierarchy by *Census_Region* and calculated means.

region within the hierarchical table. She noted there was about a $3000 difference between the salaries in South and West, and stated "that might be worthwhile."

Samantha then decided to investigate *Average_Salary* graphically by dragging it to the graph to create a dotplot of the 51 states' salaries, an *ordering* data move. She traced the shape of the graph from left to right as she said, "We might say if we plotted this as a histogram, that it was right-skewed, because the tail would be over here [on the right]." This statement suggests that Samantha may have thought that the shape of a dataset's distribution is dependent upon how the data is displayed. However, she made a claim about the distribution's shape without seeing the data in a histogram. Next, Samantha added the mean (in blue) and the median (in red) to the dotplot and stated that this is a right skewed dataset, "since the mean is larger than the median." Samantha's representation is shown in Figure 11. Samantha used CODAP as a tool to explore data, and she described an important statistical concept about the relationship between a dataset's mean, median and shape. However, these actions and the accompanying reasoning did not necessarily lead to answering her question about differences between the South and West regions.

Samantha pointed out that the mean's larger value (compared to the median) and then said, "we would also know that we might have some outliers over here [as she circled points on the right side of the distribution] that might be messing with it." She overlaid a boxplot over the dotplot and removed the mean and median. She clicked on the left whisker of the boxplot, which resulted in three changes to the workspace, as shown in Figure 12—the points below the first quartile were *highlighted as a susbset* in the dotplot, the corresponding rows in the data table *highlighted*

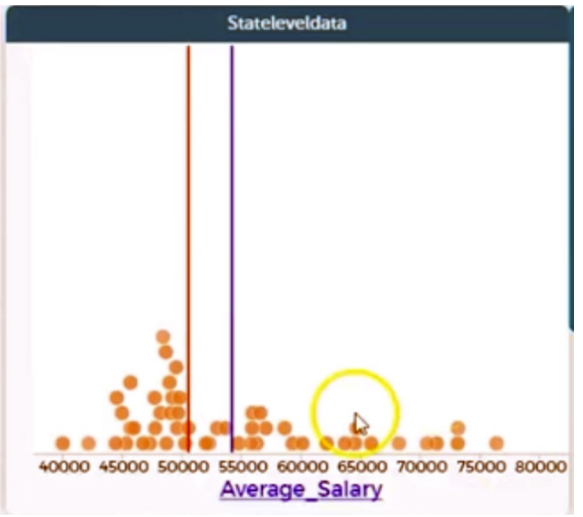

Figure 11. Samantha's Dotplot of average salaries with mean and median displayed.

as a subset, and the states associated with these points were *highlighted* using red outlines in the map. *Linking* these representations, Samantha said, "In our bottom 25% for average salary, we can see a lot in the pink states, so our South region, and four in our Western states." Similarly, she clicked on the right whisker to *highlight a subset* of states above the third quartile. She noted, "If you look at the upper 25%, we've only got two states in the West…I think we have one state in the South, Maryland, in the upper average teacher salary." Although not visible in the map, Delaware and DC also fell in the upper 25% of the data.

Samantha, similar to Laila's and Monica's data moves of *using subsets* and *summarizing* described above, continued her investigation by creating stacked dotplots for *Average_Salary* and *Students_per_Teacher*, disaggregated by *Census_Region*. She concluded, "For me personally as a teacher, I care a lot about how many students I'm going to have in my classroom, and I prefer to have a smaller number of students. So, I might choose to teach in the South, even though I have a lower salary."

Discussion

Characterizing data moves in CODAP

The current study builds directly on the work of Erickson et al. (2019) who advocated that data moves should play a more prominent role in K-12 students' exploration of data. Learning to engage in these data moves is critical for data investigations involving real-world, multivariate, and messy data (c.f. Ben-Zvi et al., 2018; Gould et al., 2018; Kahn & Jiang,

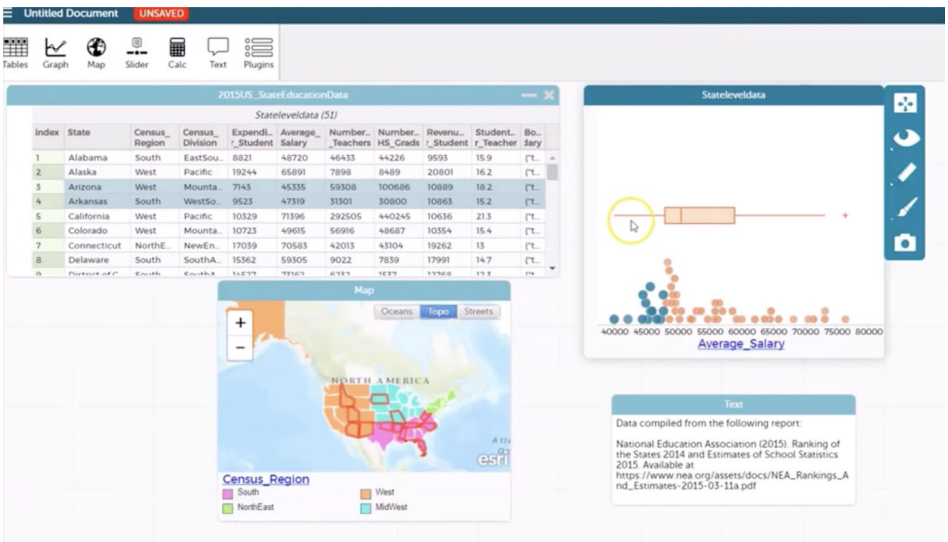

Figure 12. Samantha's CODAP workspace with cases below the first quartile highlighted in a dotplot and outlined in a map.

2020). The framework described in Table 1 clarifies and extends Erickson et al.'s initial work and expounds specific data moves by identifying a list of possible actions in CODAP aligned to each data move. The framework was refined by observing the nuanced moves of PMTs as they investigated data in CODAP. As described above, the framework disaggregates the *grouping* and *expanding datasets* data moves into sets of three types of data moves. We also observed the PMTs engage in data moves beyond those described by Erickson et al. (2019). The PMTs frequently *inspected* the data by hovering over cases, attributes, or measures to *obtain information* or *locate* cases geographically. *Locating*, as demonstrated by Eleanor, was a particularly valuable data move given the context of the dataset and CODAP's capability of using maps as a data representation. Although the action of *inspecting* does not change the structure of a dataset, it changes the representation so a user can learn more about the data. Building on others' conceptualizations of Sorting (Erickson et al., 2019; van Borkulo et al., 2023), we characterized *ordering* as a data move that almost all of the PMTs used when creating dotplots, creating scatterplots, or reordering categorical attribute names. We also identified a common data move of *highlighting subsets*, a type of *grouping* data move often paired with other data moves. For example, Monica first *highlighted subsets* to select cases prior to *filtering* them. *Highlighting subsets* may also be used when *linking* representations, such as when Samantha selected the lower whisker in a boxplot and *linked* the corresponding points to states on a map. Our definition of data moves reflects the user's intent, so *linking* takes place when a user intentionally selects cases in one representation to identify cases in other representations.

Although our lens for describing PMTs' approaches focused on their data moves, or combinations of data moves, contrasts between the four themes described above are explicitly connected to affordances of using dynamic data tools. Laila reasoned using single graphs, a process that is commonly accessible when analyzing static data representations. However, CODAP's capabilities allow users to engage with data through data moves that would not be possible with traditional, paper-and-pencil tasks (Mojica et al., 2019). For example, users often move between exploring data and considering models while investigating a question. The ease of being able to create multiple graphs (e.g., dotplots, boxplots) and calculate and plot visuals of statistical measures, allows users to make, test and possibly refine conjectures which can be confirmed or refuted quickly with evidence as demonstrated by our examples of PMTs' approaches. Similarly, maps can be used to explore and draw conclusions from geospatial data with other attributes, as Eleanor's work demonstrated. Additionally, CODAP supports developing multivariate reasoning, where users can graph and examine relationships between up to four attributes on the same graph,

and *linking* multiple graphs, tables and maps as in the example of Monica's work. Using different data moves also provides important information to the user when looking for trends or patterns, such as Laila *obtaining information* about standard deviations or Samantha *highlighting subsets* of states in a particular quartile. However, Samantha's actions were quite different than the other PMTs, because she actually restructured her data into a hierarchy and *expanded the dataset* to make comparisons, capitalizing on a significant feature of CODAP.

Although the data moves framework (shown in Table 1) was developed through examining teachers' work with a specific data tool, the data moves and purposes are also applicable to ways learners engage in processing, wrangling, and analyzing data using other tools (e.g., Israel-Fishelson et al., 2023; Jiang & Kahn, 2020; Sanei et al., 2023). For example, data is easily *ordered* in a spreadsheet using different actions to "sort" data, and *filtering* a dataset is easily accomplished in R with a filter function where a user defines the conditions that must be true for a case to remain in the dataset. All data tools support capabilities to *use subsets* and *create subsets* using various commands or actions specific to that tool. Thus, we encourage others to use the data moves framework to support their work with teachers and students, regardless of the data tool used, and for researchers to continue to expand the framework to identify types of data moves and purposes as they study how computational data tools are used to support data investigations.

Design of data investigation task

Aspects of the U.S. State Education dataset afforded unique opportunities for PMTs to investigate and reason with data in ways suggested by many (Ben-Zvi et al., 2018; Gould et al., 2018; Rubin, 2020). Questions investigated by PMTs required comparisons between the South and West census regions, which often led to the creation and interpretations of representations showing bivariate or multivariate data. Furthermore, the units within the dataset represented states, not individuals. This feature required PMTs to attend to their statistical conclusions with precision. For example, PMTs who analyzed *Average_Salary* in a dotplot needed to recognize that each dot represented an average of salaries, rather than individual teachers' salaries. In some instances, PMTs made statements that suggested they were not considering that each dot represents many salaries. Variation exists within each state's individual teachers' salaries that are not accounted for in this dataset.

The context of the dataset was of high interest and relevance to PMTs, who are planning careers in the field of education. PMTs used their contextual knowledge of education (e.g., a desire to have low student-teacher

ratios) and geography (e.g., cost-of-living) as they considered which region would be ideal. Like many real-world datasets, this dataset included attributes and cases that were unnecessary to answer the questions posed. PMTs needed to engage in contextual reasoning along with their data moves to make sense of specific attributes. Data moves such as *ordering*, *summarizing*, *creating subsets*, *inspecting*, and *linking* seemed to support such contextual reasoning,

The screencast project described in this article has potential for use in teacher education and other settings, because it permits the instructor to observe the data moves the learner (e.g., PMTs) engages in. Insights can be gleaned not only from what a learner does but also from what actions they do not make. For example, Laila did not *filter* the data or create multiple representations to *link*. Instructors may want to attend to data moves taken by individual students to guide future conversations and differentiate instruction for individual learners. However, the dataset may not include opportunities to engage in all the potential actions in CODAP identified in the Data Moves Framework. For example, one of the *ordering* examples in Table 1 is treating a numeric attribute as categorical. The task's statistical questions could be answered without engaging in this action. The data moves utilized are driven by aspects of the dataset and by the statistical questions posed.

Conclusion

Although CODAP is a powerful tool that has incredible potential to be used by students and teachers to transform and investigate data, very little research has provided evidence for how teachers interact with this tool. The current study demonstrates how teachers can use CODAP to employ data moves that are inherent in data science, such as modifying the structure of a dataset, producing linked data representations, and investigating geospatial data. Although other tools exist with these capabilities, CODAP was uniquely designed to foreground learners and learning (Finzer, 2013; Mojica et al., 2019), and it can be transformative in what statistics and data science concepts and skills are accessible within the secondary curriculum and consequently, for teacher education. However, in order for students to engage with dynamic data tools like CODAP, teachers must develop their technological pedagogical statistical knowledge (Lee & Hollebrands, 2011) to use and teach with such data tools, which will require changes to current practices in teacher education programs (McCulloch et al., 2021). The screencast assignment and U.S. State Education dataset have been designed as part of a curriculum for preparing teachers for this purpose. Teacher education programs must support teachers to learn about and engage in data moves using dynamic data

tools to ensure the next generation of teachers are prepared to adeptly provide learners with meaningful experiences involving authentic, multivariate data.Notes

Acknowledgments

The authors would like to thank Christina Bissada, Heather Barker, and Taylor Harrison for their contributions to this work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the National Science Foundation under grants DUE 1625713, DUE 2141727, DUE 2141716 and DUE 2141724 awarded to NC State University, Eastern Michigan University and University of Southern Indiana. Any opinions, findings, and conclusions or recommendations expressed herein are those of the principal investigators and do not necessarily reflect the views of the NSF.

ORCID

Rick A. Hudson http://orcid.org/0000-0002-4834-6815

References

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education (GAISE) report II*. American Statistical Association and National Council of Teachers of Mathematics. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Ben-Zvi, D., Gravemeijer, K., & Ainley, J. (2018). Design of statistics learning environments In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 473–502). Springer. https://doi.org/10.1007/978-3-319-66195-7

Burr, W., Chevalier, F., Collins, C., Gibbs, A. L., Ng, R., & Wild, C. (2020). Computational skills by stealth in secondary school data science. arXiv preprints. https://doi.org/10.48550/arXiv.2010.07017

Casey, S. A., Harrison, T., & Hudson, R. (2021). Characteristics of statistical investigations tasks created by preservice teachers. *Investigations in Mathematics Learning*, *13*(4), 303–322. https://doi.org/10.1080/19477503.2021.1990659

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2010). Developing and using a codebook for the analysis of interview data: An example from a professional development research. *Field Methods*, *23*(2), 136–155. https://doi.org/10.1177/1525822X10388468

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*, *12*(1), 1. https://doi.org/10.5070/T5121038001

Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, *7*(2), 91. https://doi.org/10.5070/T572013891

Gould, R., & Çetinkaya-Rundel, M. (2014). Teaching statistical thinking in the data deluge. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen–Using Tools for Learning Mathematics and Statistics* (pp. 377–391). Springer Spektrum. https://doi.org/10.1007/978-3-658-03104-6_27

Gould, R., Wild, C. J., Baglin, J., McNamara, A., Ridgway, J., & McConway, K. (2018). Revolutions in teaching and learning statistics: A collection of reflections. In D. Ben-Zvi, K. Makar, J. Garfield (Eds), *International handbook of research in statistics education*. Springer. https://doi.org/10.1007/978-3-319-66195-7_15

Israel-Fishelson, R., Moon, P. F., Tabak, R., & Weintrop, D. (2023). Preparing students to meet their data: An evaluation of K-12 data science tools. *Behaviour & Information Technology*, *2023*, 1–20. https://doi.org/10.1080/0144929X.2023.2295956

Jiang, S., & Kahn, J. (2020). Data wrangling practices and collaborative interactions with aggregated data. *International Journal of Computer-Supported Collaborative Learning*, *15*(3), 257–281. https://doi.org/10.1007/s11412-020-09327-1

Kahn, J., & Jiang, S. (2020). Learning with large, complex data and visualizations: Youth data wrangling in modeling family migration. *Learning, Media and Technology*, *46*(2), 128–143. https://doi.org/10.1080/17439884.2020.1826962

Lee, H. S., Casey, S., Hudson, R. A., Mojica, G. F., Azmy, C., Barker, H., Harrison, T. (2021). Enhancing Statistics Teacher Education with E-Modules. Collection of three online e-modules and two summative assignments. https://place.fi.ncsu.edu/local/catalog/course.php?id=22&ref=3

Lee, H. S., & Hollebrands, K. F. (2011). Characterising and developing teachers' knowledge for teaching statistics with technology. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 359–369). Springer. https://doi.org/10.1007/978-94-007-1131-0_3

Lee, H. S., Hollebrands, K. F., & Wilson, P. H. (2010). *Preparing to teach mathematics with technology: An integrated approach to data analysis and probability* (1st ed.). Kendall Hunt Publishing.

Lee, H. S., Kersaint, G., Harper, S. R., Driskell, S. O., Jones, D. L., Leatham, K. R., Angotti, R. L., & Adu-Gyamfi, K. (2014). Prospective teachers' use of transnumeration in solving statistical tasks with dynamic statistical software. *Statistics Education Research Journal*, *13*(1), 25–52. https://doi.org/10.52041/serj.v13i1.297

Lee, H. S., Mojica, G. F., Thrasher, E. P., & Baumgartner, P. (2022). Investigating data like a data scientist: Key practices and processes. *Statistics Education Research Journal*, *21*(2), 3. https://doi.org/10.52041/serj.v21i2.41

Lee, V. R., & Wilkerson, M. (2018). Data use by middle and secondary students in the digital age: A status report and future prospects. Commissioned Paper for the National Academies of Sciences Engineering, and Medicine, Board on Science Education, Committee on Science Investigations and Engineering Design for Grades 6-12. https://digitalcommons.usu.edu/itls_facpub/634/

Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 353–373). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-2278-6

McCulloch, A. W., Leatham, K. R., Lovett, J. N., Bailey, N. G., & Reed, S. D. (2021). How we are preparing secondary mathematics teachers to teach with technology: Findings from a nationwide survey. *Journal for Research in Mathematics Education*, *52*(1), 94–107. https://doi.org/10.5951/jresematheduc-2020-0205

McCulloch, A. W., Hollebrands, K., Lee, H. S., Harrison, T., & Mutlu, A. (2018). Factors that influence secondary mathematics teachers' integration of technology in mathemat-

ics lessons. *Computers & Education*, *123*, 26–40. https://doi.org/10.1016/j.compedu.2018.04.008

Mojica, G. F., Azmy, C. N., & Lee, H. S. (2019). Exploring data with CODAP. *The Mathematics Teacher*, *112*(6), 473–476. https://doi.org/10.5951/mathteacher.112.6.0473

Peters, S. A., Bargagliotti, A., & Franklin, C. (2024). Engaging and preparing educators to teach statistics and data science. In B. Benken (Ed.), *The AMTE handbook of mathematics teacher education: Reflection on the past, present and future – paving the way for the future of mathematics teacher education*, (Volume 5, pp. 123–150). Information Age Publishing.

Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences*, *29*(1), 154–164. https://doi.org/10.1080/10508406.2019.1705665

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE.

Sanei, H., Kahn, J. B., Yalcinkaya, R., Jiang, S., & Wang, C. (2023). Examining how students code with socioscientific data to tell stories about climate change. *Journal of Science Education and Technology*, *33*(2), 161–177. https://doi.org/10.1007/s10956-023-10054-z

Sukol, S. (2024). *Beyond borders 2024: Primary and secondary data science education around the world*. Report commissioned by Data Science 4 Everyone. https://www.datascience4everyone.org/post/beyond-borders-2024

Sundin, L., Sakr, N., Leinonen, J., Aly, S., & Cutts, Q. (2021, November). Visual recipes for slicing and dicing data: Teaching data wrangling using subgoal graphics. In *Proceedings of the 21st Koli Calling International Conference on Computing Education Research* (Article 29, pp. 1–10). https://dl.acm.org/doi/pdf/10.1145/3488042.3488063 https://doi.org/10.1145/3488042.3488063

van Borkulo, S. P., Chytas, C., Drijvers, P., Barendsen, E., & Tolboom, J. (2023). Spreadsheets in secondary school statistics education: Using authentic data for computational thinking. *Digital Experiences in Mathematics Education*, *9*(3), 420–443. https://doi.org/10.1007/s40751-023-00126-5

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1–23. https://doi.org/10.18637/jss.v059.i10

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science*. O'Reilly Media.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Zieffler, A, & Catalysts for Change (2023). *Statistical Thinking: A simulation approach to uncertainty* (4th ed.). Catalyst Press.