"It didn't really go very well": Epistemological Framing and the Complexity of Interdisciplinary Computing Activities

Serena Thoma, Elise Deitrick, & Michelle Hoda Wilkerson
Tufts University; University of California, Berkeley
{serena.thoma;elise.deitrick}@tufts.edu, mwilkers@berkeley.edu

Abstract: There are a growing number of frameworks for integrating computing into the K-12 curriculum, but these do not provide much insight into what students' thinking and learning is expected to look like at the interface between computing and the disciplines. Prior research suggests the success of disciplinary integration may lie in students' epistemological framings, their expectations about what knowledge and goals are relevant for a given activity. Here we present a detailed analysis of one students' efforts to investigate the relationship between coal consumption and production in the U.S. using RStudio to manage and visualize data. We found that the student, Audrey engaged in coding and debugging behavior that suggested that she was framing the activity as distinctly *computational* or *statistical*; when she focused on one domain, this inhibited or interfered with her sense making about the other.

The growth of computing—as a field in its own right, and as a tool used within and across professional disciplines—has been a major reason for the "rethinking of learning in the digital age". Faced with an ever-compressed curriculum and the emergence of new computational practices within the disciplines (Chandrasekharan & Nersessian, 2015), learning scientists are beginning to explore how computing can be integrated across a variety of school subjects. This has led to the development of a number of taxonomies and frameworks that highlight alignment between computational tools and methods, and the disciplinary core concepts and practices they support (Barr & Stephenson, 2011; Weintrop et al., 2016). While such taxonomies shed light on what computationally-infused *curricula* may look like, they provide less insight into how *thinking and learning* might, or should, look as students engage with these curricula.

When it comes to student learning and sensemaking, however, infusing the tools and methods of one discipline into another may not be as straightforward as it seems. For example, Bing and Redish (2009) found that undergraduates exhibited a number of different understandings of how mathematics may function as a tool for physics problem solving. These included that mathematics provides a reliable method to obtain results through calculation; mathematics can be used to model physical relationships or observations; or that citing mathematical rules invokes authority to validate a solution. Depending on which role students expected mathematics to serve in their physics problem-solving, they were likely to make use of different knowledge and resources to solve problems; in this way, these expectations could significantly limit or support their progress.

Importantly, students' expectations about the role of mathematics in Bing and Redish's studies were not static. Instead, they reflected temporary understandings of "what is going on here" (Goffman, 1974), with respect to the physics and mathematics reasoning students believed they were doing. These *epistemological framings*, or expectations about knowledge and learning, are dynamically activated in the context of different material and social configurations. Additionally, the epistemological framings available to students can themselves be tuned over time (McCormick & Hammer, 2016), and can lead to new or different connections across content areas and transfer (Hammer et al., 2002). Indeed, Bing and Redish (2009) suggested that one goal for education may be to create environments that allow students to blend or merge different epistemological framings, thus expanding their available sets of knowledge and resources.

We are interested in learners' epistemological framings concerning the role of computational tools in other disciplinary domains, and how those framings may shift when such tools are integrated into instruction. In particular, this study explores the ways in which high school statistics students might understand the RStudio data analysis environment as a tool for statistical investigation. Here we present a detailed analysis of one students' efforts to investigate the relationship between coal consumption and production in the U.S. using RStudio to manage and visualize data. We found that the student, Audrey, moved between computing and statistics in ways that suggest distinct, at times conflicting epistemological framings of the activity.

Methods

The CodeR4Stats project explores the use of RStudio, a computational data analysis environment, as part of the high school statistics curriculum. The project seeks to introduce students to the power of computing for statistics through collaborative activities grounded in large-scale social and environmental datasets about familiar and engaging topics (local ecology, Star Wars fandom, college admissions patterns). In a multi-year collaboration with an urban New England public school teacher, CodeR4Stats activities were integrated periodically into a general Statistics course throughout the academic year.

In this short piece we focus on Audrey, a high performing student who ranked in the top of her class in terms of homework completion and quiz scores and appeared especially comfortable using RStudio. We found that despite apparent success in the course, Audrey engaged in coding and debugging behavior that suggested that she was framing the activity as “computational” and “statistical” as distinct domains. When she focused on one domain, this at times inhibited or interfered with her sensemaking about the other. Audrey’s case is drawn from a 6-day unit toward the end of the year in which student groups selected a dataset from seven options, developed a research question, and pursued the question through statistical analysis. The datasets were large and complex, which allowed students to ask and pursue a variety of questions and motivated the need for computational methods made available within the RStudio environment to clean, organize, and analyze the data. Consenting groups’ screen activity and discussions were recorded; these were later synchronized for analysis.

We analyzed the data through the lens of epistemological framing as described above. Typically, research that explores framing uses linguistic markers or other discursive features as evidence for participants’ framings (Tannen 1993). However, epistemological framing can also be inferred from a variety of behavior, including para-verbal and nonverbal cues such as hand motions, facial aspect, body position and/or movement, and gaze (Scherr & Hammer, 2009). For this analysis we additionally attend to participants’ interactions with tools including the writing, editing, testing of computer code, referencing informational resources, self-talk, and verbal interactions with her class partner as sources to help identify Audrey’s epistemological framings.

Results

We examine shifts and stabilities in Audrey’s problem-solving behavior as she moves between rather disparate *computational* and *statistical* epistemological framings. We begin by establishing the nature of these two frames, and then trace how those frames affect Audrey’s debugging activities when she encounters a compilation error in her code. The compilation error leads Audrey to believe there is a problem in her code or data, but it in fact is the result of an attempt to create a histogram using plot two variables, rather than only one.

Part 1. Identifying *computing* and *statistics* as distinct frames

On the first day of the project, Audrey began by writing code to import a large data set about U.S. energy production and consumption into the RStudio environment. She created a set of two variables for production and consumption data for each energy source the students were interested in. These variables each had names that corresponded to the data their referenced; for example, the coal consumption data was stored in variable ‘cc’ and the coal production data was stored in variable ‘cp.’

Once Audrey had transformed her data into a more manageable form, she created a variable called “setcc2015” to hold the coal consumption data for 2015 that she extracted from the larger data set (Table 1, Line 118). She created another variable called “setmo2015” to hold month values for that same subset of data, with the intention to plot coal consumption data over time for the year of 2015. Audrey’s facility manipulating these large datasets and using meaningful variable names suggests she was comfortably and stably engaged in *computing*, and leads us to conclude Audrey is framing this first part of her code as a *computational* resource.

However, after creating these subset variables, Audrey redefined the variables as “x” and “y” (Lines 119, 121). She used these new variables to build her plot (Line 122). This switch in variable names is inefficient from a computational standpoint, but reproduces mathematical/statistical convention. This provides tentative evidence that Audrey understands the production of plots as a statistical activity relatively distinct from computing, an epistemological framing that we find more evidence for later in her work on the project.

Table 1. Lines 118-122 of Audrey’s R code

	Audrey’s code	What this line of code does
118	`setcc2015 = subset(cc, year == 2015)`	Creates list of coal consumption values for 2015
119	`x = setcc2015`	Stores list of coal consumption values into variable x
120	`setmo2015 = subset(month, year == 2015)`	Creates list of month values for 2015
121	`y = setmo2015`	Stores list of month values into variable x
122	`plot(x,y)`	Creates plot of coal consumption over 2015

Part 2. *Computing* frame obscures statistical misunderstanding

At the beginning of day 2, Audrey announced to her partner Zach, “Yeah, I wanna try, I wanna try to make a histogram. It didn’t really go very well, I tried it last night.” It is unclear from our records what, exactly, Audrey wanted to plot using the histogram function. She began to pursue the histogram by typing and executing the code `hist(year, cp)`. The `year` variable contained a list of the years of observation of each record in the dataset, and `cp` described the amount of coal production for each record. It seems Audrey wanted to create a histogram using two separate variables, but the histogram function in R only accepts one variable, since histograms by definition display frequencies of single measure across a collection of observations.

When Audrey executed this line of code, RStudio returned an error reading `Error in hist.default(year, cp) : some 'x' not counted; maybe 'breaks' do not span range of 'x'`. This error message is the result of the computer interpreting the second variable inside of the parentheses, `cp`, as the number of breaks in the histogram, when `cp` contains a list of values and not a single integer. Audrey said to Zach, "It just said, like, the data, like for some of the data it says, like, zeroes, so that it wouldn't work, so." Based on this, we posit Audrey interpreted the error to indicate there is something wrong with the data she wishes to graph. Audrey's assumption that there is a discrepancy in the raw data that the function hist() is unable to process leads her to interpret the error as a *computational*, rather than statistical problem. This interpretation is reinforced by the use of the variable x in the error message, which we already know Audrey interprets to mean the first variable in an ordered pair.

Audrey's first attempt at troubleshooting was to run the code `x=hist(year,cp)` (a slight revision to the original code that threw the error), and `x` (which simply prints out the data stored in variable x). The first line of code produced the same error again, but the error was quickly pushed off the screen by the list of data stored in x that was output next. Audrey scrolled through the outputted data. We presume based on our previous interpretation of her utterance, to look for zeros or null values. None were present, and she quickly moved on to the next step of troubleshooting. She attempted to create a scatterplot of `x` using the plot() function, and this generated a cloud of scattered points without any observable trend. Recall that Audrey had stored her own data using in meaningful variable names, and used x and y only transiently when she did plotting. The most recent use of x was as storage for the histogram itself. When the scatterplot appears Audrey responded, "Okie doke, that's not what I want." Given that Audrey was confident plotting data using the plot() function the day before, we interpret all of these actions as efforts to check if her data is usable by a plotting function.

Audrey next referred to online help resources for the RStudio environment using the CodeR4Stats webpage. She scrolled past the examples that show how to build a histogram using R code, which illustrate that histograms only take a single variable, and instead stopped and read examples that demonstrate how to format data for use with plotting functions. Following these examples, she typed and ran a line of code, `x=c(cp)` (cp contains coal production values), which does not produce any error or output. Audrey added two more lines, `y=c(year)` and `hist(x, y)` and executeed the three lines together. Because Audrey passed two lists of data to the histogram function, she again receives the error discussed above.

We interpret this segment of activity to reflect that Audrey's initially ambiguous framing, evidenced by her declaration that she wants create a histogram, quickly shifted to one that emphasized *computational* aspects of the problem. When Audrey received a compilation error, she believed it was due to a corrupt data structure or some other computational mis-step, rather than a more fundamental misuse of histogram as a statistical representational tool. Given Audrey's performance in the class, we have evidence to suggest that this does not reflect a misunderstanding of lack of awareness of what a histogram is. Rather, it seems that Audrey's epistemological framing more strongly activates data and syntax as features of the situation than statistical ideas.

Part 3. Re-orienting to *statistical* activity

The error produced at the end of the last subsection is from Audrey's passing multiple data columns into the `hist()` function, the same error she saw when first beginning to investigate histograms. She responds to this error again in the same way she did previously, by re-evaluating the raw data inputted into the function. However this time Audrey revisited even earlier code, code that she used to initially import and structure the dataset for analysis. She read the data in anew, and re-constructed the relevant data columns. This section of Audrey's code, however, also included the use of `x` and `y` variables that she typically used when moving to statistical (graphing) activity. She added and re-ran the histogram function again, with `x` and `y` as inputs.
This code again produces the same error. However, this time Audrey changed the command to `hist(x)`, perhaps imitating the examples on the resource website that use x as an input. Once Audrey executed this new line of code `hist(x)`, a figure was generated (Figure 2). Upon producing the figure, Audrey spoke with Zach:

```
my.data=read.csv("Energy.csv - Energy.csv.csv")
na.strings=c("energy.csv - Energ.csv.csv", "NA")
my.data=na.omit(my.data)
my.data
x=c(cp)
y=c(year)
hist(x, y)
```

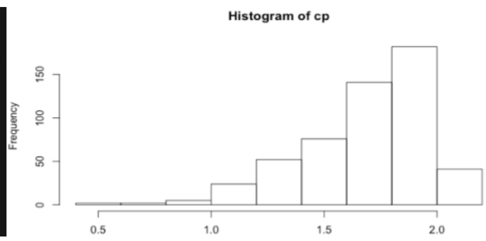

Figure 2: Audrey's R Code (from screen capture) and the resulting histogram (reconstructed to improve quality)

A: I did it! Oh my god.
Z: How did you do it?
A: Uh, let me think about how I just did that. Because that's a great question and I don't really know. [*tracing histogram with cursor*] Ok, so the x is the coal production [*x variable definition in code*], this is coal production [*x axis of histogram*] this is frequency [*y axis of histogram*]; but what we really want is to be year [*x axis of histogram*] and this to be coal production [*y axis of histogram*].

With Zach's help, Audrey shifted from thinking about the *computational* problem of producing the histogram from data via code to thinking about the *statistical* problem of what the histogram actually means. This shift is evidenced in Audrey's focus on mappings between data and the conventions of x and y axes; and her recognition that the histogram that was finally produced is not a form well-aligned with her actual goal. Indeed, Audrey had generated scatterplots of this sort earlier in the project, so the problem was not that she did not have access to the correct functions or difficulty conceptualizing of what scatterplots are. Instead, we argue, a mis-step early in the day and the error message it produced oriented Audrey's attention on *computing*, an epistemological frame that did not include reasoning through the purpose and nature of statistical plots. Since histograms only display the frequency of a single variable, RStudio had not allowed Audrey to produce a figure that did not make statistical sense, but Audrey treated the computer feedback and other available resources as if she had a computational problem rather than a statistical one.

Discussion and Conclusions

Computing tools, especially those specifically designed for data analysis, are powerful in statistics investigations—they can process large amounts of data, quickly produce plots, and have a low overhead since most commands map easily to statistical convention. And, understanding the role of computing in the disciplines is critical for understanding the nature of those disciplines as a whole. However, integration and even clear alignment between a tool and its domain of application does not mean students will recognize connections between the two, especially as they relate to the development of knowledge and understanding. Educators should focus on developing hybrid framings that allow computing and disciplinary practice to co-develop.

Without such attention, we run the risk of the computer acting as gatekeeper or source of statistical knowledge, depending on whether students choose to focus on the computer as an end in itself. For example, in the case above, RStudio would not let Audrey make an incorrect histogram. One might assume this is a good thing—it provides Audrey with feedback that a histogram displays frequencies of only one variable. However, the error led Audrey, through a shift of frames, to de-emphasize statistical knowledge. While this paper presents only a single case study, we are analyzing these dimensions of epistemological framing across students repeated exposures to integrated activity, with interest in students' development of integrated epistemological frames.

References

Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *Acm Inroads*, *2*(1), 48-54.

Bing, T. J., & Redish, E. F. Analyzing problem solving using math in physics: Epistemological framing via warrants. *Physical Review Special Topics-Physics Education Research*, *5*(020108).

Chandrasekharan, S., & Nersessian, N. J. (2015). Building cognition: The construction of computational representations for scientific discovery. *Cognitive Science*, *39*(8), 1727-1763.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* Harvard University Press.

Hammer, D., Elby, A, Scherr, R., & Redish, E (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning: Research and perspectives*. Information Age Publishing.

McCormick, M. E., & Hammer, D. (2016). Stable Beginnings in Engineering Design. *Journal of Pre-College Engineering Education Research (J-PEER)*, 6(1), 4.

Scherr, R. E., & Hammer, D. (2009). Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics. *Cognition and Instruction*, 27(2), 147-174.

Tannen, D. (1993). What's in a frame? Surface evidence for underlying expectations. In D. Tannen (Ed.), *Framing in discourse*, 14-56. Oxford, UK: Oxford University Press.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127-147.

Acknowledgements

This material is based on work supported by the National Science Foundation under Grant Number DRL-1418163. Findings or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation, Tufts University, or the University of California at Berkeley. Thanks to the anonymous reviewers for their time and feedback.