Computational Literacy in the Time of COVID-19

Andrea A. diSessa
Graduate School of Education
University of California at Berkeley

Introduction: My Life as a Computationally Literate Person

I am among the privileged few for whom exercising computational literacy is an everyday affair. For example, I keep all my financial records in a self-constructed database, and I do financial planning with tools I’ve developed for myself. One of the advantages is that I get exactly the information I want, in the visible form I want, and I understand the details of how the tools work—for example, what assumptions they make, which is seldom or never true with on-line “calculators” such as those that estimate your income tax or plot a graph of future savings for retirement. I rarely use algebra to solve everyday mathematical problems because it’s so much easier just to write a program. I prefer to write a program to compute, say, compound interest than to use a formula because (1) it’s much faster than looking up or re-deriving the formula, (2) I can decide whether I want just a single result, a chart of gain over time, or a graph, and (3) I don’t need to make any simplifying assumptions, such as a constant rate of return. To exemplify the last, in planning financially for our sons’ college I wrote a little program that included both of our sons’ expected tuition and board (and relevant dates), our estimated savings rate, raises in my wife’s and my salaries, and also expected large purchases, such as a new car.

Professionally, I keep and analyze video data with my own tools. I outline papers in the same environment in which I program because that system has an excellent hierarchical organization facility (see images later), which is well-suited for good planning and editing, organized in multiple levels. I keep notes at meetings in the same way, since it produces a much better organized summary than linear text. And so on.

Some of this may sound sophisticated, but I am not a programmer. And many of these things are trivial, given a little knowledge of programming and a good environment in which to work.¹ In one our classroom experiments, sixth grade students wrote programs that were far larger and more complex than what I normally make and use.²

But I’m sure my life is not like yours. What follows might be more vivid in your own experience.

A Vignette: Tracking COVID Infections

Here is an exercise that I did for myself, just out of curiosity. Well, curiosity and perhaps some fear of what COVID-19 meant for us and our communities. I wanted to track some critical inflection points on the progression of the pandemic. In particular, I wanted to track whether and

¹ I use the Boxer programming environment. We designed Boxer precisely to be a medium supporting computational literacy See diSessa (2000).
² They wrote video games. You can imagine how complex that might be, with multiple levels, scoring, a lot of narrative, different internal subgames, etc. Some of this is reported in diSessa (2000).
when social distancing had any noticeable effect. Technically, I wanted to see if and when there might be a deviation from the ordinary, purely exponential increase one gets from a situation where each infected person infects a fixed number of other people. With social distancing, that number should decrease.³

It took me maybe 20 minutes to assemble what I needed, including finding an old graphing utility I had laying around, and also writing some new but simple utility functions. I keep mentioning speed because: (1) People who don’t program don’t know how easily such things can be done. (2) I would not do these things with programming if I knew a faster way to do them. I have no ideological commitment to programming just because it can be done. I do it if and when it’s the fastest, easiest way I know to solve the problems I have.

I’ll cut to the chase, and then backfill details. My very first graph, aimed to compare infection rates in California, compared to the US as a whole, appears below. The graph starts on March 18—two days after social distancing was instituted—and continues until March 29 of 2020.

![Graph of US vs. CA infections](image)

This is a plot of log values of the data, since exponentials (which is the form of unfettered, constant infection rate growth) then appear as straight lines. The key lesson here is that, at the beginning of this time period, the slope of CA infections is significantly less than the US as a whole. That’s what I expected to see. Selfishly, it’s what I hoped for. To avoid confusion, note that I scaled the CA data so that it precisely matched the US at the beginning, just so that I could easily compare slope, there. (Rescaling and then applying the log function just provides a constant vertical shift; log ay = log a + log y.)

³ Down the road, I expect to look for another inflection point, where the number of cases ceases increasing.
But there are interesting complications. While the blue (CA) graph is fairly straight, with the possible exception of a small downturn in the final two days, a colleague pointed out that it is a little jagged. It turns out that the jagginess is too much to be random error (1 / root N). But, a day later I read in the paper that California had been having trouble collecting data. Apparently, some counties were not reporting, or not reporting in a form that could be imported into the state’s data bases. That fact may resolve the puzzle of jagginess. But, it’s also true that the graph is as good as the data. I used: https://ncov2019.live. Finally, I could not get to the site at exactly the same time of day each day, and the data was continuously updating.

Another observation is the arching curve in the US data. I still don’t know why that is so. But, then, that was not an interesting point for me.

You can probably see that toward later times, the two graphs appear to be more parallel, signaling a similar infection rate. In order to make this clearer, I just re-scaled the CA data so that it matched the US data later in the graph.

In this form, it is clear that for a while the US and CA had the same slope (infection rate). But now, although both CA and US appear to have tilted slightly lower (smaller infection rates) in just the last couple of days, it appears that CA is beginning to show a lower, better rate. This is what I was hoping to see, even if I expected more than this little change. On the other hand, a change in infection rate should show up after about two weeks, so we are just edging into that regime. Yesterday, the day after I noticed the slight downturn in infection rates, I saw an article in the San Francisco Chronicle, entitled “Coronavirus slowing in Bay Area? Experts track data to see whether shelter in place is working.” The article said:
By day’s end Monday [tomorrow], most of the Bay Area will have been holed up in their homes for two weeks — long enough, experts say, to see whether the unprecedented efforts to keep people apart are beginning to halt, or at least slow down, the coronavirus.

Yes, but you don’t have to be an expert, if you’re computationally literate! And I am learning so much more about COVID-19 and its tracking by doing it myself. Furthermore, I could not find any online tools to do what I wanted, much less the particular graphs I wanted. I could not even find any historical data listing, so that I have had to enter each day’s data by hand and keep my own historical dataset. These failures of the on-line world to give me what I wanted are cultural failures of our society with respect to supporting widespread computational literacy. Very few expect the public, now, to have any use for bare data, nor certainly the capability to do their own analysis of it.

What might my vignette have to do with education, other than suggesting that we should work to develop a more computationally literate public? I told someone that I could easily imagine working with a group of high school students developing hypotheses, tools, and analytical techniques like this—and in real time as the pandemic develops. I’m missing a wonderful opportunity (for not having a high school class to work with, just now). I said I was sure high schoolers could develop original hypotheses and appropriate analytical methods.

But, then, it occurred to me that we had already done something very like this, in a more difficult (if less compelling) case! We asked students to take data on the heating/cooling of two objects, at different temperatures, in thermal contact. We looked at the graphs, thought about how and why that happened (it’s exponential decay; instead of change being proportional to amount (infections), change in temperature is proportional to the difference of temperatures). We then collaboratively developed a program embodying their model. In diSessa (2017) you can read about how one group of high school students developed, on their own and with no instruction, a normative model of temperature equilibration. In diSessa (2008) you can read several cases of students building conceptual and computational models of fundamental scientific principles, including early versions of our temperature equilibration curriculum unit. Indeed, in a later edition of the same project, we did teach eight grade students (from a marginalized, immigrant population) how to think about exponential growth (in the form of spreading rumors; “each one tells two”).

The Nitty Gritty of Programming

Finally, I want to further demystify the work I did building my little COVID exploration microworld. Just below is the database for California, as it exists today. The columns are, in the sequence specified by the Key: date, cases of infection, deaths, and the number of people who recovered. The “database” is just text typed or pasted into a “box.” “X” represents missing data, all in the category of recovered cases, which I found for the US but never managed to find for CA.
Next, I’ll show the complete code for drawing one graph, revealed in stages. The first panel, below, shows the top level. Just plot a certain set of data points, which appears here as a black box.

\[
\text{plot-data}
\]

The next panel shows the black box opened (just click on it) to reveal that what’s plotted are the log values of each element of another black box of data.

\[
\text{plot-data} \quad \text{log-all}
\]

The third panel shows that black box opened up, revealing a data set consisting of yet another (black box) dataset, but scaled by a factor of 11.5. That happens to be exactly the factor that I needed to scale CA data in order to match with US data right at the beginning of my graphs.

\[
\text{plot-data} \quad \text{log-all} \quad \text{scale-all 11.5}
\]

Finally, with everything revealed, you can see that the input to the whole process is the second column of the CA COVID database, which is the number of reported infections.

\[
\text{plot-data} \quad \text{log-all} \quad \text{scale-all 11.5} \quad \text{column 2 CA_COVID}
\]

That’s the whole thing. The program to draw my COVID graphs is four commands in a nested sequence. Plot-data is a command the graphing utility understands. Log-all and scale-all are tiny programs I wrote to apply the named function to all the elements of a list of numbers. Column is an in-built primitive function of the system.

\textbf{Coda}

So, what’s the world like when every citizen can program to the (very modest) level involved in this example? How will citizens then relate to the data-filled world in which they find themselves? What will schools be like? How will mathematics and science be taught differently? What different topics will be covered, how will basic conceptions of math and science change, and what different kinds of activities will students be engaged in—such as real-world data
inquiries and modeling important scientific phenomena? That’s computational literacy. You can read some of my own expectations and hopes in diSessa (2000) and diSessa (2018).

Acknowledgments

I thank Yeping Li, Geoff Saxe, and Melinda diSessa for helpful comments on earlier drafts.

References

diSessa, A. A. (2000). Changing Minds: Computers, Learning, and Literacy. Cambridge, MA: MIT Press.

diSessa, A. A. (2008). Can students re-invent fundamental scientific principles?: Evaluating the promise of new-media literacies. In T. Willoughby, & E. Wood (Eds.), Children’s learning in a digital world (pp. 218-248). Oxford, UK: Blackwell Publishing.

diSessa, A. A. (2017). Conceptual change in a microcosm: Comparative analysis of a learning event. Human Development, 60(1), 1-37. doi: 10.1159/000469693

diSessa, A. A. (2018). Computational literacy and “The Big Picture” concerning computers in mathematics education. Mathematical Thinking and Learning, 20(1), 3-31. (Special issue on “Computational Thinking and Mathematics Learning.”) doi: 10.1080/10986065.2018.1403544