Video: “MATLABAmathematician” (4:53)

Video: (00:00)
Many funny stories sometimes begin something like “A mathematician, a statistician, and a computer scientist go into a bar...” Well, this isn’t one of those. Sorry. It would better be titled “How a mathematician, a statistician, and a computer scientist might think about data,” but you probably want to watch that movie.

(00:22)
Let’s start with mathematicians. And let me say that many mathematicians aren’t any better at balancing their checkbooks than the rest of us and they probably let the bank do it anyway. Mathematicians tend to think about structure. They develop languages or notations for describing structure, and this is probably what puts people off of mathematics - sometimes the notation looks intimidating. Once they have a framework for thinking about a structure, they try to understand structure properties and look for assumptions that make certain properties true. Let’s look at an example to make this more concrete.

(1:00)
Most of us wouldn’t have a problem accepting that 1+2 is the same as 2+1, but the idea that the operands can come in any order is very general, and called “community.” Why would data analysts care about commutivity? Turns out, we use it a lot. For example, we often add tables or arrays together. Would it matter if we added mumps to measles or measles to mumps first? A mathematician might ask whether an array addition is commutative, or wonder under what circumstances addition makes sense for arrays at all - for example, do the arrays have to be the same size? They might also ask what other operations we can apply to arrays and how these operations behave.

(1:45)
Let’s look at a more abstract example. You probably spent a lot of time in high school algebra being convinced that the words “change the sign of a value” could be written in functional notation as “f(x) = -x” or the words “throw away the sign” could be written in functional notation as “g(x) = |x|.” A mathematician might ask whether the order in applying these operations matters; that is, are the operations commutative, and they clearly are not. Why do we care? When we program, we’re constantly making decisions about what order to program operations, and if order matters and we get it wrong our code won’t work. For example, do we get different answers when we flip an array before adding or add before flipping? The point here is that functions have structure just like numbers, and we need to understand that structure in order to code correctly. From a mathematical point of view, applying one function and then another is really an operation on functions called “function composition.” We can let mathematicians worry

about it, but we do have to understand the results, and for many operations on arrays, order matters.

(3:01)
Now let's talk a little bit about statisticians. Statistics is often called the "data science." Statisticians figure out how to collect and interpret data. We can't measure everything, so you have to sample, and they figure out how many data points are necessary, and how that data should be collected and checked so we get a good answer. Statisticians often develop models for data so they can do predictions and assess error. A very common model is the linear model we worked on in class. The bottom line here is that statisticians are trying to figure out if what we're measuring is worth anything. They use tools from mathematics and computing to do this.

(3:42)
Now let's give the computer scientists a chance. A computer scientist might ask what the best way to represent the data for a computation is. We've been mostly representing our data as arrays, but sometimes a more complicated representation is better. A computer scientist also understands how to implement complicated operations, and whether those results are correct or not, or how long those results will take to compute. If we get a million times more data points, will the computation still work? And a computer scientist has to worry about how to store the results and get the information in and out of the computer for analysis.

(4:20)
Well, there you have it! Data analysis is truly an interdisciplinary activity, and mathematicians are increasingly turning to computation as part of their work. Similarly, computer scientists must understand the mathematical and statistical foundations of the methods they're implementing, or the code just won't work. Finally, data is everywhere. It pervades every aspect of our society, and if each of us does not have some idea of how data is analyzed, we cannot be informed citizens.