Want to Make It as a Biologist? Better Learn to Code | WIRED
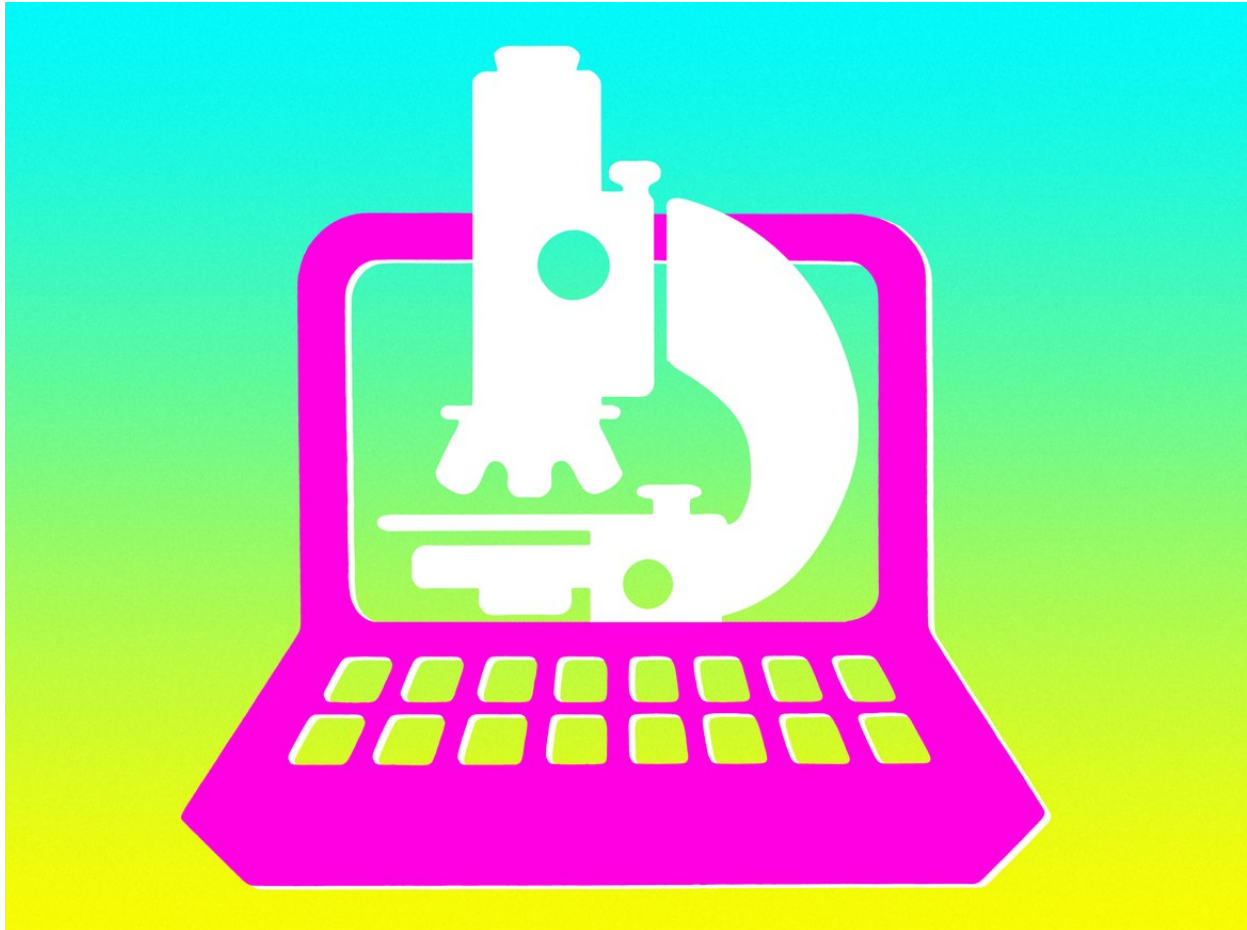
10/29/17, 6:36 AM

Want to Make It as a Biologist? Better Learn to Code

Evan Mills

Namrata Udeshi knows how to globally analyze [the proteomics](#) of human cells. You'd be forgiven for having no idea what that means or why it matters—it's a complicated technique that you'd need years of post-graduate training to master. But for now, just know it's important for disease research. Udeshi is a group leader in a proteomics lab at MIT's Broad Institute, working long days to understand the intricacies of cellular life. She's also the mother of two toddlers, with almost no free time.

And yet, every day, she spends hours learning the programming language

Python.

"Ever since I started my post-doc, I realized that it would be great to get data analysis automated," says Udeshi. "But I didn't know how to program, so I would go and find someone who knew and ask them for help." That was annoying and limiting. Now, she's enrolled in an intro to programming class through Harvard Extension School. Udeshi is hardly alone: When I asked a handful of post-doc biologists eating brunch in Boston last week how many were teaching themselves to code, every hand went up. They all realized that their curriculum was missing a core element, and they've set about rectifying the omission—on their own.

It's surprising that it's come to this. In biology, big data is the thing. Every day, biologists go into the lab to coax data out of living matter—more and more data, with the advent of biological tools like Crispr/Cas9. Udeshi used to be able to trace her data in Excel, but in the past five years, those data sets have gotten bigger and bigger. "We cannot manually look through 15,000 data points anymore," she says. To analyze it all, biologists need to write programs specifically tailored for their experiments.

Graduate programs realize that computer scientists aren't the only ones who need computational skills, and they're correcting the issue—slowly. Since 2015, the National Institute of Health has been [pushing](#) to add skills training, including coding, to biomedical graduate training, though it hasn't yet reorganized its grant priorities to require these skills. Outside of specialized computational biology and bioinformatics programs, most basic biological graduate programs don't require coding classes.

At UCSF, newly minted department head Anatol Kreitzer is trying to revamp the curriculum for neuroscience grad students. "Our curriculum is 30, 40 years old," he says—it requires some statistics and lots of speciality

neurobiology, but no coding. One of Kreitzer's first actions as department head was to assemble a committee to figure out the best way to incorporate coding into the neuroscience program's core curriculum. It might take a while, but it's a start.

On Their Own

In the meantime, working scientists who need to know this skill now turn to books, online courses, and night classes. And mostly, to each other.

Udeshi chose to take a formal course. Sam Myers, a bio-analytical chemist in Udeshi's lab, is teaching himself R by simply "Googling everything." Taking an online course is the middle ground option.

Adam Granger, who graduated from UCSF's neuroscience department three years before Kreitzer took over, would have jumped at the chance to learn coding while he was earning his PhD. Instead, he enrolled a few months ago in an online Python class through the website Code Academy. When he leaves his bench at Harvard, where he's a post-doc in electrophysiology, he opens his laptop at home and goes into a coding vortex. Arpiar Saunders, a genetics post-doc at Harvard, did the same when he learned the language R, though he took a class offered by competing site Code Camp.

Beyond the basics, all of them end up relying on an informal apprenticeship within their labs. Whoever knows the secrets of coding becomes the wizened elder who schools the younger folk—except often the age dynamic is reversed.

"It must be a huge pain in the ass for the coding experts in the labs," says Saunders. When he first started his neuroscience PhD program years ago, he improbably became that person—simply because he'd bought a book on the language Perl over the summer and taught himself the syntax. People in the lab treated him like the expert. "And I'm not a good programmer. I am a

barely proficient programmer," he says.

When Saunders became a post-doc, he found an actual expert to help him. "I realized that just the way he held his laptop was completely different from me. His fingers were spread wide open over the keys in this diagonal format, and I just knew I'm fucked, I'm fucked in this whole field," Saunders says. "I type like an old person. These kids, they interact with their computers in a completely different way." Saunders is in his early 30s.

But he's right that this problem is generational. People getting a PhD in neuroscience from Harvard now can take a bootcamp in MatLab in their first year—though it's still optional. As these biologists can attest, it shouldn't be. Not only is coding a core skill that gets the basic work of biology done, it's also taught them to look at problems in new ways. Above all, they agree, coding liberated them.

As tools evolve to allow biologists to gather ever-more-massive quantities of data, people like Kreitzer will find a way to make coding a core part of scientific education. Until then, the biologists will have to go it alone.