

[CHAPTER 1 - INTRODUCTION: PIZZA DATA]

Imagine you're conducting a survey in the school cafeteria, asking questions and taking detailed notes about how people like their pizzas. We're talking favorite toppings, what kind of crust, and the ultimate question: do you eat it folded or flat.

Personally, I prefer it on a bagel, cause then you can eat pizza anytime.

By the end of this survey, you'll probably have gotten some weird looks, but at least you'll have a big pile of data. On its own, your notebook of pizza preferences and tally marks isn't gonna tell you much. To make sense of those numbers, you'll have to analyze them.

This could be as simple as counting up your numbers and learning that more people are rockin' thin crust than deep dish. But what if you wanted to answer a more complicated question, like how someone's stress levels affect how much or how little they eat? Or, what if you wanted to take your survey further and analyze the pizza habits of the whole school district? You're gonna need more than tally marks in a notebook.

Biologists, and really most scientists, use computers to analyze data all the time – and for way more than planning pizza parties.

Hi, I'm Dr. Sammy, your friendly neighborhood entomologist, and this is Crash Course Biology. Hey Callie, could you serve me up a pizza that theme music?

[THEME MUSIC]

[CHAPTER 2 - BIOINFORMATICS]

When biologists use computer science to analyze and look for patterns in data about living things, as in our pizza survey, we call it Bioinformatics.

Bioinformatics can help scientists sort through data about everything from DNA to weather conditions, to the number of organisms on a beach three Tuesdays ago — all information that can be useful for different areas of research and conservation.

You start with a question or problem. It can be as simple as pizza preferences, or more complex, like, are two different species of fish experiencing different amounts of stress?

Then, you get a collection of numbers or a dataset. It could be an enormous one showing, say, the stress hormones of every species of fish in the African great lakes. Or, it could be a much smaller dataset about the five jellies in your home aquarium.

How ya feeling, buddies? Oh, they're not...they're not real. We got ten episodes left and now y'all tell me.

Why have I been feeding them?!

Ahem, anyway. So, the data could be something you and your team personally collected, or it could be from a larger, public dataset. The bioinformatics community is pretty open; they typically love to share information.

In fact, scientists are often supported by data that ordinary people have collected through community science projects to help fill in gaps. All that collected data can be analyzed to help answer the original question.

[CHAPTER 3 - ALGORITHMS]

To find those answers, you take that data, and you use a specific set of instructions to analyze it, called an algorithm. In many cases, you communicate those instructions to a computer, which uses the algorithm to solve a problem.

You've probably run into your fair share of algorithms around the internet, like on TikTok or on YouTube. Like, if this video appeared on your YouTube homepage, it wasn't a random action. It's because an algorithm analyzed data about other videos you've watched and predicted that you might like this one, too.

In this case, the "instructions" are to learn your viewing habits, and the "problem" is, well, how to keep you there as long as it can.

So be sure to smash that subscribe button!

Sorry...we're obligated to say it that way at least once per series. Could you –uh–could you also click the little bell too?

But algorithms can be fed all kinds of instructions that help us learn some pretty amazing information about living organisms, the environment, or even the patterns of disease spread.

[CHAPTER 4 - THE HUMAN GENETIC CODE]

And bioinformatics can help us in nearly every stage of the scientific process. Like, let's start with acquiring the data itself. For example, if we want to know the entire genetic sequence of an organism, we can't just read it from one end to the other. Because it's so long, we can only get it in small, overlapping chunks.

But we can then use computers to organize those chunks and piece them together in the correct order, like assembling a jigsaw puzzle but in the end, you get a complete dataset – and unlike a puzzle, your cat can't knock that off a table.

Using that method, in 2022, researchers officially completed one of the biggest scientific undertakings in human history. For the first time, we'd assembled an almost complete list of every letter in human DNA.

They'd been working on this for decades, and it was like they'd finally unlocked the instruction manual that made humans tick. The possibilities for medicine seemed endless! Except, there was a catch. No one was fluent in the language this instruction manual was written in. Turns out, it's one thing to know an organism's DNA sequence and a whole other thing to figure out what those building blocks do.

And this isn't exactly a code you can crack on paper. Because if you were to print out the roughly 3 billion letters stored in your DNA, they would fill hundreds of thousands of pages. And cracking the code requires comparing the DNA sequences of many individuals. We're talking an absolute ocean of data.

Enter bioinformatics, again! Scientists can use computers to compare a bunch of DNA sequences until a pattern emerges that lets us predict how certain genes are working in an organism's DNA.

[CHAPTER 5 - THE BRCA1 GENE]

And these patterns can help us understand and treat diseases, including cancer. For example, there's this gene called BRCA1, which codes for a protein that keeps cells growing normally, suppressing tumors and potentially cancerous masses. Which is a pretty big job! In fact, if even a single letter in someone's BRCA1 gene is off, it can lead to an increased risk of breast or ovarian cancer.

Overall, variations in this gene are linked to a chance of developing one of these cancers by age seventy. But not every variation carries the same risk. So, some scientists have turned to bioinformatics to try to understand how dangerous each of the gene's mutations are.

They compared more than 3,600 BRCA1 variants — trying to figure out what the gene variants look like, what they do to the tumor-suppressing protein, and what overall effects that might have in the body.

Scientists then compiled all of this annotated data into a bioinformatics tool for others to use, and the information it contains helps doctors make more informed treatment plans, which could mean better outcomes for the patient overall.

[CHAPTER 6 - TRANSCRIPTOMES]

Bioinformatics can also help us sort through which genes are being expressed, or turned on, in an organism. You see, while some genes may be present, only the ones that are expressed result in protein production that impacts the organism.

To figure out which genes are being expressed, we can look at all the RNA molecules in an organism's cells at any given time, called the transcriptome. This is an indicator of which genes are being transcribed and used to make proteins.

As with DNA, it would be impossible for a human to analyze and compare organisms' transcriptomes manually because they're so complex. But with bioinformatics, we can figure out which genes are being expressed to make certain traits happen. And if those traits are beneficial — like say, disease resistance in crops — we can learn how to use genetic modification techniques to make other organisms similarly resistant.

Bioinformatics has super broad applications, in nearly all areas of biology. For example, it can help us gain a better understanding of evolution. By tracking how genes differ among closely-related organisms — like different species of birds or primates — we can gain insight into how these organisms might have evolved.

[CHAPTER 7 - THE ZIKA VIRUS]

And, we can study random changes in organisms' DNA, called mutations. Which not only helps us understand species evolution better; it can also help us see how viruses spread through a population. And that can be lifesaving information.

For instance, in 2016, the mosquito-borne Zika virus swept through the Brazilian state of Minas Gerais. And although Zika virus disease usually has mild or nonexistent symptoms, it can lead to problems with brain development in newborns or neurological diseases, like Guillain-Barre syndrome, where the body attacks its own nerves, in adults.

So, an international team came together to use bioinformatics to try to understand how this virus got to Minas Gerais and how it evolved over the course of the epidemic.

The researchers took samples from patients with Zika virus disease. And then, they did something pretty amazing: Right on the spot, they were able to analyze the DNA in those samples using handheld nanopore technologies. Which is about as sci-fi as it sounds.

These are essentially tiny computers that let scientists sequence—or figure out the precise makeup of— DNA and RNA, wherever they are. There's no need to ship anything to the lab — you just feed the device a few drops of sample and hit go. Then, those sequences were uploaded to a larger database, and compared with each other to build a picture of how they were related, a little like a virus family tree.

Ultimately, this DNA analysis told the team that the virus had been circulating in Minas Gerais for at least sixteen months before it was confirmed in the lab. And by gaining a deeper understanding of the outbreak's journey and timeline, researchers were better equipped to slow the disease's spread.

[CHAPTER 8 - BIOINFORMATICS & PROGRAMMING]

No matter what you're studying, with bioinformatics, the computer is doing a lot of the math. So while it's important to understand what an algorithm is trying to do, you don't necessarily have to be an expert in math or programming to use one.

That said, despite being able to sort through piles of data faster than you can say "bioinformatics," algorithms aren't all-knowing overlords.

You see, algorithms work precisely because programmers give them limits and assumptions, which are set values that the system assumes are true.

For instance, if we were to create one for our pizza survey, we might give it the assumption that anyone who didn't like their pizza folded ate it flat. This helps it make calculations more quickly because now it knows that any answer that doesn't fit the "folded" value must be "flat."

Limits and assumptions are really important, and if you don't know what they are, your program might give you misleading results, or might not work at all.

Like, say I was sorting through that pizza database, and I asked a computer to give me all the data about Deep Dish pizza. If it was programmed to limit results to just data about topping preferences, it might throw me an error. "Sorry, Dr. Sammy. Deep dish pizza doesn't exist." But that's not true, I'm just looking in the wrong data set.

[CHAPTER 9 - REVIEW & CREDITS]

So, just like a lab coat, a clipboard, and a microscope – computers and algorithms are tools in the biologist's utility belt. They can be used to analyze huge sets of information that would take us humans many lifetimes to work through on paper.

Thanks to bioinformatics, the fields of biology have advanced by leaps and bounds, and continue to grow as engineers develop better computers, and programmers build better algorithms that allow biologists to ask more and more complex and fascinating questions. And speaking of complex questions, next time we're going to answer a weird one: why aren't we made up of just one big cell? I'll see ya then! Deuces!

This series was produced in collaboration with HHMI BioInteractive. If you're an educator, visit [BioInteractive.org/CrashCourse](https://www.biointeractive.org/CrashCourse) for classroom resources and professional development related to the topics covered in this course.

Thanks for watching this episode of Crash Course Biology which was filmed at our studio in Indianapolis, Indiana, and was made with the help of all these nice people. If you want to help keep Crash Course free for everyone, forever, you can join our community on Patreon.