

# THE NUMBER CRUNCHER

Data scientist Rayid Ghani, C'99, leaves the corporate and political worlds behind to focus on using big data for the public good.

By Rod O'Connor  
Photos by Matthew Gilson



**R**ayid Ghani doesn't want to discuss what happened in Charlotte last night. It's a misty September morning, and over Ghani's shoulder, the window in his small office provides a perfectly framed view of the ivy-covered Gothic stone buildings of the University of Chicago campus.

As director of the prestigious institution's Center for Data Science & Public Policy, he's been working with the Charlotte Police Department to harness the power of big data to develop an early warning system to try to predict which cops might engage in an inappropriate use of force. Unfortunately, his algorithms couldn't prevent last evening's bloodshed, as one of the department's white officers shot and killed an African American man, Keith Lamont Scott, who may or may not have been holding a gun.

"I don't have enough information to comment," he says politely, when I bring up the shooting as I settle in across his remarkably uncluttered desk. Today is Ghani's first day back on campus for the fall semester—he spends his

summers leading a fellowship program that meets in a downtown Chicago office—so the piles of paper and reams of research have not yet accumulated around him. There's a copy of *Police Chief Magazine* off to the side. One shelf behind him is filled with books on probability, statistics, and machine learning, including a copy of the *Perl Cookbook*, which isn't a cookbook at all, but rather, a collection of computer programming "recipes." On the wall is a whiteboard with the fading remnants of last spring's scribbled-out equations, soon to be replaced with new ones.

searches, and arrests, Ghani created a prototype program to try to flag those cops most likely to snap and, possibly, use lethal force. In turn, the department had initiated changes in terms of how they responded to calls and the types of support provided to officers involved in emotionally charged events.

As a self-proclaimed "reformed computer scientist turned wannabe social scientist," Ghani is anxious to get his hands on more data about the latest in a rash of deadly interactions around the country. For example: Did the cop in question have a history of unjustified

apples, although those cops do exist," he explains. "It's that they don't have the right support programs in place."

Helping curb police misconduct is just one example of how Ghani, a tirelessly talented Pakistan-born Sewanee alumnus, is collaborating with nonprofits and government agencies to use computer science to take on society's ills.

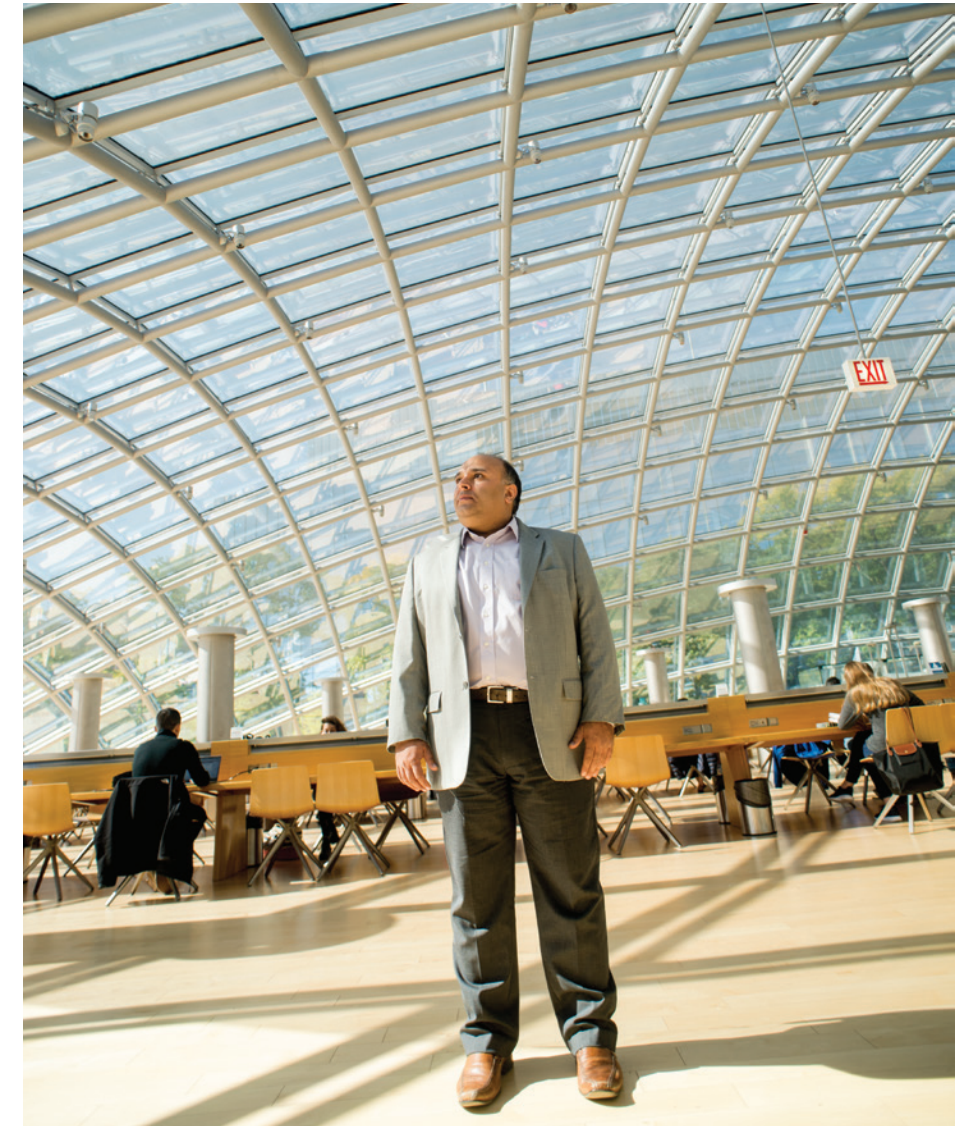
"At a high level, I help organizations use data to improve what they do," says Ghani, who previously served as the chief data scientist for President

things to improve lives for people. And they mean well," he says. "But they don't have the right infrastructure or the right tools. Some of the things they don't use often are evidence and data."

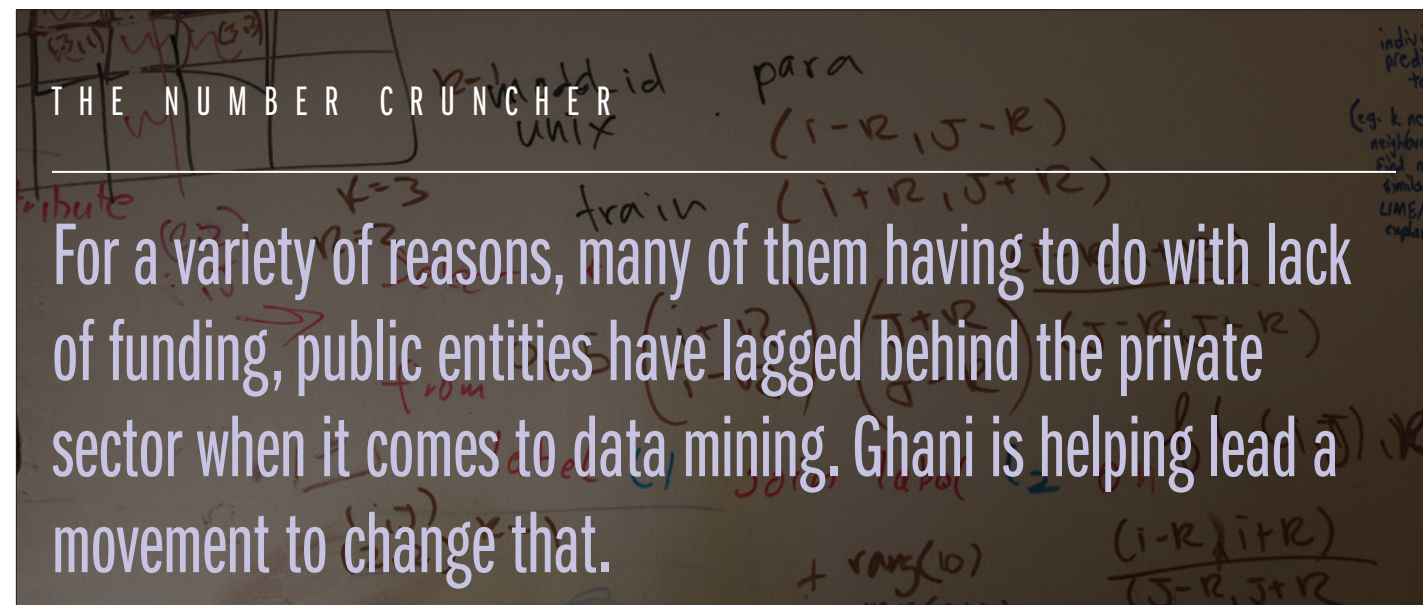
Most of Ghani's projects focus on establishing early-warning systems to avoid the types of things communities don't want to happen. For example, scouring data for clues that could help pinpoint those grammar-school kids who might not graduate from high school. Much like the signs that show which stressed-out cops could be bubbling under the surface, ready to snap, Ghani found that looking at things like test scores and whether they participated in afterschool activities could help identify at-risk students as early as the third grade—when there is still a chance to steer them on the right track. Other projects try to solve civic infrastructure challenges, like the predictive models he is currently using to help officials in Syracuse, New York, detect signs of a water main break before it happens.

"What [most governmental agencies] do now is, they wait for something to break," says Ghani. "If someone complains, they go and fix it. But with early-warning signs, we can predict things early—and fix them earlier. And prevention tends to be cheaper and less damaging, obviously."

While it's easy to paint Ghani as a compulsive do-gooder, the truth is, he hasn't always been so altruistic. From 2001 to 2011, he worked as senior research scientist and director of the analytics group for the consulting company Accenture. There, from a bustling technology lab in Chicago, he led a team that used data mining to support a long list of blue-chip companies, from Boeing and Best Buy to Pfizer and Ford. Most of his efforts focused on helping brands personalize their ad campaigns and predict consumer behavior. Put another way: He helped companies sell things. Among the successes attributed to him during this chapter of his career:



Rayid Ghani says he was initially drawn to work in the corporate world not because he cared about increasing companies' profits but because they had large amounts of data for him to play with.



Ghani, whose thinning dark hair is graying at the temples, looks both professional and professorial in his crisp white oxford shirt and dark slacks. He seems troubled by the incident that occurred in North Carolina, but his tone is also straightforwardly objective. For the past two years, Ghani and members of his 14-person staff have been wading chest-deep in a massive pool of data provided by Charlotte's police force, which agreed to hand over the information after meeting him at a Washington, D.C., law-enforcement summit. After digging through the data associated with 15 years' worth of officer stops,

force? Were there previous citizen complaints against him? But he also wants to know more about the officer himself. Was he frazzled because he was at the end of a long shift? What was going on in his personal life? Could some sort of intervention from a peer or supervisor have been the difference between him pulling and not pulling the trigger?

According to Ghani, it's this nexus between empirical data—and the people behind the numbers—that could be the difference-maker when it comes to curbing police abuses.

"The real problem in police departments isn't necessarily that there are bad

Obama's 2012 re-election campaign, when I ask him to summarize his job. "I used to do it for organizations that aren't doing anything particularly useful for most people. Now, I'm trying to do it for groups that I care more about."

For a variety of reasons, many of them having to do with lack of funding, public entities have lagged behind the private sector when it comes to data mining, a computer science term for finding patterns and other useful information within large amounts of data. Ghani, C'99, is helping lead a movement to change that. "Most government agencies and nonprofits are trying to do

He figured out a way to predict the final price of an eBay auction with 96 percent accuracy.

But he's quick to point out that he was first drawn to the corporate world not because he particularly cared about companies' goals to increase profits or raise shareholder value. They simply had large amounts of data for him to play with.

"The reason I got interested in data and computer science and artificial intelligence was the intellectual challenge of solving the puzzle," he explains. "And wherever there was access to problems and data and people willing to try out

what I do, that's what I was attracted to. At some point that changed. It wasn't enough to solve a problem because I could. I wanted to do something more useful."

In 2011, he decided he would quit his job and try to figure out how to apply his skills for good. The plan was to take some time off to ponder his future. But then, on his last day at Accenture, a representative from President Obama's re-election campaign called with an offer that would shape it for him: Would he like to come on board as the chief data scientist for Obama 2012?

"I had no idea what that meant," he



Ghani's work on President Obama's 2012 re-election campaign made him a sought-out numbers guru just as recognition of the value of data mining was becoming widespread.

says, laughing. "I had no idea how they use data. But I couldn't think of anything potentially more impactful." So he joined the team and spent the next year and a half in campaign mode, working 24/7 to figure out how analytics could help recruit volunteers, mobilize voters, and target those most likely to write donation checks.

"The biggest use of data was in tar-

getting, where we could use information from social media and other sources to identify those who were supporting President Obama, but needed a nudge [to actually get to the voting booth]," he explains, as he waits for his cappuccino at a bustling café on the University of Chicago campus, surrounded by the sort of millennial voters he targeted so successfully during that election cycle. "Or

empirical analysis of baseball statistics) into the national vernacular. The statistician Nate Silver, who used computer models to predict 49 of 50 states in the 2008 presidential election, had already become a data-wonk rock star. Unmarried and 35 years old, Ghani had the flexibility and reputation to do almost anything he wanted.

Suddenly, he was fielding calls

those who definitely planned to vote, but were on the fence about who to vote for.

"For the first category, we would tell volunteers to talk with them about getting to the polls," he continues. "For the second group, we tried to persuade them to support us. So, depending on who they are, we can target them in a certain way." The campaign did the same things with fundraising. We used targeting to determine who we could we ask for money, and how much money should we ask them for. Because if we ask for \$20,000, you may go, 'eh, no thanks.' But maybe \$3 is the right ask."

So, does he think his efforts helped get President Obama a second term? "That's one question that I get asked a lot: How much difference did the data make?" he says. "And the tricky part is, I don't know. But I do know this: Data won't make up for a bad candidate."

One thing that is known: His role on the Obama campaign got him national recognition as a numbers guru. And the timing couldn't have been better. The Brad Pitt movie *Moneyball* had come out the year before, inserting the term "sabermetrics" (the

from everyone from the NFL's Miami Dolphins (who wanted him to coax a winning team out of their player data) to major corporations with tempting offers to return to the for-profit world.

"I was back in the same place as I was before: How do I do what I want to do [use data to solve societal problems] longer-term? But now I had at least a rough idea of what could be done. I spent a few months catching up on sleep and talking with governments, nonprofits, companies, foundations. I said, 'I'm

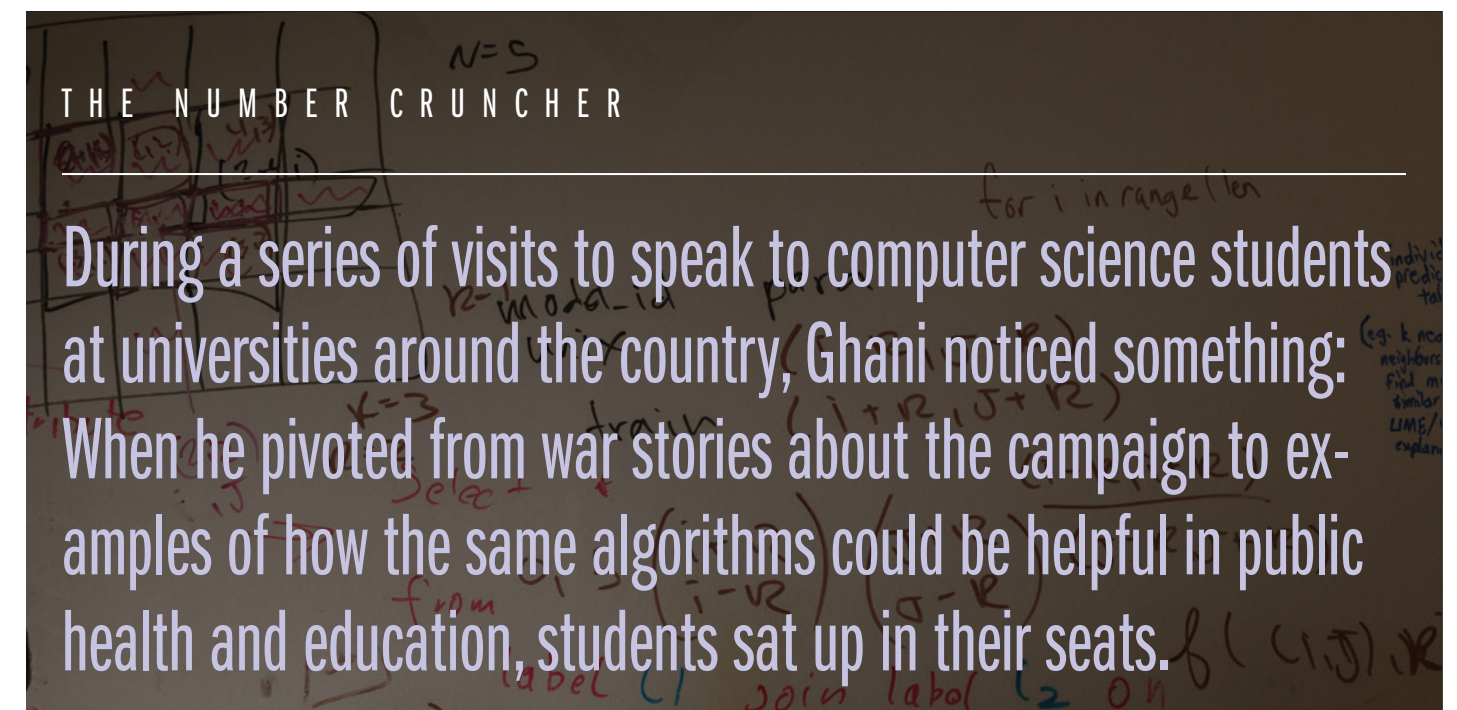
that Sewanee taught me. It's something I learned here.'

"And now, his follow-up hasn't been to stay in politics, but to try to find ways to benefit humankind. He saw new ways to use these large collections of data for good, which is really amazing."

**W**hen Rayid Ghani arrived in Sewanee from Karachi, Pakistan, back in 1995, he wasn't sure what to make of it. In fact, he acknowledges

ries: When he first got to his dorm, an administrator had left a message on his answering machine. Ghani could barely understand the slow-as-molasses delivery and thick Southern drawl. "I was like, why isn't he speaking faster?" he chuckles. "Doesn't he know that I speak English?"

Ghani acclimated quickly to his academic surroundings and believes that living in Sewanee's self-contained world allowed him to focus on his studies without having to concern himself



going to talk to anyone and everyone, because I can."

"He had come back to visit about a year before the election, and he told me about this possibility [of working on the Obama campaign]," recalls Linda Lankewicz, his former computer sciences professor at Sewanee. "He had a very good career with a large company, and you don't always step away from that. But he said, 'I don't know where it will lead, but I think I can make a difference. I'm going to jump off that cliff.' And I remember he said, 'That's something

that he didn't really know what a liberal arts curriculum was. His schooling in Pakistan focused almost exclusively on science and math. The primary reason he applied to Sewanee, among many other schools, was because he saw it had cracked the top 25 in *U.S. News and World Report*.

His grades resulted in several acceptance letters. But Sewanee was the only school that offered him a full scholarship. "So, I was like, OK. That's where I'm going," he says.

One of his first campus memo-

with also acclimating to an entirely new country. "If I had gone somewhere else, like, Chicago, I would have to figure out Chicago, the U.S., and college at the same time. [At Sewanee], I didn't have to figure out transit. I just had to walk from my dorm to my classes. I didn't have to learn about the food, because I had a meal plan. I think a lot of those things were kind of nice."

What he didn't care for, at least initially, were classes in a broad range of fields. "I registered my first semester for math, physics, chemistry, and computer

science. I was like, I'm not going to do liberal arts," he says. A few months in, he was convinced Sewanee was the wrong choice and he began preparations to transfer.

"I can recall him coming to my office with all these application forms he wanted to submit," says Lankewicz, who taught Ghani all four years of his undergraduate career. "He had this track that he wanted to pursue in computer science. And he realized it might be a shorter path if he could go to an engineering school."

A few things changed his mind. One was the free-thinking atmosphere that the campus provided. The other was the paid internship program that Sewanee offered, which allowed him to study with renowned researchers at more traditional data-science powerhouses. In fact, it was a guest lecture by Carnegie Mellon Professor Tom Mitchell that led Ghani to an internship and, ultimately, a 2001 master's degree in knowledge discovery and data mining from the esteemed Pittsburgh-based university.

"It's tough—you're on the other side of the globe and you are at this place [you don't really understand]," says Lankewicz. "But Sewanee looks for ways for students to flesh out opportunities based on your interests. He was able to say to a leading researcher: 'I'd like to come study there with you this summer.' Sewanee made it possible to find a path."

After a while, Ghani found himself enjoying classes in a variety of disciplines. Especially religion and art. And he savored the freedom to figure out his own path, which is something he probably would never have experienced in a more traditional engineering program. As an added bonus, he developed the communications skills that he must now use constantly to promote and push forward his data initiatives.

"The culture at Sewanee included a lot of writing and presenting," he recalls. "There were times that I'd be giving two

or three presentations every week. And I thought that was normal until I went to grad school.

"I still find it all painful, you know, dealing with people," he says. But, with his casual demeanor and easygoing personality, you'd never know it. In fact, the more I read up on the scope and ambition of his projects, the more impressed I am with his ability to combine social skills with a genius for numbers and data. Imagine if Alan Turing, the code-cracker from *The Imitation Game*, was also a hit at fundraisers and cocktail parties.

"I think the broad exposure to many subjects helped him see the application behind the science," says Lankewicz. "There are applications and patterns everywhere. But if you haven't been exposed to it, you don't think in those terms. He saw that [at Sewanee] he could have the best of both worlds."

After Ghani's success on the Obama campaign, it took a few months for the pieces to fall into place that have provided him the platform that he has now. The real light-bulb moment occurred in 2013 when, during a series of victory-lap visits to speak to computer science students at universities around the country about his work in the election, he noticed something: When he pivoted from war stories about the campaign to examples of how the same algorithms could be helpful in areas such as public health and education, students sat up in their seats.

"I expected them to say, 'Yeah, whatever,'" he recalls. "Or, there's no money. But their response was, 'Hmm, that's interesting.' [At the time], most students were only exposed to the advertising and marketing [applications of data science]. They didn't realize what they were studying could be useful for these problems.

"For me, that was the key point. It was like, OK, I want to do the work

myself, but I also want to get people to understand how to do this work and how to get excited about it. And that's what brought me to the University of Chicago. I wanted to be in those two worlds, not only teaching, but also doing the work myself and getting other people interested."

The University of Chicago hired him in 2013 to head up its new Center for Data Science and Public Policy. And he had lots of ideas. But now he needed the funding to bring them to life.

"The interesting thing about working for a university is, nobody is telling you what to do," he says. "Which is kind of good because it gives you the freedom to explore, but it's bad because you have to do it yourself. In a company, you've got big machinery behind you—and you have funding. Universities don't necessarily have a lot of cash sitting around they can dedicate to individual projects."

Fate intervened again in the form of one of President Obama's biggest Silicon Valley supporters: Eric Schmidt, then-CEO of Google (and currently executive chairman of Google's parent, Alphabet, Inc.). "He spent a lot of time with the campaign and I had got to know him. We were chitchatting about what I was going to do. And he said, 'Go figure out what you want to do and I'll fund it.'"

What sprang from that conversation was Data Science for Social Good, a grant-funded summer program from the University of Chicago that trains selected fellows how to implement the socially minded data mining that Ghani has made his life's work. For the past three years, 126 students (each summer, the program accepts 42 students from hundreds of applicants) have taken on real-world problems in areas such as education, health, public safety, economic development, and more.

By combining lectures and classroom learning with team-oriented problem solving that includes visits out in the field (whether it's ride-alongs with

police or sit-downs with government officials), the students can tackle problems using the same philosophies that Ghani has developed starting with his undergraduate years on the Mountain.

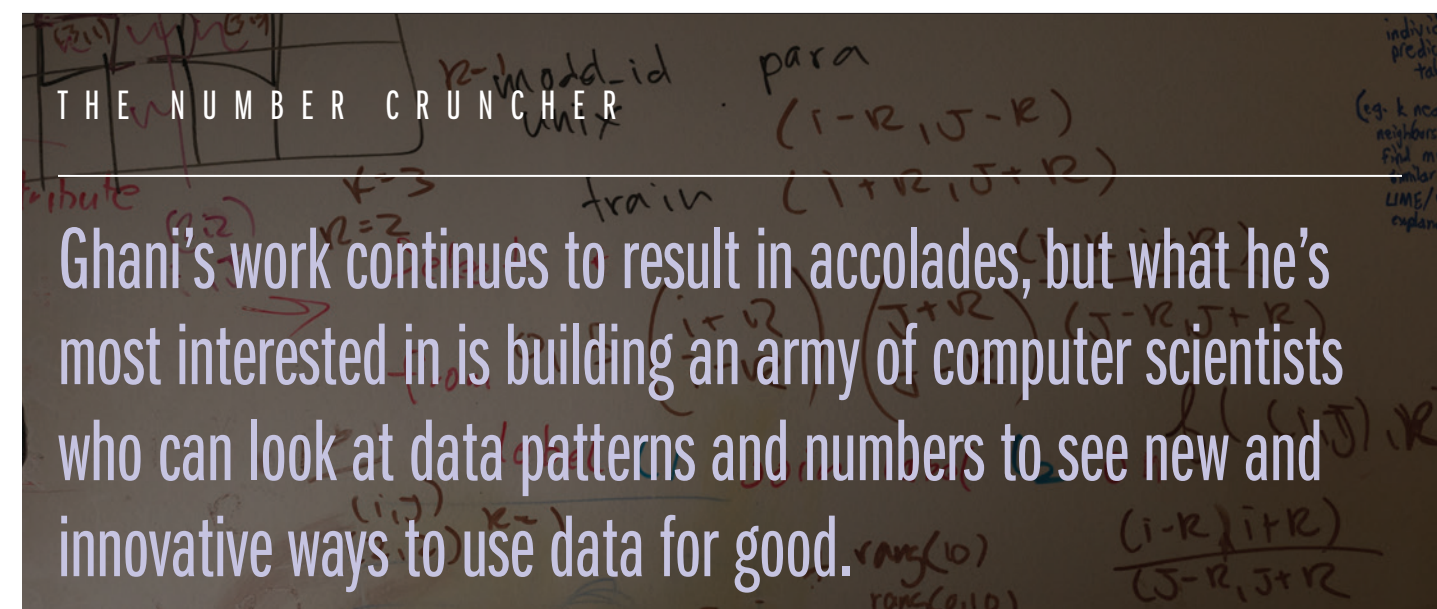
"When you work with nonprofits and governments, you have to understand that when you're making these decisions it's not just about the numbers," he says. "It's about the people behind the data points. That's something we spend a lot of time on. I always torture my students with this. It's like, OK, so who are we talking about here? Tell me about those people.

in dozens of communities nationwide, as well as inspiring similar initiatives at other universities, such as Georgia Tech and the University of Washington.

And Ghani's ever-increasing body of work continues to result in more accolades, including being named a "Young Global Leader" by the World Economic Forum in 2014. But what he's most interested in is continuing to build an army of computer scientists who can look at data patterns and numbers to see new and innovative ways to use data for good. A big part of that is the mentoring work he's done with fellow Sewanee

the City of Chicago's Department of Public Health asked if Ghani could help reduce lead poisoning in kids, which has to do with lead paint in walls, by creating computers models to find likely candidates before they get it.

Currently, kids are tested and if they have a high lead level, then a team is sent to check for lead hazards. But the fact is, the department already knew a high percentage of cases were coming from older homes and in poorer neighborhoods. The problem was that they didn't have the resources to fix every home that could be a source of lead poisoning.



"If you just look at the numbers, you might not see the biases that exist in the data," he says. "For example, if someone didn't fill out part of a questionnaire, what are some reasons that a particular person might have left that question blank? This sort of thinking allows the students to get a sense for what's happening on the ground as opposed to just sort of staying in front of a computer."

Ghani's hands-on and human-centered approach to data science is spreading, with Data Science for Social Good working on data-driven projects

computer science alumni, helping them build a bridge from the familial, liberal arts experience to top-tier science schools for their postgraduate studies.

"Before, Sewanee wouldn't have resonated as a leading computer science undergraduate college," says Lankewicz, his former professor. "Now, we have this stream of folks who have gotten advanced degrees and gone into industry following in his footsteps."

As we part ways, Ghani is visibly excited as he shares details of another project he and his various disciples have been working on. A couple of years ago,

"So, we said, 'Let's take everything we know about the homes that had exposure, and everything we know about the kids, and put it into this predictive model and see what comes out,'" he says. "By combining these factors, we were able to come up with a risk score. Now, the department can allocate a percentage of its staff to go and do these preventative inspections. It helps them prioritize."

And helping cash- and resource-strapped organizations better use their resources is perhaps the best equation Ghani could ever hope to solve.