Statistics, Computer Science Students Collaborate on Real-World Data Problems through Mini-Think Tanks

What is the difference between statistics and data science—and, perhaps more importantly, why do we have two fields with what seems to be the same focus? The best way to understand the emergence of data science as a separate discipline, explains Herman "Gene" Ray, director of the Center for Statistics and Analytical Research at Kennesaw State University, is to see data science as the merger of computer science and statistics. "Most traditional statistics programs teach you a lot of theory and how to work out problems by hand," he says. "Computer applications are something of an afterthought. But businesses aren't going to analyze 100 million records by hand; they're dealing with huge convenience samples. And that's where data science steps in."

And that's where the academic infighting starts: Statisticians say data scientists lack the statistical or mathematical foundation to understand data collection and analysis, and data scientists roll their eyes at statisticians for their lack of programming savvy. This, says Ray, was the biggest obstacle they faced in creating one of the first US PhD programs in analytics and data science: How *do* you combine statistics and computer science? "Each one thinks they can do it without the other," he says. "But the reality is that most statisticians are not very good programmers, and most computer scientists don't really understand some of the nuances of statistics. Our goal is to bridge that divide."

Their solution, in part, leveraged the increasing awareness among Atlanta-area businesses of the importance of data. The Analytics and Data Science Institute created nine sponsored research laboratories, each focused on data problems facing a business or public service or nonprofit, and each with one to four PhD students led by a faculty member. "They're like miniature think tanks exploring real-world problems," says Ray. "And in doing so, students get to understand the problem from the computer science and the statistical perspective." A more traditionally minded statistics student might be encouraged by a colleague to explore neural networks, while a more traditionally minded computer science student might be encouraged to see why they have to use representative sampling over convenience sampling.

One recent project involved working with Cobb County Fire Department, a suburb of Atlanta, which was not meeting the national metrics for fire standards. "We took all their data for fire and ambulance events—the time of the first phone call to the time the ambulance left the

firehouse to the time it took it to get to an event. We looked at the routes and traffic patterns, and then optimized response times using graft theory and Google Maps." Routes were changed, fire zones reallocated, and response times were cut. "The Cobb County fire chief is very data savvy," says Ray, "so he's implementing incremental changes and then seeing how the data updates."

The research laboratories also add another dimension—and an increasingly important one—to student experience: how to talk to people who aren't statisticians or data scientists.

"When I was trained, the expectation was that I would work with other statisticians and present at academic conferences," says Ray. "So, we all spoke the same language. Today, a data scientist could be speaking with an executive, or client, or policymaker, who has very little statistics background at all. They must be able to read this really quickly, and make sure the right message is still communicated at the appropriate level. That's one of the beautiful things about these labs—they force everyone to learn how to speak in a way for the lab to be successful."

JSM Talk:
http://ww2.amstat.org/meetings/jsm/2018/onlineprogram/ActivityDetails.cfm?SessionID=215542

For details, contact: Herman Ray, *hray8@kennesaw.edu*

About JSM 2018

JSM 2018 is the largest gathering of statisticians and data scientists in the world, taking place July 28–August 2, 2018, in Vancouver. Occurring annually since 1974, JSM is a joint effort of the American Statistical Association, International Biometric Society (ENAR and WNAR), Institute of Mathematical Statistics, Statistical Society of Canada, International Chinese Statistical Association, International Indian Statistical Association, Korean International Statistical Society, International Society for Bayesian Analysis, Royal Statistical Society and International Statistical Institute. JSM activities include oral presentations, panel sessions, poster presentations, professional development courses, an exhibit hall, a career service, society and section business meetings, committee meetings, social activities and networking opportunities.
http://ww2.amstat.org/meetings/jsm/2018/index.cfm

About the American Statistical Association

The ASA is the world's largest community of statisticians and the oldest continuously operating professional science society in the United States. Its members serve in industry, government and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare. For additional information, please visit the ASA website at *www.amstat.org*.