



STAT1ST1CAL SUCC3SS

\$1.5 MILLION
GRANT HELPS
SOPHOMORES
MAKE SENSE OF
BIG DATA



One of the first assignments in Mark Daniel Ward's Introduction to Big Data Analysis class was to analyze billions of data points from a recent data exposition: 12 gigabytes of airline data, including 29 variables (takeoffs, delays, arrivals, locations, etc.) for every airline flight in the United States from 1987 to 2008. Ward, associate professor of statistics, asked the students to communicate their results in succinct, insightful and understandable ways. The task of distilling the data to a handful of plots and descriptions was simultaneously challenging and rewarding.

The 20 undergraduate sophomores in Ward's course are the first class of students to benefit from a \$1.5 million National Science Foundation grant announced in 2013 and implemented this fall. The grant fuels the creation of a learning community where students live together on a common floor of a residence hall, participate in the same coursework, dine with faculty, attend professional development seminars, work together in teams during a new DataFest competition and conduct research in several different fields. The funding allows students like Felix Francisco-Sanchez to concentrate on classes and research without relying on part-time jobs to pay bills.

Like his 19 classmates, Francisco-Sanchez chose a lab where his statistics expertise would be valued and applicable. His

project began during summer 2014, under the direction of R. Claudio Aguilar, associate professor of biological sciences. Francisco-Sanchez will gain experience in Aguilar's lab working with big data and applying his statistical analysis talents in a life sciences field.

"I'm analyzing biology research with (programming platform) R," Francisco-Sanchez says. "I'm speeding up the process" for Aguilar's research team.

In addition to Introduction to Big Data Analysis, the 20 sophomores take Probability together. While only a few weeks in, students grouped together and compared notes during a Wednesday morning class in Stanley Coulter Hall. Large computer monitors were crammed with charts, graphs, code and extensive notes about the data.

Ward's excitement for the opportunities his students have is palpable. During class, he enthusiastically helped them finish a project while introducing the hefty airlines assignment. He did so in a way not to overwhelm them. He spun the billions of pieces of data in the airline data set as a unique opportunity and a grand challenge.

"We're having a great deal of fun; at least I am," Ward laughs.

It's been over a year since Ward took on the initiative to establish the learning community for statistics sophomores and obtaining the NSF grant. Now that it is being implemented, he sees bright opportunities for his students. The grant is especially geared toward students making the transition through the many rigors of their sophomore year.

"No matter what they're studying, they're going to benefit from working with data," Ward says.

Christina DeSantiago is an applied statistics major paired with Lisa Goffman, a professor of speech, language, and hearing sciences. DeSantiago is looking forward to becoming more experienced with research.

"We are working with students who have language impairments and comparing them to normally developing children to see how they grasp learning words," she says. "It was a little overwhelming when I first thought about it, but I'm really excited to get into the project and see how I can help."

Emily Martin concurs.

"I've learned a lot," says Martin, a mathematics sophomore whose research mentor is Cleveland Shields, associate professor of human development and family studies. "We're doing health care research about the communication between doctors and patients, particularly diabetes patients and their spouses."

Ward's students are entering an era where big data will be a part of every field. The earlier undergraduates have the opportunity to take on such projects, the better.

Ward believes an undergraduate's second year in the College of Science is a crucial one. He sees numerous programs designed to help first-year students get settled and acquainted in their studies. Upper-class students enjoy a wide range of research opportunities. So, what about sophomores? Ward wants to help them to overcome the "sophomore slump," a phrase used to characterize the attrition of many students in the sciences during their sophomore year.

"It is uncommon for a sophomore to participate in research, especially for a full year, but we want to change that," Ward says. "We want students to know they don't have to be an expert before joining a research team. They can jump right in and learn as they go. Purdue students are enthusiastic and capable, and we want to create a culture and climate that is supportive of undergraduate research."

The grant will fuel the program for five years (20 sophomores per year through 2018-19); each student receives a \$9,400 stipend for the work. That makes billions of pieces of data a little easier to tackle.

"I don't have to have a part-time job," DeSantiago says. "The stipend is nice so I can just focus on my studies and my research."

TIME FOLK

