**150 Years of Women in Statistics by Sandrine Dutoit**


This year marks the 150th anniversary of the UC Regents' landmark decision to approve the resolution that allowed women to be admitted into the university on the same terms as men. Berkeley Statistics, as well as the rest of the Berkeley community, celebrates this historical milestone by commemorating the year-long 150 Years of Women at Berkeley celebration.

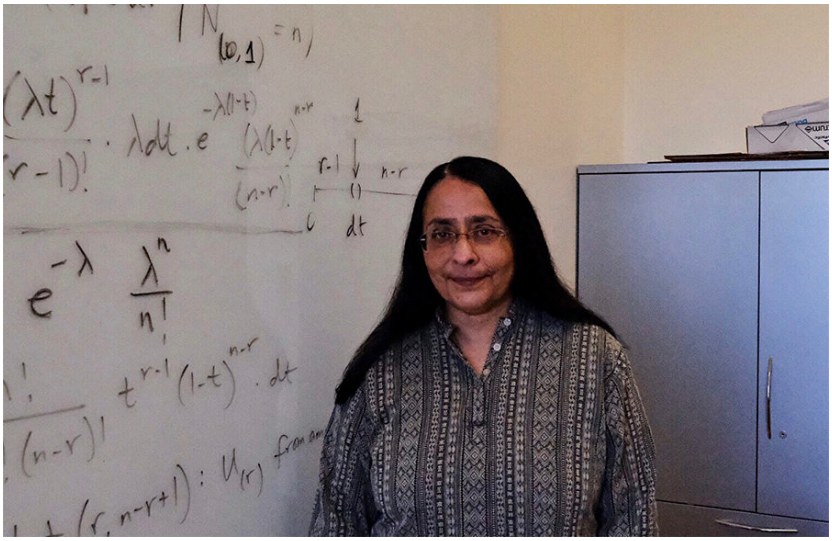
For decades, our female students, faculty, and alumni have consistently been at the forefront of the groundbreaking work of Berkeley Statistics. Together with their staff colleagues, women of our department continue to leave their mark on their fields, the university, and beyond. In celebration of 150 Years of Women at Berkeley, we are highlighting the stories of some of the women luminaries of Berkeley Statistics who have been lighting the way for over a century.

Special thanks to Amanda Glazer and Lauren Pitcher for bringing this project to life!

**Works Cited:**

[150 Years of Women at Berkeley Logo]. (2020). 150 Years of Women at Berkeley. https://150w.berkeley.edu/

**Not just CS: Professor Ani Adhikari talks distinction of data science major at UC Berkeley**
*By Arth Vidyarthi, Staff, The Daily Californian*



*Ani Adhikari*

Four years ago, about 110 UC Berkeley students enrolled in the first edition of a class called Computer Science 94. Today that class is better known as Data 8, one of the largest and fastest-growing classes on campus. Serving as the introductory class to UC Berkeley's new data science major, Data 8 has become symbolic of the growing interest in the field of data science, which combines computer science principles with statistical analysis. The major, offered through the Division of Data Sciences, is on track to become one of the most popular majors on campus.

Today, we engage in conversation with one of the architects of the program — professor Ani Adhikari. Currently the chair of the Undergraduate Program Committee in the department of statistics, Adhikari has been at the forefront of both designing and popularizing the data science division and major. The passion and exuberance with which she approaches her teaching do not go unnoticed by the student body, which voted her "best professor" in The Daily Californian's 2019 edition of Best of Berkeley. In our conversation, we discussed her academic background in the field and insights she has for prospective students, as well as some parting thoughts for upcoming data science graduates — the second class to graduate with such a degree from UC Berkeley.

Adhikari spent her undergraduate years at the Indian Statistical Institute in Kolkata, taking classes across mathematics, statistics and probability. She then continued her education at UC Berkeley, completing a doctorate in probability theory.

**The Daily Californian:** Berkeley has several renowned departments, including CS (computer science), statistics and economics. In that regard, where do you see data science fitting in/forming its own unique sort of discipline? Do you see it as an overlap?

**Ani Adhikari:** You know, I don't actually think it is unique. I mean, statisticians have been doing data science for a long time. What has changed now is that tools that we now have made it possible for us to answer and also ask questions which we could not have conceived of asking before. And so, it is now a collaboration between the statistics department and computer science for developing the fundamental theory, and collaboration with every other domain for developing a theory that actually makes sense for domain applications. Now, the latter has always been true for statistics — that you're not just doing mathematics for its own sake, but that it has to be meaningful to

somebody else. I think it's a much bigger change of mindset for people in computer science as opposed to people in statistics.

All this that is (considered) very new (is) not new for us at all. It's that computer science, maybe even very recently, until about 10 years ago, was primarily used for building things such as software. It's an engineering department. But increasingly, computer science is being used for analysis. So I think for computer scientists, seeing this new direction for their field is actually genuinely new. For us, for statisticians, it seems like what we've always been doing, but on a much bigger scale, which actually informs what we can conceive of today.

**DC:** So I'm a freshman myself, and this next one is a question that a lot of incoming freshmen — including myself — this last year had. How does the data science major differ from a double major in CS and statistics?

**AA**: You know, statistics majors have, for decades, been joint majors. It used to be that econ/stats was the most common combination, but increasingly it has become CS and stats. For us, it's just not new — this idea that students will learn inference and they will learn computing. It has been happening for quite some time. So the data science major is one where you are formalizing some things for students who are interested in a broader experience than they could have with CS and stats. The CS-and-stats double major is often somebody who is exceptionally qualified at developing the field — not just running programs, but understanding the underpinnings of data science. But there is an increasing number of people who use data science and who bring good questions to data science and who have a very perceptive sense of what is the right thing to do and how to interpret it — who do not necessarily need to develop all the theoretical underpinnings but whose insight develops the field. That person might want to be a data science major.

A properly designed CS-and-stats major not only gives you theoretical underpinnings in both disciplines, but because stats has asked for an applied cluster, it also gives you some sense of where you are going to apply these insights. But a data science major could also be somebody who is primarily a social scientist, but someone who is going to make decisions based on vast amounts of data. And hence, they have to understand how these processes work. So that social scientist will have some parts of the CS curriculum and some parts of the stats curriculum but will also have a domain emphasis that allows them to apply this knowledge to whatever area they find interest in.

**DC**: People often associate data science with buzzwords such as "machine learning." What are your thoughts on this?

**AA**: Methods for statistical inference have always had the same base, but they have continued to evolve. Starting from something like basic linear regression models to more complex ones today, statistics has always been about trying to make sense of data in the best and most appropriate way possible. I think every generation or time has its own buzzwords, but the foundations behind these always remain the same.

**DC**: Do you think that being a "division" instead of a "department" disadvantages the data science program in any manner?

**AA**: Data science being a division instead of a department is actually advantageous because it is a combination of applications in so many fields. Being a division is beneficial since it allows for increased collaborations between professors and students from across the traditional departments.

**DC**: What expectations would you have for someone graduating with a degree in data science? Any parting words?

**AA**: Understand the data, have respect for where the data has come from, have awareness about biases within the data, and truly understand the ethical implications that the power to analyze data provides. Berkeley students no doubt have the caliber to achieve all of the above.

**Works Cited**

Shan, Y. (2019). [Photograph of Ani Adhikari]. The Daily Californian. https://www.dailycal.org/

**Q&A with Jennifer Chayes, Associate Provost of the Division of Computing, Data Science, and Society (CDSS) at UC Berkeley**

*By Jon Bashor, Science Writer, UC Berkeley Division of Computing, Data Science and Society*

*Jennifer Chayes*

Jennifer Chayes is Associate Provost of the Division of Computing, Data Science, and Society (CDSS) at UC Berkeley, which comprises EECS, Statistics, BIDS, the Data Science Education Program, the Center for Computational Biology, and the School of Information, for which she is also Dean. Chayes is Professor in four departments and schools: EECS, Information, Mathematics, and Statistics. For 23 years, she was at Microsoft, most recently as a Technical Fellow, where she co-founded and led three interdisciplinary labs in Cambridge, Mass., NYC, and Montreal.

She is a member of the National Academy of Sciences, and the American Academy of Arts and Sciences. Chayes has received numerous awards and honors, including the 2012 Anita Borg Institute Women of Vision Leadership Award, the 2015 John von Neumann Award of the Society for Industrial and Applied Mathematics (the highest honor of SIAM), and an honorary doctorate from Leiden University in 2016. Chayes is deeply committed to diversity in STEM; she has participated in numerous activities and served on many committees for gender and racial diversity.

In this Q&A she talks about the need for an organization like CDSS and how data science can help tackle some of the world's most pressing problems.

**Question:** You certainly picked an interesting time to start a new job pulling together a new division at UC Berkeley. How are things going after your first nine months?

**Jennifer:** People have said to me, “Oh my, you arrived in January and the world collapsed in March.” But I think that times of great disruption are also times of great opportunity. Disruption allows us to rethink everything. Could we use the disruption of online education to think through new ways to make education more inclusive? Can we identify some of the sources of inequities that played out so devastatingly in COVID, and come up with ways to address these going forward? We are looking to push some conventional boundaries in order to have bigger impacts.

There is such depth and dedication among the people at Berkeley. But we need to make sure that the whole is greater than the sum of the parts. As I have virtually met more and more of the CDSS staff, I’m impressed by the breadth of expertise they have and their commitment to the organization. We continue to hire people for critical positions--many of them from campus. We are also striving to ensure we have a culture reflecting how we value diversity and inclusion. If we just talk about it, but don’t demonstrate that commitment they are just empty words.

Among our key hires are members of our executive staff. Kathy Yelick is our Associate Dean for Research and Deborah Nolan is Associate Dean for Undergraduate Education; we've just been joined by Oliver O'Reilly as Associate Dean for Graduate Education and Nathan Sayre as Associate Dean for Faculty. Rebecca Miller has been our Chief Administrative Officer since February, and Cynthia LuBien joined us in May as our Chief Development Officer.

In many ways, CDSS is like a startup. We're moving fast and adapting quickly. It's both exhausting and exhilarating and I am having the time of my life.

**Q:** When colleagues ask about your new position, what's your "elevator pitch" reply?

**A:** Our job is to weave together the riches of the university to solve societal problems. We are focusing on climate and sustainability, biomedicine and health, and social welfare and social justice. We're not here just to advance our core capabilities (though we will do this too), but to integrate expertise from around the campus to advance new research agendas and to ensure that students come out of Berkeley thinking masterfully and ethically about data so they can transform whatever field they choose to enter.

**Q:** You mentioned COVID, which is constraining so many things, but it also presents an incredible opportunity to show how the science of data can help slow the virus and help us create a more just, more resilient future. Can you give a few examples of how CDSS is helping battle the pandemic?

**A:** We have many examples, but here are a few. In mid-March when the outbreaks began, Prof. Bin Yu of the Statistics Department worked with many students and outside collaborators to create predictions of hospital demand seven days out. She and her team used 20 databases from counties and hospitals to come up with five models to accurately learn where to ship PPE and ventilators. They then created visualizations so that non-experts could understand the results. Over 1 million face shields were delivered around the world as one result of this effort.

In another project, a team from Lawrence Berkeley National Laboratory and campus used Natural Language Processing to search existing databases of scientific literature to look for information relevant to treating COVID-19. Now, it isn't labeled as such because COVID-19 didn't exist until recently. By adapting a method used in material sciences, they uncovered 80,000 papers, patents and other info and are adding about 1,500 more each week. At the same time, about 1,000 unique visitors are using the site each week to learn about proteins, genes, symptoms and possible drugs that already exist.

In the area of drug design, Jennifer Listgarten of EECS has been using machine learning to research the use of small molecules in drug design. Once the pandemic hit, she started looking for small molecules that could prevent the binding of the COVID-19 virus to human cells.

One other project sits closer to our home. Maya Petersen and Art Reingold from the Department of Epidemiology and Biostatistics in the School of Public Health and their collaborators are working on a project to help Berkeley and other campuses figure out when it will be safe to reopen. They want to take in many sources of data, such as for symptoms, exposure, mobility, place-based

data such as from air filters, and aggregate public health data to create a machine learning-based risk prediction of who is most likely to contract the disease.

Most epidemiologists agree that the best way to slow a pandemic is to not just test symptomatic people, but asymptomatic persons too. Should we test everybody all the time? Some models say every two or three days, while others say once a month. But testing is a limited resource so how do we use testing and surveillance with so many people to test? They are doing an experiment using two matched cohorts. The first group is of students, with 1,000 who are living in dorms and another 1,000 who are living elsewhere. The second cohort comprises faculty and staff and they have one group who have gone back into their labs and a second group that is sheltering in place. The goal is to use machine learning to find a smarter way of identifying who to test.

That's just a subset, but I think it shows the breadth of our efforts and collaborations.

**Q:** In both external and campus presentations you've talked about how data that are skewed -- either intentionally or not -- can have devastating effects on people in areas such as health care, economics and social divisions. How does this problem intersect with the recent surge in protests for equal treatment for all across social and economic levels? How is CDSS working to address these issues?

**A:** As I mentioned, structural inequities in our health care systems played out in devastating ways during the pandemic. But this was in many ways predicted in a paper written by Ziad Obermeyer of Berkeley's School of Public Health and his co-authors and published last October in Science magazine. Using machine learning, the authors found the racial inequality in how health care is allocated. The authors found racial bias in one widely used algorithm because it uses health costs to measure health needs. Because less money is spent on Black patients, the algorithm wrongly concludes that they are healthier than equally sick white patients on whom more money is spent. This resulted in reducing suggested care to Black patients who need extra care by more than half. We saw similar things happen when the pandemic struck. That paper predicted it to a large extent.

The hideous murder of George Floyd is also proving to be very disruptive and, again, is an opportunity to make significant changes in society. I originally considered having "fairness" be one of CDSS' main foci, but in talking with Linda Burton, Dean of School of Social Welfare at Berkeley, we decided to go beyond striving for fairness and commit to improving human welfare and increasing social justice.

I see us reaching out to people who can effect change. Public defenders, social workers, child welfare workers, policy experts, K-12 educators and have rich experiences that inform the way we can look at the relevant public data. These people are our Berkeley alumni. If we work with the people on the ground who see and live with the effects of racial and economic injustice, we can do proper causal inference, not just causation. We want to understand the effects of interventions. I think we can define a new field around the concept of human welfare and social justice.
We can try out this approach by working with our alumni and hopefully it will then go farther and ultimately benefit everyone in California. I want this to be what we're known for in five to ten years.

**Q:** In an interview before you joined campus, you talked about your commitment to getting more women interested in STEM careers and in getting more students overall to take an introductory class in data science. About 6,000 students take such a class at Berkeley each year and more than half of them are women, which is a good start. What do you think is motivating these students and why do you think such a class is important for students, whatever their major?

**A:** One descriptor of CDSS is "leading in a data-driven world." Students are increasingly aware of the role of data in their daily lives. But awareness only goes so far -- we want to touch every student who comes through Berkeley so that when they leave, they can think confidently, critically, and ethically about data. I like to think we are inoculating them against misinformation.

It's not just about Berkeley. Our Data Science Education Program is being used as a model by other colleges and universities, including community colleges and Historically Black Colleges and Universities. We want all students to think about and question data because it comes into play with whatever their field of study. The more you have this ability in your toolbox, the more you will be able to learn using data and transform whatever field you are in.

When I was at Microsoft, I would meet with groups of young women in middle and high schools through our DigiGirlz program. We would talk about how knowledge of data and computer science could enhance whatever they wanted to do with their lives; not replace what they were thinking of doing, but enhance it.

**Q:** CDSS recently received a $252 million gift toward a new building, the aptly named Data Hub. As the largest gift in the history of UC Berkeley, what does a gift of this magnitude say about the importance of CDSS and its research and education missions?

**A:** First, it indicates there is an extremely generous donor who shares our vision. It's clearly a strong endorsement of the multidisciplinary and inclusive vision we have laid out for CDSS. We're calling it a Data Hub and not the Center for Data Science for a reason. A hub is a central entity with spokes or branches radiating out in all directions. In the case of CDSS, these arms will link us with many other parts of campus, helping us build partnerships with social scientists, economists, mathematicians, sustainable business experts, computer scientists, public health and public policy experts, and others. The new building will both house and further those partnerships. And it will enable education in all of these areas, helping us to foster the growth of future leaders.

**Q:** It sounds like there has been significant progress. CDSS is drawing in expertise from a number of colleges and schools, which gives it a unique flavor on campus. After aligning these pieces, what will the resulting organization look like in a research sense? Where do you see yourself and CDSS in five years?

**A:** I think people should be able to move fluidly, to move where their passion takes them, where the problems take them. For example, in CDSS we are bringing together faculty and staff from computer science, from statistics, from the School of Information and we hope that the rest of the university will be able to feel that same fluidity.

We are building something at Berkeley called the Data Sciences Commons. It will house those who are willing to take the risk of being not "just" a computer scientist or not "just" a biologist. It's my job to set up the structures that de-risk this, which allow people to be fluid in their careers. I think that's the way for people to be most fulfilled and have the most impactful careers.

There's a tug between the fact that we need to run things, but on the other hand we need to make it easy to blur things so people can follow both their passions and the passions created by the world's problems.

I hope we will begin to see the effects of this culture very soon. To go back to an earlier point, I believe that having a diverse staff, in which each of us understands how we contribute to this larger goal and have a stake in our collective success, will be a key driver in this effort.

**Q:** Is there a question I didn't ask that you'd like to ask and answer?

**A:** Ok, here's one: "Jennifer, what convinced you to leave an established company like Microsoft after 23 years to create what amounts to a startup organization at Berkeley?"

That's a great question and one I've been asked many times. It's because when you look at the world, you see so much potential with data and how we interact with data. We have the opportunity to integrate data science across campus to develop end-to-end solutions to some of the world's most pressing problems. How could I say no to such an amazing opportunity? For me, being here makes me feel like a kid in a candy shop.

**Works Cited**

[Photograph of Jennifer Chayes]. (2019). UC Berkeley Computing, Data Science, and Society. https://data.berkeley.edu/

**Betty Scott: Professor Pioneered Using Statistics to Address Bias in University Hiring**
By Jon Bashor, Science Writer, UC Berkeley Division of Computing, Data Science and Society

*Elizabeth Scott*

By the time Elizabeth "Betty" Scott became the first female professor in the Statistics Department at UC Berkeley in 1962, she had already made quite a name for herself. Although she joined the mathematics faculty at Berkeley in 1951, she had actually earned her Ph.D. in astronomy, her first love.

In 1957, she identified a bias in the observation of galaxy clusters, realizing that distant clusters could only be found if they contained brighter-than-normal galaxies, as well as a large number of galaxies. She submitted a formula to correct for what became known as the Scott effect. She wrote 42 research papers on astronomy; her 13th one, written in 1948 with Jerzy Neyman, was her first statistical treatise.

Elizabeth Leonard Scott was born in Oklahoma in 1917 and moved to Berkeley with her family when she was four so her father could attend law school. Known as Betty, Scott earned her bachelor's and doctorate degrees in astronomy at Berkeley, but seeing that the field was essentially closed to women at the time, she chose mathematics as her academic career path.

Her statistics career began during World War II when she was employed at the Statistical Laboratory that Neyman had just started at Berkeley. Her work furthered the war effort, mostly concerned with improving the precision of air bombing. Their collaboration across a wide range of disciplines, from skin cancer to insect populations, continued until Neyman died in 1981.

During the 1960s, Scott combined her research and advocacy skills to address social issues, such as fundraising for Martin Luther King Jr.'s civil rights movement and raising bail for students arrested during the Free Speech Movement.

In the early 1970s, Scott applied her expertise in statistics to analyze career patterns of men and women in academia, documenting discrimination affecting the advancement of women graduate students and women faculty. She then set to work fighting this issue. She was co-chair, with Elizabeth Colson of Anthropology, of a special committee set up by the Academic Senate to study the status of women on campus.

Scott organized the collection of data showing the discrepancy between the availability and proportion of women hired at each level in the University, from majors to graduate school and from graduate school to the professoriate in each department. The resulting report was reprinted by the U.S. Congress in the report "Discrimination against Women" and widely used in the early

1970s as a model for study of possible discriminatory patterns based on sex and ethnicity at other universities.

For a national multi-university study sponsored by the Carnegie Commission, Scott created “the AAUP kit” (for the American Association of University Professors), a set of guidelines for constructing and using a regression model to flag possible cases of wage discrimination. The criteria she developed, published by the commission in 1973, were used by a number of universities, including Berkeley, to provide salary adjustments to temporarily remedy inequities. In the late 1970s, Scott was asked to serve as a consultant to a committee appointed by the American Anthropological Association to examine patterns of hiring and advancement of men and women within anthropology. She worked closely with that committee for several years, while also helping other academic groups.

In addition to her excellence in research, Scott was a superb teacher who cared deeply about her students. She single-handedly ran the statistics master's program for about 15 years, and, in her last seven years alone, she supervised the theses of 11 Ph.D. students. She also co-chaired the biostatistics program for over 16 years.

She devoted a great deal of time to various campus committees and played a prominent role in the design and maintenance of Evans Hall, where statistics is presently housed. Members of the department used to joke that, “Betty owned Evans Hall."

Scott was a Fellow of the Institute of Mathematical Statistics. The Committee of Presidents of Statistical Societies awards a prize in her honor, the Elizabeth L. Scott Award, for "fostering opportunities in statistics for women."

Elizabeth Leonard Scott died in 1988 at age 71. In her memory, the Elizabeth Scott Fund was established on campus the following year and awards a scientific book and a citation "to the M.A. student in the Statistics Department showing the greatest promise in statistical research." The award web page notes that, "Starting out in the 1930s in a university that was almost entirely bereft of women role models, Scott worked tirelessly and determinedly for equal opportunities and equal awards for women in academia. It is no exaggeration to say that today many women owe their position in academia as equal partners among men to Scott's commitment to this cause and her outstanding achievements, as an academic, polemic, warrior, and friend."

**Works Cited**

David Blackwell, Elizabeth Colson, Susan Ervin-Tripp, Lucien Le Cam, Erich Lehmann and Laura Nader, Elizabeth Leonard Scott, Statistics: Berkeley, Calisphere: The Online Archive of California, California Digital Library, University of California, 1991
Rebecca Conley, Review of Equivalence: Elizabeth L. Scott at Berkeley by Amanda Golbeck, MathematicalAssociation of America, July 18, 2017.
Uncredited, Elizabeth Scott (mathematician), Wikipedia
Uncredited, Summary of Equivalence: Elizabeth L. Scott at Berkeley, Amazon.com, 2017
[Photograph of Elizabeth Scott]. (n.d.). UC Berkeley Statistics. https://statistics.berkeley.edu/

**Statistics' Cari Kaufman Honored for Extraordinary Teaching in Extraordinary Times**
*By Jon Bashor, Science Writer, UC Berkeley Division of Computing, Data Science and Society*

*Cari Kaufman*

The UC Berkeley Academic Senate's Committee on Teaching has presented Statistics Lecturer Cari Kaufman with an Extraordinary Teaching in Extraordinary Times Award in recognition of her innovative work to keep students engaged while learning remotely during the pandemic.

Kaufman's award was one of two presented to instructors in the Division of Computing, Data Science, and Society. The other award went to Data Science Undergraduate Studies graduate student instructors Philippe Boileau, Anna Nguyen, Suraj Rampure and Allen Shen. Both Statistics and DSUS are part of the Division of Computing, Data Science, and Society, which leverages Berkeley's preeminence in research and excellence across disciplines to propel data science discovery, education, and impact.

The Academic Senate recognized the awardees at its April 29 meeting.

"We received almost 500 nominations for this award, signifying how our campus community embraced the instructional challenges posed by the COVID-19 pandemic and ensured that our students remained engaged and supported," wrote Glynda Hull, Chair of the Academic Senate's Committee on Teaching.

Kaufman, who joined the Berkeley faculty in 2008, teaches Stat 2, an introductory course about statistical reasoning that typically draws more than 300 students.

"I've taught it for seven semesters, and I deeply love helping students--especially math-phobic ones!--explore a new, more precise way of thinking about learning from data," Kaufman said.

Kaufman and her team moved the class to a Zoom format for the Spring 2020 semester and were planning to continue in the same vein for the Fall 2020 semester.

"Like many of my colleagues, I resisted fully embracing online learning, and I disliked the idea of pre-recording lecture videos," Kaufman said. "However, as the pandemic wore on, and my own 'Zoom fatigue' became more and more obvious to me, I began to wonder about this stance."

Her decision to change her approach came after she sat down to watch one of her own 80-minute lectures, but could only make it through 15 minutes.

“The style that had worked very well in-person, and at least moderately well for those who were attending live on Zoom, became almost unbearable for me to watch after the fact,” she said. “Despite receiving very positive course evaluations in Spring of 2020, this experience left me feeling humbled and ready to try a new approach.

She was selected to participate in Berkeley’s Semester in the Cloud program and credits it with helping her think outside the classroom in adapting to remote learning. She restructured her existing material into modules and eliminated a lot of detail that wasn’t central to primary ideas she wanted the students to learn. This allowed her to break the lectures and other materials into smaller, more easily digestible chunks.

“The Semester in the Cloud program gave me the time and resources to re-envision the core ideas of this course and rebuild it from the bottom up,” Kaufman said. “Interacting with students during this time has also deepened my awareness of student hardship in its myriad forms and prompted me to revise course policies to prioritize flexibility and empathy.”

But even when students and faculty return to classrooms, Kaufman said she plans to keep some of her new approaches.

“Teaching during this challenging time has driven home for me the idea that our academic culture has done harm to students by dismissing attempts to help them overcome barriers as ‘hand-holding’ or ‘spoon-feeding’,” Kaufman said. “Since I view teaching as a process of iterative improvement, I will carry this lesson forward in the future.”

Statistics Chair and Professor Sandrine Dudoit announced Kaufman’s award to faculty, staff and students, writing “I am so glad that the University has recognized Cari’s outstanding contributions to our teaching mission, her innovative pedagogy, and her commitment to the well-being of our students. Congratulations and thank you for all you are doing, Cari! We are fortunate to have you as a colleague.”

**Works Cited**

[Photograph of Cari Kaufman]. (2021). Cari Kaufman.

**A certain "je ne sais quoi": An interview with Sandrine Dudoit**
By Amanda Glazer, PhD Candidate

*Professor and Department Chair Sandrine Dudoit*

Sandrine Dudoit is a professor of Statistics and Biostatistics at UC Berkeley and current chair of the Department of Statistics. She earned her PhD in statistics from UC Berkeley in 1999 and joined the faculty in 2001. Her research focuses on statistical methodology and computing with applications to biomedical and genomics research. Sandrine is a beloved member of the department, not just for her high quality research and teaching, but also because of her dedication to creating a warm and welcoming community for all. When asked to describe Sandrine, students describe her as "amazing," "someone who really cares about everyone in the department," and someone whose "commitment to diversity and equity is inspiring." In this interview, Sandrine tells us about her path into statistics and why she has long loved Berkeley.

**Amanda Glazer (AG):** How did you first get interested in math and statistics? You got your undergraduate degree in math and then switched to statistics for your PhD, right?

**Sandrine Dudoit (SD):** That's right. I have an undergrad and a masters degree in math, more like probability theory, and then switched to statistics for my PhD. But it was never really a planned career in the mathematical sciences. It sort of happened randomly based on people that I met. Back in high school I was really all over the place. I was ok with math, but I wasn't really that much into it. I think what really happened was I went to high school in France, and when you're a good student they put you automatically on the math track.

**AG:** Do you get to choose your track?

**SD:** You get to choose, but there's a lot of expectations. You're not forced to do it, but you're young, and your teachers go, "oh you're a good student you should really do math / physics." In my days, they used to call it "la voie royale," the "royal path," because it opened the doors to the best schools and the most prestigious careers. So, I ended up doing that, and it was ok with me because I kind of liked everything. And then my family moved to Canada.

**AG:** You were in Canada before France too?

**SD:** Yes, I grew up in both. I was born in Montreal to French parents. My family moved to France when I was a teenager, and then back to Canada right after I graduated from high school. So, I did all my schooling in the French system. The typical path when you do math / physics in France is you go on to engineering school. That's what I was geared to do, but then my family moved back

*Sandrine at three years old*

to Canada, and I enrolled in an engineering school. And in Canada, like in the US, they're "real" engineers. They're not mathematicians like they are in France. So that was the biggest shock to arrive in Canada in aerospace engineering. I had no clue what it was. I had a drafting class; it was horrible. My first two weeks I just panicked. I could do math but not that. Luckily, I had a good staff advisor that realized, "ok, this is not good for her. Let's put her in math and physics." So, I ended up back in math and physics. I did a lot of physics as an undergrad. That was good.

A lot of my choices have been based on people I really got along with and that mentored me. This one mentor told me, "well you should probably try statistics. It's more applied." So, I took a summer class in statistics, and I hated it. I think it was the worst class I had in my whole career. Statistics tends to be taught really poorly. It was taught by rote: here's this catalogue of tests in a vacuum. So, I absolutely hated it and thought, "I'm not doing that ever again." So, I went back to do math and then probability theory. Then again, through another teacher, I started doing more statistics and I thought, "actually this is kind of nice."

It was a probability theory and a stats teacher that encouraged me to apply to grad school and to Berkeley. I really doubted myself, but they said, "No, no, you should really try it. What's the worst that can happen? Just go for it. Apply to all these places in the US." I was at an average university in Canada, but they said, "No, the worst that can happen is they won't take you but just give it a shot." They really supported me.

When I arrived at Berkeley, I had two things in mind. I was either going to do probability theory (I had been talking to Steve Evans), or something more with genetics because in the back of my head I had always thought genetics was so cool. So, I took 205 (Probability Theory) and 215 (Applied Statistics). I took 215 from David Freedman. That's really what taught me statistics. Statistics is more than just a form of math. It's a way of thinking, a way of reasoning about data. It was probably the hardest class I've ever taken, but unlike the other classes it stuck with me. The other classes were methods, and methods disappear. You forget about them. But 215 was about the philosophy of statistics and how to approach problems.

**AG:** Why did you decide to apply to primarily American grad schools?

**SD:** I studied at Carleton University in Ottawa. At that point, I'd met my ex-husband. We were both planning on going to grad school together. For him, France would not have worked at all. He didn't know French, and it's really hard to go study in France when you're a foreigner. By that time I felt comfortable with living in the US or Canada. We made the choices of grad schools together. I already had a bias towards the West Coast. In the end, I had narrowed down the places I wanted to study at to UBC, Stanford and Berkeley. I lived in Davis actually the first four years

of my PhD, because my ex-husband was a student at Davis and he was doing lab work, which made commuting more difficult for him.

**AG:** What do your parents do?

**SD:** My dad is a retired diplomat. He studied political science and law. My mom is a stay-at-home mom, but she studied political science, that's how she met my dad.

**AG:** So, I'm assuming you moved because of your dad being a diplomat?

**SD:** Yeah, that's right. My dad was a public servant, but he started moving only when I was a teenager. So, I only went on one posting with them, because I was living on my own after that. But I got to spend vacations in these cool places, like Madrid and Prague.

**AG:** Then you stayed in the bay area after your PhD?

**SD:** Oh yeah, I fell in love with the Bay Area right away. My first visit, which was in April or March before my first year, I thought, "wow I want to live here." It's funny, because before coming to Berkeley, on paper and based on what my mentors were telling me, I was leaning towards Stanford. I went to visit Stanford first and I loved it. I had a great time. So, I was pretty sure I'd go to Stanford. Then I spent two days in Berkeley and I was like, "ah this feels better." Berkeley always felt like more of a fit for me. I always tell applicants that it's really important to visit the departments they apply to: See how you feel, project yourself into a place, because it can be good on paper but it's really how you feel in the environment and how you interact with the people that matters. Every time throughout my career I've always gone back to Berkeley. When I had job offers after my postdoc, again I had to make the choice between Berkeley and Stanford, and it just felt right in Berkeley. I was lucky that the timing worked to be able to get a job at Berkeley.

**AG:** Did you ever think about not going into academia?

**SD:** Oh definitely! It was never part of my plans. Well, even math was not in the plans at all. When I went to undergrad, actually I was not happy at all to be back in Canada. I wanted to stay in France.

**AG:** Why didn't you stay in France?

**SD:** My parents didn't want me to. They just thought I was too young and didn't want a 17-year-old to be by herself in a big city like Paris. So, I was not happy at all to leave France. I would have liked to have studied political science or engineering in France. But instead, I was back in Canada and the default was that I'd just do engineering school. So, that was already not a planned thing. When I started my undergrad studies, I didn't even know about grad school. I was completely ignorant. Really, really, like no career plan at all. It was only maybe in my fourth year of undergrad that I was like, "I'm actually enjoying it so sure let's try two more years." I also felt that it might be a bit too early to get a job because I didn't know what kind of job I would want, especially with an undergrad in math; at that time I hadn't done statistics. I liked studying math, so I thought, "let's keep doing that." Then in the second year of my masters, I still hadn't settled on a career plan.

Again, I was fortunate to have really good mentors that believed in me and pushed me, because on my own I was just very, very ignorant about careers.

**AG:** It's amazing the role and influence mentors have in people's lives.

**SD:** Yeah, definitely. Without them, I wouldn't have thought of applying to grad school. I didn't have a direction.

During the PhD, again I wasn't sure if I would stay with statistics or with academia. I also had a bit of a rough time as a postdoc. My PhD was motivated by a genetics question but it was very theoretical. It was like Markov chains on graphs, no data. It was fun but still very far from data. Then for my postdoc I thought let's do some real applied work. So, I did a 180 and I did my postdoc in a biochemistry lab. I didn't do the wet lab part, but I was where the data were being generated. It was the early days of DNA microarrays. I don't know if you've heard of Pat Brown? He's known right now as the person who created the Impossible Burger. That was my postdoc mentor. So again, I was lucky with who influenced me. I was very lucky to meet him. He had developed this new...have you heard about microarrays? That was about 25 years ago. It's a high-throughput, biological assay that lets you measure the expression levels for entire genomes at once. Before, you had to do that one gene at a time, but Pat and a few others came up with this technology where you can efficiently scale up to an entire genome with thousands of genes at a time.

At the beginning, there was Pat's technology which was all open-source. He had put out plans of how to build a microarray on a website. He claimed you could build it in your garage. There was also a version developed by a company that was all closed and proprietary. So, Pat already back then was all about open science. He gave a talk at Berkeley, at I-House, and I went to listen to his talk and I thought it was so cool: They're generating tons and tons of data and there's so many methods that need to be developed. There's more data and more questions than we have the ability to answer. I thought that it'd be really cool to work with him. By that time, I was a little less shy, so I just walked up to talk to him and I said, "I would love to do a postdoc in your lab. I'm a statistician. I don't know any biology." But Pat thinks out of the box and takes chances and he said, "yes, let's do it." That was an amazing experience. It was a huge lab, so I was a bit lost at first having gone from our department where we're pretty small to now a huge biology lab where I had to learn the language of biologists and the technical aspects of the experiments. I would go to group meetings and at first have no clue what they were talking about. But I learned so much from this experience and I feel it was an amazing preparation for the rest of my career in data science. Taking 215 with David Freedman and then being on my own working with biologists was "real" applied statistics. I would go in the lab and see how the data were generated, be involved in experimental design and have huge, messy datasets to analyze.

So, that was Pat that believed in me. Then I became much more applied and involved in software development. Towards the end of my postdoc, I met Robert Gentleman, who is one of the two original R authors (along with Ross Ihaka, a Berkeley Statistics graduate). At that time, Robert had already written R, and it had already become popular. Robert was getting involved in computational biology; I was aware that there was a great need for good statistical software in computational biology and that I had some serious learning to do in that area. Again, like statistics, I had taken programming courses as an undergrad, but the languages were taught in a vacuum and

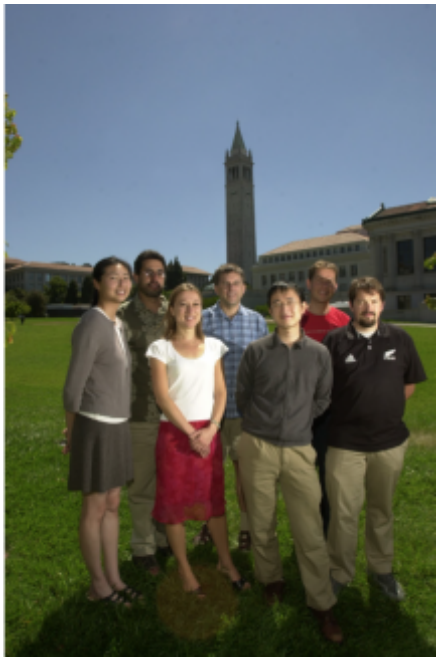
the focus was on syntax rather than general concepts. So, I felt very ignorant and insecure about computing. I started working with Robert, and, with a few others we cofounded the Bioconductor project, which is an open-source software project for biological data analysis. People say that when you play tennis against someone that's better than you, you learn really fast. That's how it felt with Robert and software; just being able to see how he was writing and developing software helped me a lot. So, that's how I got into statistical computing. All about meeting the right people at the right time! I was lucky.

**AG:** And also following all of your different interests and not being afraid. That's pretty cool.

**SD:** It was fun. I was lucky. I was still not sure if I wanted to go into academia when I was in my postdoc. I learned a ton, but it was also very, very stressful. Just realizing what it takes to be in academia and you know the lab environment. Pat was an amazing mentor, but it was a lot of pressure. Not by him at all but just the whole idea of academia about publishing and doing certain things. So, I'd considered going into industry or just doing something completely different. And then again, I thought, "what is there to lose? let's apply for these jobs in academia." There was an opening at Berkeley in biostat. They were looking for someone exactly in my area. That's when computational biology was expanding and Berkeley wanted to have a bigger footprint.

**AG:** So, let's talk a little bit more about Berkeley. What are your fond memories from the department or things you like best about the department?

**SD:** There's so many of them. I think ultimately it's the people that make the department. It's a vague answer but that's really how it is. You know, the atmosphere, the people. Of course, also, the level of the work and the quality of the intellectual environment. The department's atmosphere is unlike what a lot of people thought, including myself, before coming to Berkeley. In my days Berkeley had this image of being really cutthroat. That you would be there and everybody would be competing against each other, but it was exactly the opposite of that when I arrived. I felt really supported. So, I think it was really the atmosphere of the department and a few key people that made me feel welcome: David Freedman, Terry Speed, Philip Stark, Deb Nolan, Steve Evans, and Nick Jewell.



*Sandrine with the Bioconductor Project team in August 2003. Left to right: Yee Hwa (Jean) Yang, Rafael Irizarry, Sandrine Dudoit, Robert Gentleman, Cheng Li, Wolfgang Huber, and Ben Bolstad*

**AG:** It's really interesting how cultures form -- how we have our department that has such a good community but other departments that seem similar can have such different cultures.

**SD:** That's right. I'm trying to put my finger on it but it's intangible. It's just the feeling that is still there when we have events in 1011. It still feels very much the same as when I was a student. It probably goes back to before I arrived, to Betty Scott and people like that, that created a certain vibe.

**AG:** I think one thing that is really apparent to people, especially with you as Chair right now, is how much you care -- about the department and the people and everything. Where does that come from?

**SD:** Thank you. I do care a lot and I'm glad that it shows. I think some of it is probably because the department is sort of like a family. I feel a family connection with the department. I kind of grew up in the department. More than half of my life has been associated with the department. Those of us that were students in the department, like Peter Bickel and Bin Yu, we probably have this extra connection. We feel like the department gave a lot to us, and we want to give back and for the students to have the same experience as we've had. But it's not just being a student at Berkeley. Deb, Philip, and Steve who came to the department as young faculty, also care so deeply. I guess it's part of that intangible vibe. It feels like --

**AG:** Je ne sais quoi!

**SD:** Exactly! Je ne sais quoi! You want to give back and maintain this culture.

**AG:** Tell me about your current research.

**SD:** I love doing research and I wish I had more time to do research these days. A lot of my research is still genomics and high-throughput biological assays.

I'm still working with a collaborator from Berkeley that I started working with when I was a postdoc. His name is John Ngai, and he's in MBC. And this guy's just been amazing. He really appreciates working with statisticians, so we have this really good communication and we work together closely throughout a project. The ultimate goal is to understand the brain. This sounds super general and way too ambitious, but the way he and most biologists approach this is to divide and conquer. John is a neurobiologist and he uses the mouse olfactory system as his model system. The experiments he's doing lately have been concerned with studying how stem cells differentiate in the olfactory system. He's also interested in discovering novel cell types in the brain. The way he approaches that is using these single-cell high-throughput sequencing technologies. Nowadays, with these technologies, you can measure the expression levels of entire genomes at the resolution of single cells, as opposed to a collection of cells. With a collection of cells, you'd be measuring some sort of average, which is informative in some settings, like if you want to predict patient response to treatment. But if you want to classify cells or look at how a stem cell differentiates, you need to look at one cell at a time. So, that's the technology he uses. These experiments are really good examples of data science workflows where you have complex data. These are very high-dimensional datasets, with 20,000 or 30,000 features. There's a whole bunch of preprocessing steps. There's sparsity issues. Often, preprocessing has a much larger effect than the choice of downstream machine learning method. So, we spend an awful lot of time early on "looking" at data, understanding where our data come from. You talk to biologists or a domain expert and they tell you, "oh here's my question," and you're like, "ok, so what does this mean in statistical terms …" So, there's also a lot of back-and-forth framing the domain question and translating it into something that's a statistical or data enabled question. My methodological work falls in the area of high-dimensional statistics. Things like prediction, feature selection, cluster analysis. Upstream, there's a lot of exploratory data analysis and visualization. It's exciting because I get to

play with really cool biological problems, like understanding how the brain works, and at the same time develop statistical methodology.

I really appreciate being a statistician, because there's a lot of variety in our job. I happen to be working on biology today, but tomorrow I could say I want to work on something totally different. Of course, I'd have to learn the subject matter, but the methodology we develop is general, so studying the next domain you're interested in and working in that is cool. My work has been a mix of theory, methodology, and computing. I don't do much computing these days. I miss it. I used to love that back when I was a postdoc or young faculty. Now it's more supervising students on computing projects. It would be nice to have more time for research. I'm lucky to have a great group and I miss seeing them in person.

*Sandrine with former students at userR!, Stanford, June 2016. Left to right: Kelly Street, Davide Risso, Jim Bullard, Sandrine Dudoit, and Kasper Hansen.*

**AG:** How have you seen things for women change, specifically in statistics and biostatistics?

**SD:** I think women are definitely more visible and more confident. That's definitely a change that I've seen. In my cohort as a PhD student and as a postdoc, there were always quite a few women. I think it's more the attitude of women but also the attitude of men with respect to women that has changed. That's good to see. In the department itself, it's nice that we have women in leadership positions. Deb's been chair twice; she's an associate dean now. Bin's been chair once.

**AG:** Do you like being chair?

**SD:** Overall I do, because I feel it's meaningful and because I care. But it's been humbling and challenging. I could've done without the pandemic, but I care so I'm happy to do it.

A big challenge for me as a new chair is to move from what's urgent and not so important to what's important but not so urgent. I feel like I'm not doing very well on that tradeoff. Especially since the pandemic, I feel like I'm stuck in Zoom meetings all day and dealing with moving targets. I would love to be able to, and I need to work on that, step back and work more on big picture issues. The DEI action plan, the applied stats curriculum, hiring, the move into the Division of Computing, Data Science, and Society (CDSS): those are things I care deeply about and I would like to spend more time on. There's so many exciting things to be done with our department. I like variety, so being chair is a nice change. I've been on the faculty since 2001, so almost 20 years. I like to learn. I like to be challenged.

**Works Cited**

[Photograph of Sandrine Dudoit]. (n.d.). Sandrine Dudoit.
[Photograph of Sandrine Dudoit as a child]. (n.d.). Sandrine Dudoit.
[Photograph of Sandrine Dudoit and Bioconductor Project Team]. (2003). Sandrine Dudoit.
[Photograph of Sandrine Dudoit and userR! students]. (2014). Sandrine Dudoit.

**Renowned Statistics Professor Bin Yu Calls Change in Career Direction "My Most Successful Failure"**

*By Jon Bashor, Science Writer, UC Berkeley Division of Computing, Data Science and Society*


*Bin Yu*

After enduring 10 years of social upheaval, including the death of her father at the hands of the Red Guard, as a young girl living through Mao Tse Tung's Cultural Revolution, UC Berkeley Statistics Professor Bin Yu found both peace and order in a math textbook given to her by her cousin.

That immersion in math, coupled with an unplanned detour into probability and statistics, has led Yu to her current position as a Chancellor's Distinguished Professor in the Departments of Statistics and Electrical Engineering and Computer Sciences, and a member of the National Academies of Sciences. She is also the Class of 1936 Second Chair in the College of Letters and Science. Along the way, she has held positions at the University of Wisconsin, Yale University and Lucent Bell Labs.

She was three years old when Mao launched his effort to purge traditionalist and capitalist traits from Chinese society. Because her parents were intellectuals who worked for an agriculture university in Harbin, the Yu family was branded as "bad." Her parents were taken away by students who had joined the Red Guard and locked in makeshift cells at the university. She never saw her father again.

*Bin Yu pictured with her family*

Not long after he died, Yu's mother was released and the family, including Yu's older sister and their devoted nanny, were sent, along with other university faculty, to a farm near the Russian border where they were to be re-educated by farmers. After spending their first night in a cramped room during a blizzard, Yu's mother and nanny pasted newspapers over holes in the ceiling to keep out the cold winds.

"During our time in Xianglan, I watched my mom in awe," Yu wrote in a book chapter about her life. "As the manager, she turned a troubled dining hall around in no time for the youngsters from the Shanghai region who were 're-educated' by the farmers. Her versatility was truly impressive, and her success had a lot to do with her warm and disarming personal style."

Yu began elementary school in a school set up in the other half of the dining hall, while her sister had to walk 45 minutes each way to attend school in the farm headquarters. But as the Cultural Revolution began to wane, the family was sent back to Harbin, where Yu's mother became a department chair at a university devoted to Chinese medicine and life was easier. It was around this time that she realized she was pretty good at math.

**Mentors in Math**

"My cousin, Dawei Huang, gave me a book called Foundations of Applied Math," Yu remembered. "It was a refuge and a place where things were clearly right or wrong. I still remember the cover and tables."

Yu used the exponential and logarithm tables to convert numbers back and forth. "There was no computer, just the tables in the book," she said.

Her next mentor was Jianye Chen, who was hired to be a substitute math teacher in her middle school. Although she didn't have a class with him, he organized math competitions for the students, as well as a math club where they practiced polynomial factorization problems.

"My mom went to talk to him and asked him to tutor me after school, and he did," she wrote. "For the summer of my second year in middle school, he lent me a plane geometry book printed before the Cultural Revolution by Chunfang Xu. I did all the problems in the book that summer and loved it!"

Yu and her family moved to Beijing in 1978, but she returned to Harbin for the summer before high school and Chen taught her the entire high school math curriculum in one summer.

"That really put me ahead of my class in high school and set me on the path to become a math major in Peking University in 1980," she said. "I put PKU math as my first choice on my college application in order to realize Teacher Chen's dream – he was not able to go to PKU because his father was labeled a 'rightist' in the 1950s."

**Looking east to America**

While at the university, she was visited by her uncle, who had been in the U.S. before World War II and was persuaded to enlist in the U.S. Army (or face deportation) and served as an interpreter in China during the war. Afterward, he attended Harvard on the GI Bill and had a long career with IBM.

On his visit, "He gave me my first bicycle. At the time you needed a special permit to buy a bicycle," Yu said. At a store where items could only be bought with foreign currency, he came out with a "beautiful blue bicycle -- almost all the bikes in China were black." The bicycle provided her transportation during her years at the university.

Chen, her former teacher in Harbin, urged her to attend graduate school in the United States. With that in mind, Yu continued to study English on her own and with friends after completing the required two years of English instruction.

Yu placed first in the math parts of the entrance exams for graduate school in math at Peking University, but the professor she hoped to work with would not accept her. Because of that, she decided to switch from math to probability and statistics. Although she considers it a forced move, it has allowed her to do applied work in areas such as neuroscience, genomics and precision

medicine. "I call it my most successful failure," she said. "I feel I was rejected because I was a woman and he didn't think I would go very far."

**Heading to Berkeley**

Her opportunity to study in the U.S. came in the form of the Chern Math Exchange Program between China and the United States. The program was named for Shiing-Shen Chern, a mathematician born in China who also taught at UC Berkeley and was the first director of the Mathematical Sciences Research Institute in Berkeley.

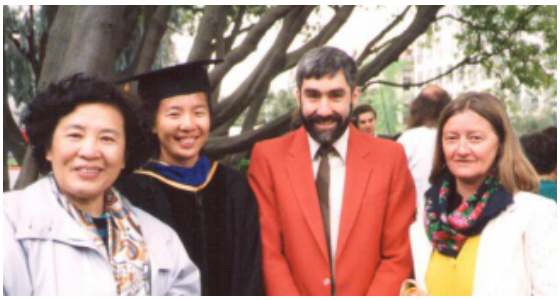
*Bin Yu in graduate school at Berkeley, pictured with Professor Lucien LeCam*

In the beginning, she was not nominated to participate in the Chern program by Peking University, although four male students were. But she got her chance when some students from other universities did not pass the on-site interview at her university. She was interviewed in China by two members of the American Mathematical Society (AMS), one of whom was T. S. Yau, a Harvard professor. Yu was one of 15 students accepted in 1984 and with a recommendation letter from the AMS she applied for and received a fellowship from UC Berkeley. Prof. Lucien Le Cam was the admissions chair that year and he and Yu corresponded often and Le Cam sent Yu her first statistics book by Bickel and Doksum.

"He was my mentor from day one," Yu said. She continued to study English throughout the year before heading to Berkeley in August 1985. Despite her studies, she found the first year of speaking English a rough one and life in the Bay Area was full of new discoveries.

"I had $500 in my pocket and one friend from the math department at UCB met me at SFO," Yu remembered. "On the drive up 101 I thought the sky was so much lower than in China, because of the clouds."

She first stayed with this friend from the math department in a place on Regent Street as she looked for her own apartment. She learned how to open a checking account and find furnishings at garage sales for her new home on MLK Way that she shared with a friend in biology who worked nights in a lab at UCB. She borrowed her friends' car to learn to drive and on the day she barely passed her driver's license test, she parked the car near campus and left the car lights on. When her roommate went out to the car to come home that night, the battery was dead. Lesson learned.



*Bin Yu pictured with her mentor, Terry Speed (red coat), on her graduation day*

Yu continued her studies, making friends with other students from around the world. A new faculty member, Terry Speed, arrived in 1987 from Australia and became her co-advisor with Le Cam. By that time, she had already extended some results in empirical process theory from the independent case to the dependent case under the supervision of Le Cam. Her research with Speed took a different direction into the minimum description length (MDL) principle at the interface of information theory and statistics. Moreover, she credits Speed with extensively mentoring her on critical thinking skills and applied statistics. Her interests in science grew as well, especially through her many conversations with Speed.

She also played soccer with the statistics team called "Bootstrappers," and found she was better suited to defense. In 1987, she returned home to marry her boyfriend and a year later he joined her in Berkeley. Thinking they would return to China, she didn't worry about finding a job in the U.S. But when she finished her Ph.D. in 1990, they made a last-minute decision not to go back.

Instead, she got a position at the University of Wisconsin in Madison. After a year, Berkeley called and she interviewed and accepted a position. But she stayed on at Madison for another year so she could fulfill her commitment to spend a semester at Yale as a visiting professor before returning to Berkeley.

In 1996, Yu was up for tenure but it didn't go smoothly. The Friday before the Monday meeting on her tenure case, a letter was put in all voting faculty's mailboxes with four signatures urging them to deny Yu tenure. The vote was close -- 11 to 9 in favor, with two more votes in favor coming later. She was pleasantly surprised when campus approved her tenure very promptly in February 1997.

Feeling somewhat bruised by the experience, Yu took leave to work at Lucent Bell Labs in New Jersey.

"I found it very liberating and we could pursue our own ideas," Yu said. "That's where I got into machine learning, and started collaborations with Mark Hansen on MDL and Peter Bühlmann on machine learning."

Having kept in touch with her cousin who gave her the math book about 25 years ago, Yu also found an opportunity to co-author a paper with him. Now back on campus for 20 years, Yu and her group have spent the past seven months on a COVID-19 project that builds on predictive work that she did with her cousin and two others at Bell Labs.

She has also authored a paper and is now finishing a book describing a framework that she believes will improve the field of data science and lead to a more rigorous and trustworthy use of methods such as machine learning. Yu laid out her framework for integrating predictability, computability and stability, which she calls PCS, in her paper "Veridical data science" co-authored with her former student Karl Kumbier (now a postdoc at UCSF) and published in the Proceedings of the National Academy of Sciences in February 2020. Read more about the PCS framework.

She and Rebecca Barter, Yu's former student and current postdoc, are now adapting the material into a textbook to be published by the MIT Press in 2021. They also plan to make the material available online at no cost.

**Helping Shape Data Science at Berkeley**

In addition to her work to develop and propagate her work to develop the PCS framework to make data science more trustworthy and transparent, Bin Yu has also played an important role in establishing UC Berkeley's data science program.

In 2014, she was a member of the campus' first data science education committee. The following year, she chaired the Statistic Department's internal review committee under department chair Michael Jordan as part of the department's 10-year external review. That process produced the first written document proposing a stand-alone school for data science.

To bring data science to the classroom, Yu and three of her colleagues Joseph Gonzalez and Joseph Hellerstein (both Computer Science) and Deborah Nolan (Statistics) co-created and co-taught the first instance of Data 100 in Spring 2017.

Yu served on the faculty advisory board chaired by Cathryn Carson that wrote a report outlining the intellectual and organizational vision for today's Division of Computing, Data Science, and Society (CDSS). She helped bring CDSS to fruition as a member of the faculty advisor committee for Interim Dean David Culler. Yu was also a member of the hiring committee that selected Jennifer Chayes as the Associate Provost for CDSS and Dean of the School of Information.

**Works Cited**

[Photograph of Bin Yu]. (2014). Bin Yu.
[Photograph of Bin Yu and family]. (n.d.). Bin Yu.
[Photograph of Bin Yu and Lucien LeCam]. (n.d.). Bin Yu.
[Photograph of Bin Yu and Terry Speed]. (n.d.). Bin Yu.

**Shanna Swan**
*By Amanda Glazer, PhD Candidate*

Shanna Swan (then known as Helen Wittenberg) earned her PhD in Statistics from UC Berkeley in 1963, after earning an MA in Biostatistics from Columbia and a BA in Mathematics from the City College of New York. After proving an extension of the Central Limit Theorem and graduating with her PhD, she stayed at Berkeley working at the seismographic station. Today, Shanna is a renowned epidemiologist and professor of environmental medicine and public health at the Icahn School of Medicine at Mount Sinai in New York City. She has had a successful career working at Kaiser Permanente, the California Department of Health Services, the University of Missouri and the University of Rochester. Her much anticipated book, Count Down, about how chemical exposures are affecting our health, will be available next month. I sat down with Shanna to hear about her amazing career path and time at Berkeley.

*Shana Swan*

**Amanda Glazer (AG):** How did you first get interested in math when you were younger?

**Shanna Swan (SS):** This is a weird story. I went to a school for intellectually gifted children. It's called Hunter Model School in New York City. I wasn't aware of a particular interest in math. When I was around 10 years old, I was going to the 92nd street Y, and I started acting and I loved it. I became a child actress. I was on TV, radio. I went to the High School of Performing Arts. I was an actress, and math was nowhere in the picture. But I was clearly smart. After two years in the High School of Performing Arts, for various reasons, such as the academics, I decided to switch to Bronx Science. It was a real difference academically. At Bronx Science I got the first C in my life, and I had to start working. When I went to City College I majored in math with a minor in logic. I think it was because I had some very good teachers at Bronx Science that I got really interested in mathematics.

**AG:** What happened to your acting dream?

**SS:** It was wonderful for the time. But I'm 5 feet tall, so I played children all the time but pretty soon you can't play children. I was rejected so many times, making the casting rounds, and I just couldn't take it. It's not my personality to be continuously rejected. So I just figured, I can't do it, I'm not going to make it in this field. And that's true, it's very, very hard to make it in acting. I did the right thing, but I still feel very fondly about acting. But I have to say the fact that I did that means that I can talk to 2000 people and not care. I can talk to a TV camera and not care. So, it was very good training for that.

**AG:** How did you end up at Berkeley for your PhD?

**SS:** My path has been very unusual in many ways. My interest was originally in foundations of mathematics and logic. I got the Ward Medal in Logic from CCNY. Then the summer after graduating I had no money, so I took a job at Health Insurance Plan of New York as a medical coder. It was very boring but fine. It was a job. The head of the group called me into his office and he said, "What are you doing here? Why aren't you going to school?" And I said, "Well I don't have any money." And he said, "Just apply to the biostat program at Columbia. I'll write a letter, and you can get a NY Regents Fellowship." And that's what happened. I got my Masters in biostat from Columbia. While I was there I studied with somebody named Agnes Berger. She was a friend of Neyman's.

Then in February of '61 I took off and went to Berkeley. I went first to see Tarski because I was interested in foundations of logic, mathematics. But he couldn't get me any money, so I went to Neyman.

The semester had started. I never applied to the Statistics PhD program. This is how I got into the program: Neyman said, "Go to the board." So, I went to the board and he said, "Integral 0 pi/2 sin(x) dx." I wrote that down, and then he said, "Now which is bigger, in the sense of Riemann or Lebesgue?" And I looked at it, and it was so simple. I said, "Well, it's the same." He goes, "Good. Now can you make tea?" And that was my entrance exam. Plus the letter from Agnes Berger. I got a key to Campbell Hall, a desk, and a teaching fellowship, and then I was good to go. I was in the program.

Coming from foundations of mathematics, I was not actually interested in statistics. My interest was in probability theory, so I studied with Loève and Le Cam. My thesis was a theorem which was published in Loève's book. The theorem was an extension of the Central Limit Theorem for randomly stopped sums of independent random variables.

**AG:** Tell me more about your theorem.

**SS:** I love the central limit theorem (CLT), which to me is one of the most beautiful theorems in mathematics. So, I was thinking about the CLT and thinking about the sequence n, n+1, n+2,... and then I thought, well in a laboratory that doesn't happen. You lose some rats, some rats have a litter; the sequence is not actually integer. So then what happens to the CLT because it postulates an integer progression. I went to Le Cam and said, "I'd like to think about this, maybe for my thesis." And he said, "Okay. Just bring it down. Make it really simple." And so I did. I started with what happens when that stopping variable converges to an integer valued variable and then looked at more complicated cases. The theorem is, "if the stopping rule is independent of the underlying variable the CLT holds." I think it's beautiful.

**AG:** What are some of your favorite memories from Berkeley?

**SS:** Oh my god, Amanda, the department was the most wonderful department I have ever seen in my whole life. First of all, we were very few students, maybe two students per faculty. The faculty, just think about who was there, David Blackwell, Henry Scheffe, Hodges, Lehmann, Neyman, Le Cam, Loève. I mean just imagine being with those people all the time! It's like heaven.

Then there was Neyman. Neyman's influence cannot be underestimated. He brought all those people there, and then he set up the coffee room. This place, every day at lunch, everyone came there. There was coffee, a fridge and a blackboard. People wrote things on the blackboard. They would leave a problem and say, "Here I'm stuck." Then somebody else would write in the interim. It was continually interactive, students and faculty. There wasn't a line between students and faculty, it was like we were all in this together.

I remember sitting next to David Blackwell and asking him about a problem, and he'd just sit back and go, "So, let's bring it down to the unit interval." He would make everything simple. It was just the most wonderful, nurturing place. Everyone was supported. Then at 3 o'clock Neyman would walk through the halls, and he'd say, "Come! Come, come, come! There is cake!" So, once a week we had cake. It was a family. It was just absolutely extraordinary.

**AG:** What did you do after you earned your PhD?

*Shanna in the early 1960s*

**SS:** As soon as I finished my PhD I was given an assistant professorship. I got pregnant just at the time that I defended my thesis. I went to Le Cam and said, "I'm pregnant. I'm going to get married." And he said, "Don't do that." He said wait until you retire. There was a couple in the math department that had done that. Because there was a nepotism rule at that time, so one of us would have to go. David Freedman (her husband at the time) was hired earlier, so I left.

I left the department but not the university. I got a job in the seismographic station. They had a grant, and they were looking for a statistician, so they hired me. They were looking at the travel time of p-waves which had always been an interesting question. Until atomic testing they didn't have a place where they knew the signal came from in time and space. Because you have an earthquake but you don't know exactly where it is. But with atomic testing you know exactly. So they wanted to use the data from atomic testing to update what's called the Jeffrey-Bullen's tables, the tables of p-wave travel times for earthquakes. That was the project, and I actually published several papers which were pretty well received in that topic. I did that maybe 5 years, and it was interesting. With a statistics degree you can have a lot of applications. I always enjoy new problems.

**AG:** What did you do after that?

**SS:** I went to Kaiser Permanente as a statistician. They were doing a project on the health effects of oral contraceptives, which is starting to get close to what I'm doing now. That was really interesting, and I did that for quite a few years. First, I was the statistician on a particular study. The project was to see what the health impacts of these contraceptives were on a whole array of women's health measures -- their lung function, blood pressure, cell counts, etc. As I was doing that, I started to get interested in the science of it. Because that's the way I am, I can't just do the

numbers. So I started studying the subject matter, and I started to write papers on my own. I gradually morphed into an epidemiologist, because I was publishing on epidemiology. Then I heard that the Health Department was hiring, and they wanted someone who could fill the role of both biostatistician and epidemiologist. And I said, me! So, I was hired, and it was wonderful. I stayed for 17 years in the California Department of Health Services. It was a new program in environmental health and within that, I got to form a reproductive health group.

*Shanna in the late 1960s in Strawberry Canyon*

**AG:** At what point did you move back to New York?

**SS:** That was a pretty long trek. I was in the health department. I was appointed to a National Academy of Sciences committee on a class of chemicals called endocrine-disrupting chemicals (EDCs) that affect the body's hormones. I went to this committee knowing nothing about these chemicals at all. I went because it sounded like an interesting thing to do. On that committee I was the only statistician or biostatistician. A paper had just come out saying that sperm count had declined 50% in 50 years. So, the committee asked me, "What do you think of this paper? Do you think we should consider it in our deliberations?" I looked at it, and I thought, "I don't think so." It looked pretty thin. I had some time, so I said, "I'll just spend the next 6 months and I'll address that question."

I got the underlying data. I looked at all the confounders and risk factors. What could have artificially created this trend? I had two colleagues work with me on that, and when we were all done, it was really quite astounding, that the point estimate for the slope of the decline was identical to the decimal place. It was stunning. Biggest shock of my life. I took it back to the committee and said, "I think there's something here. I cannot make this go away." The paper had been attacked worldwide, and I had taken every bias into account, and I couldn't make it go away. It wasn't the age, sampling method, or way of recruiting. Nothing changed it. So, I thought, well this is pretty strong. I went back to the health department and I said, "EDCs are serious risk factors." But they said, "We're not interested." So then I moved back into academe with fulfilling positions at the University of Missouri, the University of Rochester, and Mt. Sinai, continuing my research on EDC's.

Actually we're going to go back to California in May. Because in the 37 years I lived in California I had 3 children, and I have 6 grandchildren and my sister's there and she has 4 children and so we're like 25 people, family. And I'm here and they're not. All in California. I feel like I'm going to go home, because that was my home.

When I went out to Berkeley for the first time, which was the first time I ever flew, I took that bus from the airport to Berkeley. I went to the International House. I figured I could get a room there. I didn't have a reservation. I knew nobody. This was really crazy. It was just a great time. But when I took that bus into Berkeley, and I saw the palm trees, and I'd come from NY in

February, I thought I'd died and gone to heaven. I thought, "Oh my god, I am never leaving this place."

**AG:** After working so long in public health and writing a book, how do you think about science communication?

**SS:** I spent 17 years in the Health Department and that really changed me. That made me a public health scientist. Our job was to respond to the concerns and needs of the California population. That was our primary goal. It was not to do research, but we did do research in the service of that. The first project I took on in the Health Department was in Santa Clara where there had been a spill of toxic chemicals from a semiconductor plant, and the community was concerned about the water. I went there to determine if there was really a problem and whether there was a possibility of a study. And the answer was yes, and I did that study. What was driving my science at that point was concern of the public. That has remained my overarching concern. I don't believe it makes sense to do science in a vacuum.

I told you how in '95 I got on this committee when they asked me to look at sperm decline. Well my book starts with the story of sperm decline where I repeated that study in a much more sophisticated analysis. That study went viral. It was the 27th most cited study in the world in 2017. It was a very big thing. Because of that, someone came to me and said, "Will you write a book about this?" I thought about that a lot, and I thought, well why should I write a book? And what made me write the book was the fact that I had talked for 20+ years to scientific and government committees and written over 200 papers, and people still didn't know what an endocrine-disrupting chemical is! So, what I had done had not worked. It didn't cause change or public health to improve. So, I decided to write the book. Now, I'm launching a campaign which we're calling the Action Campaign, to follow up on the publication of the book and get communities involved. I feel like I need to do something that makes a difference. Public health does that, and I hope what I'm doing now can do that.

**Works Cited**

[Photograph of Shanna Swan]. (n.d.). Shanna Swan.
[Photograph of Shanna Swan from the early 60s]. (n.d.). Shanna Swan.
[Photograph of Shanna Swan at Strawberry Creek]. (n.d.). Shanna Swan.

**Play for each other: An interview with Denise Yee**
*By Amanda Glazer, PhD Candidate*

*Denise Yee*

Denise Yee has been at UC Berkeley for fourteen years, working as the Undergraduate Student Services Advisor and Course and Curriculum Advisor in the Statistics Department. She graduated from UC Santa Barbara with a BA in Asian American Studies and Mathematics, and from University of Maryland with an M.Ed. in Counseling and Personnel Services. Her expertise makes the undergraduate statistics program run smoothly and reassures and motivates students to pursue a career in statistics. Undergraduates describe Denise as "kind, genuine, and helpful," "incredibly patient," "very approachable, honest, and intelligent," and someone who "puts in so much work every day to make the department a more welcoming, inclusive, and less intimidating place." In this interview, Denise tells us about her job, career path and words of wisdom for students.

**Amanda Glazer (AG):** Could you start by telling us about your job and what your day-to-day looks like?

**Denise Yee (DY):** My official title is "Undergraduate Student Services Advisor and Course and Curriculum Officer." My job encompasses working with our undergraduate students in statistics. I work with both declared majors and minors, and also students that may just be taking our courses and have related questions. That's one part of my role. Then the other part is doing the scheduling and enrollment management for our courses. I work with faculty to schedule courses, help them manage enrollment, troubleshoot, and serve as a resource for them. I also refer them to other resources on campus when issues come up with students in their courses. That's the snapshot of what I do.

In terms of working with the statistics majors and minors, there's some administrative work that happens on the back end to get them declared and confirm that they've completed all of the major's requirements so they can graduate. But before that, there's the developmental part of seeing where they're at, why they're interested in statistics, helping them refine their goals, and asking them a lot of probing questions like "what got you interested in statistics? Do you have a particular career or post graduation goal in mind?" Then I try to connect them with resources that can help them make the most out of their time here.

**AG:** What are common concerns that come up with students?

**DY:** For pre-majors, it's whether they can make it into the major, because we have pretty rigorous prerequisites. So, I reassure them that there are additional options in case it doesn't work out the

first time, and that they can still apply the following semester and make up any prerequisites that they didn't satisfy.

That also ties into imposter syndrome which many students, particularly new students like incoming freshmen or transfer students, experience across campus. A lot of the time they are very successful prior to coming to Berkeley and sometimes they haven't had to put as much time into their academics as is needed when they get here. They experience this stark difference in academic rigour and expectations. I reassure them that it's very common that students experience imposter syndrome and have doubts about their own capabilities. I try to be a reassuring voice to them and say: "You can do it. It may take more time, energy and practice than you may have been used to, but you can do some self reflection and assess if there's anything that might need to be changed or other resources that you haven't accessed in the past like tutoring or office hours."

Another thing is students in the major deciding what they should do after they graduate. But I think people, not even just college students, are always re-assessing what their calling is in life and what their interests are. The important thing is to reassure them that it's perfectly normal not to know what you want to do exactly and encourage them to pay attention to all the experiences they are having -- whether it be in the classroom, a conversation they're having with a friend or a family member, an internship or volunteer experience, or their reactions to the news. I tell students to pay attention to those feelings and the things that pique their interest and that can help them narrow down what they might want to go into.

**AG:** That's great advice. I've definitely noticed there can be a lot of uncertainty and anxiety for students.

**DY:** I feel like in our department, and in the field of statistics, it's really easy to tell students, "don't worry, you're going to get a job." Data is being collected everywhere. There's a reason why there's this steep increase in the number of statistics majors. It's because data is collected everywhere. Every industry, government, and organization has some sort of data that they want analyzed, so students shouldn't be too worried about finding a job. Finding a job is still a process and can take an investment of time but students can rest assured that their statistics degree is going to give them a lot of opportunities.

*Denise Yee at Cal Day*

**AG:** Absolutely. So, what is your favorite aspect of your job?

**DY:** I do like talking with undergrads. I identify as more of an introvert so when I have lots of appointments it can be draining, but at the same time it's pretty rewarding and satisfying to be able to help students. Especially when they're about to graduate or have graduated and have a dream job lined up or are ready to go onto their graduate program, that can be pretty satisfying. Then more on the course and curriculum side of things, scheduling isn't always the most fun thing but I

do appreciate being able to use a different set of skills in terms of problem solving. It is super satisfying when things line up together. When you're thinking about scheduling, we take into consideration the preferences of the instructors, the timing of graduate courses, that sort of thing. We don't want too many courses conflicting with each other if they're going to draw the same types of students. It's nice when things are settled and faculty seem satisfied.

**AG:** Putting all the pieces together sounds really challenging.

**DY:** I think I've been doing it so long that I don't get as worried about things and have gotten to know some of the preferences of our faculty so that's been nice. Then with remote instruction, we're not restricted so much by the physical classrooms, so we have a little more flexibility in terms of the times. Things are always changing, sometimes instructors' assignments are swapped kind of last minute or it's difficult to find a qualified GSI and we need to change a lab time to make it work, so we're constantly making adjustments here and there. It is nice when things have kind of settled down.

**AG:** You've been with the Statistics department at Berkeley for a long time. What was it like when you first started? How has it been over the years?

**DY:** I think it's me and Ryan who've been in the department the longest. I've been here since October 2006, so 14 years. When I came here my position was sort of new in that we originally had one student services advisor who advised PhDs, MAs and undergrads. This is when we had maybe 60 undergrads, 5-10 masters, and maybe about 30 PhD students. So, it was a lot smaller when I started. They siphoned off the undergrad part from that original student services position and then added the course and curriculum stuff that had been part of this other administrative position. I was replacing, at least in terms of the course and curriculum work, Pat Hardy, who had been in the department almost 40 years. So, I had some big shoes to fill.

Then also just trying to figure out how these two areas worked together. It took a little time to figure some things out. The staff size was I guess somewhat similar. I think we probably had about 8-10, but there were only two advisors. There was the grad advisor and then me.

We saw a huge growth shortly after I started in 2006. It was maybe 60-80 undergrads at a given time, and then within a few years it was like probably double that. A couple years ago we probably had over 400 students in the department major. Ultimately we were able to hire another advisor which was good. But for a while I kind of felt like I was treading water just trying to get the basics all taken care of. When the growth came as quickly as it did without the additional resources, not just staff, but also the faculty and GSI resources to be able to teach more and more courses, that kind of limited our ability to be able to serve all these students. So, we implemented these more rigorous prerequisites to help slow the growth. For a long time we were growing very quickly. We had a ton of double majors, probably over half of the students, which I think is totally natural because students saw how useful it was to have the tools and knowledge you learn in statistics.

**AG:** How did you decide to go into advising and higher education?

**DY:** I did my undergrad at UC Santa Barbara where I was a double major in Asian American Studies and Math. With math, I was on a teaching track, and I thought I might teach high school. I liked it, and I felt pretty confident in my mathematical abilities, but I sort of shifted gears a little bit after a math teaching internship over the summer between my junior and senior year. I shadowed a classroom with a credentialed teacher. This was summer school so it was two algebra classes. There was a credentialed teacher doing the majority of the work, but I felt really overwhelmed by the demands of being a teacher -- writing quizzes, correcting homeworks and projects, and all that. But I still really appreciated the purpose of education and public education in particular.

So, I came back home after graduating. I didn't end up going into high school teaching. I got a job at Stanford in their advising center as a program assistant. It was mostly a lot of administrative work but in support of work that their central advising center provided. I supported some of the professional advisors that ran programs like the Freshman Advising Program and an expanded advising program which was several mentoring programs that connected different groups of undergraduate students with graduate students. So I got to see how it worked and I had some great mentorship from my immediate supervisor. I worked there for about four years, and then I decided I would go back to school to get a masters. It seemed like a lot of the advisors in the advising center had advanced degrees, and I felt like I needed that extra background and experience because Asian American studies and math didn't give you a whole lot of experience with counseling and psychology.

I went to grad school at the University of Maryland and got a Master of Education in counseling and personnel services. I enjoyed my time working at Stanford, so I decided to further pursue working in higher education. As an undergrad, I don't know that you really see that this is a career field. It was weird because I don't know why I didn't think it could be, because in my student organization involvement I was always working with counselors and student org advisors, but it just didn't occur to me that you could do this as a profession. But once I got my program assistant job and then went to my grad program I was like, "oh this is a viable career path." It merged my appreciation for education, and in particular higher education, and then an interpersonal aspect, the interactions you get with students and others, in a helping profession.

In terms of sticking with statistics, it's been pretty good. I do appreciate our medium sized department where it's intimate enough where I think I can confidently say all the faculty know me.

**AG:** Why did you choose to apply to the advising job at Stanford?

**DY:** It was in higher education and advising. As an undergrad, I was an Educational Opportunity Program peer counselor. When I initially applied, the particular program assistant position was to support this expanded advising program which was run through the undergraduate advising center. The program was Partners for Academic Excellence (PAE). It was a mentoring program that targeted support for different groups: student athletes, Black, Chicano / Latino, LGBTQ+ students. It was in line with some of my values in terms of supporting underrepresented groups. I liked the way the program was run.

**AG:** How have you seen things for women at Berkeley change?

**DY:** A lot of my inspiration comes from the women leaders that are around. Just looking at immediate supervisors, La Shana's my current supervisor, and there's Laura. Then having Sandrine as chair, and Deb and Bin as chairs in the past. Hearing Cathy Koshland's name, Frances Hellman, Carol Christ and now Jennifer Chayes, along with countless other women in the advising community, it just seems like there's more women. I don't know if it's true that there's a lot more women in upper administration, but maybe it's just that they're more visible. At least recently, I know I have definitely acknowledged to Laura, La Shana and Sandrine how much I've appreciated, especially because of the pandemic and going remote, how our leadership has really advocated for us, for me, as a staff member. I've really appreciated that. They're working to represent us and our concerns really well.

**AG:** Final words of advice or hopes for students at Berkeley?

**DY:** When we have in person orientations for the new transfer students, one thing I try to impress upon students is to really take advantage of the network of people and resources they have access to at Berkeley. To not just take courses here, which are great and a good way to get to know faculty, but to also take that extra step to actually get to know the faculty and not just put your head in your books.

It's related to something I heard from the President and CEO of the San Francisco Giants, Larry Baer (also a Cal alum), who spoke at my sister's commencement at San Francisco State in 2013. I was still at Berkeley, and the Giants had just won the World Series. Larry Baer talked about how the world's most complex problems aren't going to be solved by individuals. Akin to what the Giants were doing in terms of playing for each other, you're going to have to work with other people, and the most complex problems will likely only be solved by working together. I try to impress this upon our new students: to get to know each other for that reason. Because you never know who's going to be your next boss. The person next to you in class could create this huge company. You could be working with your groupmate, and you could just brainstorm some idea that you could actually pursue and make into a great thing and impact the world. I think data science is really founded on the principal that you're bringing together experts and people from different disciplines to address issues in all different areas and it's very interconnected. I refer to Larry Baer multiple times in these transfer orientations and tell students, "play for each other!"

**Works Cited**

[Photograph of Denise Yee]. (n.d.). Denise Yee.
[Photograph of Denise Yee at Cal Day]. (n.d.). Denise Yee.

**150W: Women of the Berkeley Statistics PhD Program**
*By Amanda Glazre, PhD Candidate*

In celebration of 150 years of women at UC Berkeley, we asked some of the current women pursuing a Statistics PhD at Berkeley to share with us what they love about statistics and their hopes for the future. These women are conducting impactful research in a wide range of areas, including biostatistics, causal inference and theoretical statistics. In addition to their top tier research, they serve on different Statistics Graduate Student Association committees and help create the Statistics department's wonderful community. We strongly value and appreciate the strength of the women in our department and look forward each year to more women joining us.

Tell us a little about yourself.

**Miyabi Ishihara (MI):** I'm Miyabi Ishihara, a fourth year PhD student in statistics. I do research in causal inference, remote sensing, and spatial analysis and use them to study natural disasters and their impact on the economy. I like making t-shirt designs featuring cats doing statistics.

**Emily Flanagan (EF):** This is my first year at Berkeley! Previously, I majored in statistics and pure math at the University of Washington. I'm not sure what I will be researching yet. I tend to find most things interesting once you start looking into them, but lately I have been most intrigued by causal inference, and working with real, messy data to solve real problems. In my free time, I like puzzle games, and reading short stories.

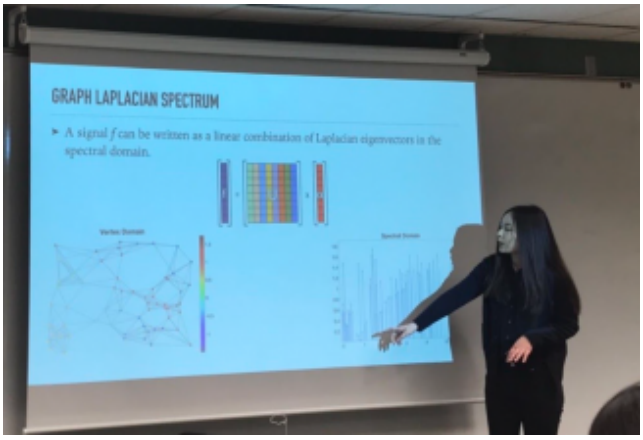
*Zoe Vernon playing basketball for Washington University in St Louis.*

**Zoe Vernon (ZV):** I'm a fourth year PhD student in statistics. My research is in developing methodology for genomics data. My first project, which I am in the process of wrapping up, involved creating a new statistic to measure the dependency between gene expression profiles. The statistic is specifically designed to assist in the discovery of novel treatments for diseases. In my free time you can probably find me at the RSF playing some sport, and I help coach a high school girls basketball team, which is very fun. I also love to bake!

**Huong Vu (HV):** My name is Huong Vu. I am an immigrant from Vietnam. I am a first-year PhD student in the Statistics department. I graduated from UC Davis in 2018 and majored in Mathematical Analytics and Operations Research. I like gardening and DIY ideas.

**Tiffany Tang (TT):** I'm a third year PhD student in statistics. My research interests are primarily problem-driven and lie broadly at the intersection of applied statistics/data science and medicine.

**Shuni Li (SL):** I'm Shuni, a 3rd year PhD student in the Statistics department. I'm broadly interested in statistical machine learning (ML) and optimizations with applications in computational biology. Currently, I'm working with Haiyan Huang on modeling heteropolymer sequences. We combine traditional statistical approaches and state-of-art machine learning techniques to help the design of a novel polymer material that mimics protein behaviors.



*Shun Li presenting research*

**Corrine Elliott (CE):** I am a second-year doctoral student in statistics. Prior to entering UC Berkeley, I studied mathematics, statistics, chemistry, and computer science at the University of Kentucky, and then spent a year working as a statistical consultant in the design of clinical trials. My past research includes predicting the performance of organic components in lithium-ion batteries; analyzing protein structural sequences; and characterizing an algorithm for semi-automatic design and validation of statistical models. My interests lie broadly in the use of computational and applied statistics to support research in the natural sciences, particularly medicine. Outside of academia, I am an avid swing dancer, connoisseur of coffee, and neophyte cook.

***How did you first get interested in statistics?***

**EF:** I got interested in statistics after taking AP Statistics in high school. I loved the versatility of the subject. It felt like any problem could be translated and tackled using the language of statistics. My favorite part of the course was probably thinking about confounding variables and all the ways things could go wrong. I think this shift in mindset has stuck, and it encourages me to question and probe the world around me.

**ZV:** I first got interested in statistics through a research project as an undergrad. I was a math major and had done some math research, but happened upon a project that was more statistical and I really enjoyed it. After that I took a couple of statistics classes in undergrad and decided to apply to grad schools.

**SL:** I actually first became interested in machine learning when I was an undergrad. At that time, ML was a mysterious but powerful black box to me. As I was more exposed to the area, I found that statistics and optimization were the foundation to ML's power. That was when I decided to study statistics in grad school. To me personally, statistics has the charm of rigorous mathematical reasoning, which I have always been passionate about, and the super-power of solving complex real-life problems.

***Why do you love studying statistics?***

**EF:** I love how it can be used anywhere. I love how we can translate seemingly crazy, messy, nasty data into meaningful patterns, even if it's not the pattern we wanted to see!

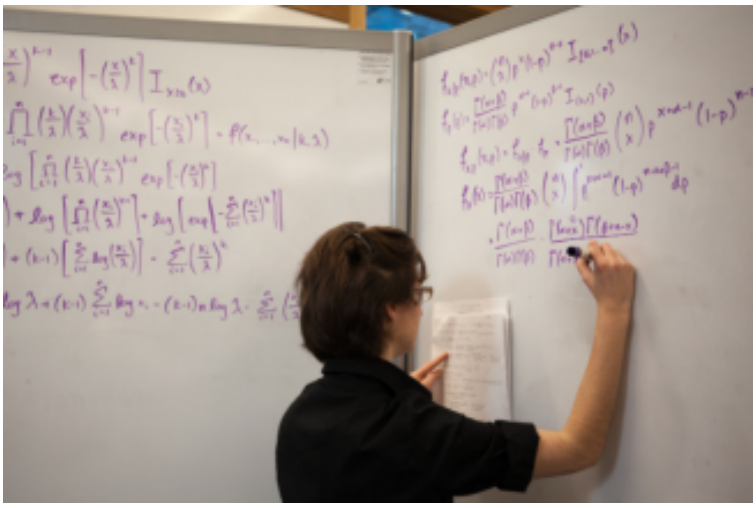
**MI:** I think that it is important to be intentional about the world we live in and engage in society in our own ways. Being an applied statistician lets me be an active member of society by giving people more tools to solve the problems they are most interested in. I was inspired by my master's advisor to view statistics in this way, in relation to the world.

I hope that the statistics community can extend access to truly nourishing learning and research environments, so that more people can cultivate their statistical talents to make their original contribution to solve pressing issues that we face on this planet.

**HV:** I love studying statistics because in statistics, every number has a story to tell. For statisticians, the goal is not only producing a high-accuracy model, but also being able to explain the model in a meaningful way.

**ZV:** I love working with real data and making cool, informative figures.

**TT:** I really enjoy meeting and working with scientists, doctors, and experts from other domains to make sense out of the data. It's exciting to be able to learn a fairly general set of statistical tools that can be applied to a wide variety of real-world problems.



*Corrine Elliot doing statistics*

**CE:** My regard for [applied] statistics lies in its being simultaneously a science and an art. Statistical theory shares the beauty and logic of pure Mathematics, but to reap actionable knowledge about the natural world, we must interpret the results with creativity and an appreciation for the domain of application. Even well-established statistical methods can gain new life by their administration to new research questions, rendering our discipline a source of eternal fascination and potential for insight.

***What's one of your proudest accomplishments?***

**EF:** My proudest accomplishment would be starting the Statistics Directed Reading Program at UW with Anna Neufeld. I am very passionate about connecting everyone, regardless of background, with statistics since I believe it is a tool that can be used to empower people. This program connects up a grad student in statistics with an undergraduate student to embark on a small reading project about an advanced topic in statistics. One student built an R package that went viral on Twitter, so much that their hosting site crashed! That was super cool.

**HV:** Being at UC Berkeley and doing a PhD program is my proudest accomplishment personally and professionally. As an immigrant and a female in a STEM field, I hope that my accomplishments inspire others to keep pursuing their dream. I think my family is really proud of me and that makes me proud of myself.

***What are some hopes you have for the future of women in statistics?***

**EF:** My hope, I guess more generally, is that our strengths that come from being women are embraced and valued.

**HV:** I would like to see more women not only in statistics, but other STEM majors as well. I hope that we would be stronger and believe in ourselves more.

**TT:** To be leaders not just in the field of statistics but also in our community through outreach.

**SL:** More encouragement and support for young girls interested in statistics and more representation in grad schools and the field in general. I hope that we can all stand together and ultimately bridge the gender gap in STEM fields.

**Works Cited**

[Photograph of Zoe Vernon]. (n.d.). Zoe Vernon.
[Photograph of Shun Li]. (n.d.). Shun Li.
[Photograph of Corrine Elliot]. (n.d.). Corrine Elliot.

**Florence Nightingale (F.N.) David**
*By Lauren Pitcher, Communications Manager*

*F.N. David*

Florence Nightingale (F.N.) David was an internationally acclaimed statistician and professor emeritus in the Department of Statistics at UC Berkeley. David's research resulted in advances in combinatorics, including a clear exposition of complicated methods. She studied the correlation coefficient and computed solutions of complicated multiple integrals using the distribution of the correlation coefficient.

She is the author of 10 books, two monographs, and over 100 papers in scientific journals. Many of these are still actively referred to today. David's research and innovation, and trailblazing as a woman in statistics, were highly regarded. This is evidenced by her numerous awards and accolades. She was elected as a Fellow of the American Statistical Association and a Fellow of the Institute of Mathematical Statistics. She won the first Elizabeth L. Scott Award for her efforts in opening the door to women in statistics, and in 2001 the Florence Nightingale David Award was established in her honor.

David once had this to say about the state of affairs of women in academia, "I always say, 'Well, in my day you had to do 200% better than the nearest man they could possibly give the job to before you got it.' Now I think it's come down to about 150%. But I still think there is prejudice. There always will be."

Born in England, she began making regular visits to the United States in the late 1950s, principally as Visiting Professor and Research Statistician at UC Berkeley with the Department of Statistics, and Applied Climatology and Forestry Divisions. She permanently moved to the United States in the 1960s to become a professor. In 1970 she was selected as the chair of the Department of Statistics at the University of California, Riverside, and was the book review editor for the journal Biometrics for four years.

After retiring from UC Riverside in 1977 she moved to Berkeley where she continued to be an active Professor Emeritus and Research Associate in Biostatistics until her passing in 1993. Although David's contributions to statistics will be remembered for a lifetime, she was not focused on the influence of her work. When asked if she was perceived as influential during her time at

UC Riverside she said, "Oh, I think I was. But I don't think my job in life is to be influential. Rather it was her job in life to "ask questions and try to find the answers."

David was born in Ivington, England, and named for Florence Nightingale, the English founder of modern nursing and a friend of David's parents. David's mathematical training began at an early age in Ivington She previously shared, "I was born in 1909 which would make me five when the first World War broke out, and we lived in the country. I went to do private lessons with a British parson who said, 'Well, you'll of course have to know arithmetic so you better start on algebra. And you can speak English so you'd better start on Greek and Latin.' So I learned Greek, Latin, and algebra." She died in 1993.

She received her BS in Mathematics in 1931 from Bedford College for Women. Shortly afterward, she became a Research Assistant to Karl Pearson at University College, and then she was appointed Assistant Lecturer in the Statistics Department, University College, London. She went on to receive her doctorate in Statistics from University College.

~ Lauren Pitcher

**Works Cited**

"A Conversation with F.N. David," Statistical Science
"Florence Nightingale David," Wikipedia
"Dr. Florence N. David, Statistician, Dies at 83, New York Times
[Photograph of F.N. David]. (n.d.). UC Berkeley Statistics. https://statistics.berkeley.edu/

**Patricia Hardy**
*By Lauren Pitcher, Communications Manager*


*Patricia Hardy*

Growing up in the small rural town of New Edinburg, Arkansas, Patricia Hardy never imagined that one day she would work for the top public university in the nation. As a child, Patricia and her brother would watch their dad plow their fields behind a mule. Their mother stayed home to take care of the household. Patricia's parents instilled the values of hard work, family, and education; lessons she would later instill in her own children. The Great Migration during the Dust Bowl of the 1930s led Patricia and her family to Oakland where she would meet her husband, start a family, purchase a home, and begin her career at UC Berkeley.

In 1967, Patricia decided to transition from her work in the banking industry to apply for a role at UC Berkeley. She had heard that Cal would be a great place to build a long-term career. Her first role on campus was in the Graduate Division. Then she moved on to Berkeley Summer Sessions before eventually securing her career role in the Department of Statistics as a clerk/senior typist. One of Patricia's fondest memories from her career in Statistics was with Professor Elizabeth "Betty" Scott.

"Professor Scott took us to Evans Hall in the 1980s to ask us (staff) to look for changes to improve the building. Because of Professor Scott's vision to include staff in that tour of the Evans Hall construction, changes were made to include women's restrooms on every floor and at least two restrooms that included interior rest areas with a small sofa," she recalled. "Scott knew it was important to have women's restrooms on every floor. From this, I learned there were many ways to contribute to campus, even as a staff member."

The Statistics faculty provided immeasurable support to Patricia throughout her career. She recalls never feeling like her role was less important than faculty and that she was a valued member of the department. "Deb Nolan, David Blackwell, Jerzy Neyman, David Freedman, and Erich Lehmann were some of my biggest supporters."

In 2003, Patricia earned her BA in Sociology from UC Berkeley, fulfilling her lifelong dream of becoming a college graduate. "I was inspired by our children. My husband and I had a dream to put our children through college without any student loan debt and we did it. After our kids completed college, I decided it was time for me to pursue my education." Just three years later in 2006, Patricia retired from the department; commemorating a 39-year career with UC Berkeley.

“I am so happy I worked for UC Berkeley. It was the best thing ever! I attribute my connections to the community to UC Berkeley and Statistics.”

Patricia continues to reside in Oakland as the proud mother of three adult daughters and six grandchildren. She enjoys spending her time volunteering in the community. She supports a variety of organizations such as the Reclaim Your Vote Campaign, Oakland Promise, and she serves as the president of the Oakland Women’s Rowing Club.

**Works Cited**

[Photograph of Patricia Hardy]. (n.d.). Patricia Hardy.

**Deborah Nolan**
*By Lauren Pitcher, Communications Manager*

*Deborah Nolan*

Deborah (Deb) Nolan is a Professor of Statistics and the Associate Dean for Undergraduate Studies in the Division of Computing, Data Science, and Society at UC Berkeley where she holds the Zaffaroni Family Chair in Undergraduate Education. Nolan's work has contributed to the advancement of statistics and data science education as evidenced in the four books she has authored highlighting the practice of teaching statistics with case studies and hands-on problem solving. Nolan is a fellow of the American Statistical Association and the Institute of Mathematical Statistics. She has served as the Statistics Department Chair for five years and Associate Dean for Math and Physical Sciences for nine years.

Nolan's academic career has been filled with a number of firsts. As an undergraduate, she attended Vassar College in the early days of the campus becoming co-ed.

"It was an exciting time to be at Vassar because men were first admitted four years previously and they were going to graduate that year," she said in a previous interview with the Journal of Statistics Education. "There was a lot of buzz in the news and on campus about it. There was a strong sense on the campus of the importance of women's education and people wondered how this might change now that Vassar was co-ed. I found it exciting to be a student in this environment."

After completing her Ph.D. in Statistics at Yale University, Nolan began her teaching career at UC Berkeley. It was not until after a few years into her role as an assistant professor that she learned she was the first female professor hired in Statistics since Betty Scott was hired in 1951.

"For me, the transition from being a student to an assistant professor was not easy. My colleagues at Berkeley were always supportive of me and my career, but being the only woman in the department was difficult," she said. "The connections that colleagues make over the tennis court and soccer field didn't happen for me. I felt this so keenly that I wrote an article on the topic, called "Women in Academe: Mentors Matter" (Nolan 1990). That's why I co-founded the IMS New Researchers Meeting in Statistics and Probability in 1993."

Nolan was awarded the Berkeley Distinguished Teaching Award for excellence in teaching and is noted for working with and encouraging all students in STEM. She helped design the Data Science Major and develop and teach Principles and Techniques of Data Science (Data 100). She has created and led programs to encourage students to pursue their education in STEM, including the Summer Math Institute (1991-97), Explorations in Statistics Research (2005-12), CalTeach (2006-20), and Berkeley Unboxing Data Science (started in 2020). Her pedagogical approach connects

research, practice, and education, and she is co-author of four textbooks: Stat Labs, Teaching Statistics, Data Science in R, and Communicating with Data.

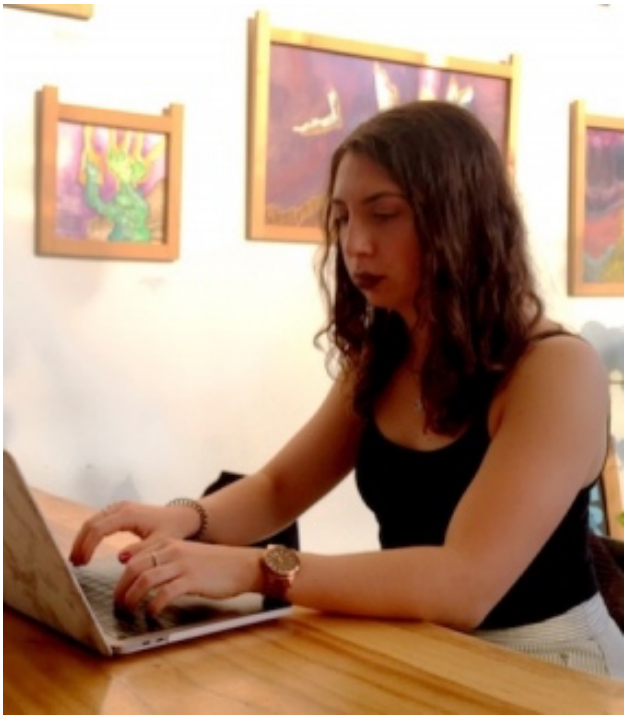
After teaching for over 30 years, Nolan is now focusing on preparing the next generation of students for a successful undergraduate career at Berkeley.

“One piece of advice I would offer someone interested in statistics and the prospect of teaching is that statistics is a field in which it takes a long time to develop expertise and to be a good teacher they need to work at developing this expertise,” she said. “I would encourage them to be lifelong learners in the field, to try their hand at data analysis, and learn how experts approach statistical problems.”

**Works Cited:**

All quotes were taken from Nolan’s previous interview with the Journal of Statistics Education, 2015.
[Photograph of Deborah Nolan]. (n.d.). Deborah Nolan.

**Kellie Ottoboni**
*By Lauren Pitcher, Communications Manager*



*Kellie Ottoboni*

Growing up in the Bay Area Kellie Ottoboni always knew she wanted to follow in her parent's footsteps to attend UC Berkeley. It was in high school that she developed her love of math as she found that it explained every physical process in the world. Kellie's childhood dreams led her on a path to receive her B.A. in Applied Mathematics and Statistics and a Ph.D. in Statistics both from UC Berkeley.

As a Ph.D. student, Kellie was attracted to data science at the early onsight of its rising popularity as an emerging STEM field. She saw the potential of data science and all the ways data could be used to impact society. Professor Philip Stark became Kellie's Ph.D. advisor guiding and mentoring her throughout her doctoral journey.

Kellie was drawn to his diversity of research interests. His research was directly aligned with Kellie's interests. He focused on a variety of topics and social issues such as the U.S. census, election auditing, earthquakes, the lottery, student evaluations of teachers, and more.
Kellie describes her Ph.D. work as the "intersection of Social Good and Statistics." She focused on data that specifically impacted people such as gender equality and health making an impact on well-being.

Kellie's path to discovering her passion for social good and statistics was not necessarily linear. She learned through self-discovery that her career path needed to be one that allows her to make an impact in the world.

"To feel fulfilled, I wanted my work to directly impact the lives of others, even if it's in small ways." She worked with Professor Stark to identify research that intrigued her. "I had to find my voice to say no to things I didn't want to do and find the courage to dive into new problems."

During her doctoral journey, Kellie was also a Fellow at the Berkeley Institute for Data Science (BIDS).

"BIDS was the greatest part of my Ph.D. I was exposed to a diverse community outside of Statistics," shares Kellie. "This was where I became excited about open source software and I had my first experience working on open-source Python packages. At BIDS I saw the broader impact

that data can have and I learned to understand the fundamental differences between data science and statistics."

Today, as a data scientist, Kellie is interested in causal inference, experiments, hypothesis testing, and the nuances of data about people. She is currently working on Data and Experimentation at Pinterest where she leads a team that focuses on experiment quality.

"My role is really focused on oversight and governance. Any change to the Pinterest app goes through an experimental process," explains Kellie. "When I joined the team, it became my job to figure out if Pinterest's experiments were being designed optimally to get the best results possible."

Looking to the future, Kellie is excited about what the future holds for data science as it permeates all facets of society. People want to make informed decisions such as understanding all the information around COVID-19 or even the upcoming election but don't know where to turn.

"With data science, we are looking at what data do we need, what are the problems with the data, and technically how do we acquire the data through the pipeline. Data intuition is so valuable in data science."

[Photograph of Kellie Ottoboni]. (n.d.). Kellie Ottoboni.

**Ashia Wilson**
*By Lauren Pitcher, Communications Manager*

*Ashia Wilson*

Ashia Wilson will join the MIT faculty as an assistant professor of Electrical Engineering and Computer Science (EECS) in 2021. She received her BA from Harvard University with a concentration in applied mathematics and a minor in philosophy, as well as a Ph.D. from UC Berkeley in statistics. Ashia held a postdoctoral position in the machine learning group at Microsoft Research, New England. Her journey comes full circle as her mother, Carol Espy-Wilson became the first Black woman to earn a Ph.D. in Electrical Engineering from MIT in 1987.

"When I was growing up, my mom was one of very few examples that helped conceptualize that it was possible to be in these spaces. However, she was such an ever-present example that it really helped me along the way," reflects Ashia. As a new professor, diversifying the STEM field is at the forefront of her mind.

"I'm hoping that as we start penetrating these spaces that it (diversity) will change. I think a large part of retention is seeing yourself in these spaces. Diversifying at the faculty level is in part the reason why I decided to go into academia rather than industry because it will allow me to mentor students of color with whom I can relate to being a person of color in STEM."

Ashia's research focuses on the methodological foundations and theory of various topics in machine learning. She's interested in developing frameworks for algorithmic assessment and providing rigorous guarantees for algorithmic performance.

"I've now pivoted a little bit from classical optimization into thinking through issues in quantifying uncertainty. I've done some work that looks at cross-validation, which is a very fundamental statistical tool.

I've also started thinking through issues of fairness which is growing as a field in machine learning. This includes thinking through the decision-making pipeline in machine learning, including how problems are formulated, data is collected, and analyzed and how errors are treated, all with issues of power and justice in mind."

Ashia grew up on the MIT campus with her parents serving as assistant head-of-house in a dormitory there for several years.

“I always had this idea as a child that maybe one day I could be a part of the MIT community in a meaningful way. I’m definitely excited to begin my career there.”

**Works Cited:**

[Photograph of Ashia Wilson]. (n.d.). Ashia Wilson.