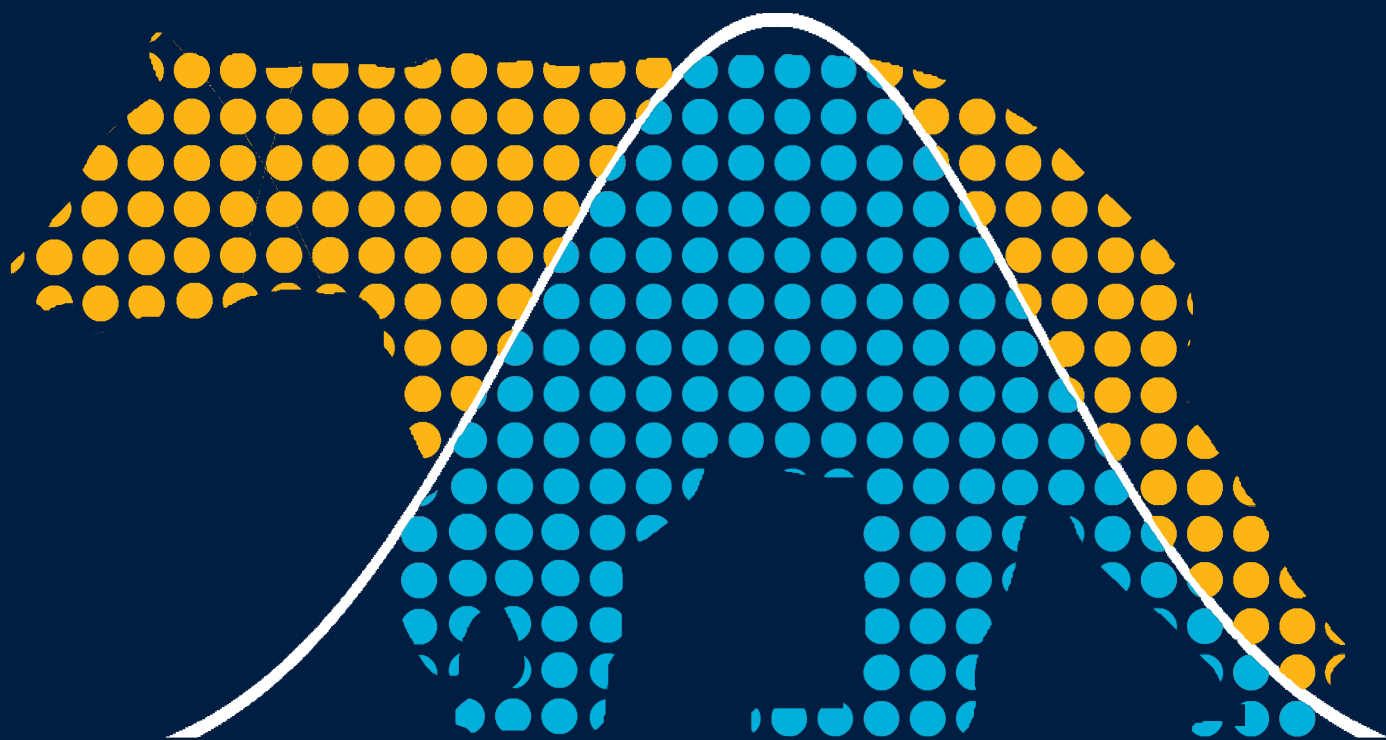


# STATISTICS



## STATE OF THE DEPARTMENT

with Chair, Deborah Nolan

## FEATURED FACULTY RESEARCH

Prof. Haiyan Huang shares some of her recent research

## CREATION OF CONSULTING

Juliet Shaffer's impact on the Department



# Table of Contents

- 3** State of the Department
- 4** The Implications of Data Science
- 6** Featured Faculty Research
- 8** Creation of Consulting
- 10** DataFest @ Berkeley
- 12** New Faculty in Statistics
- 14** Faculty Awards
- 15** Spring 2017 Graduation
- 16** Hearing from our Alumni
- 18** News from the SUSA and SGSA
- 20** Statistics Staff Update



As I begin my second stint as Chair of the Statistics Department, I am keenly aware of the progress and changes in our field. Our department is staying true to its roots and also remaining at the forefront of these innovations. As our Spring 2017 commencement speaker Steve Stoute told our graduates, “let the numbers help you tell your story. Create businesses, create hope, create progress. Data is the instrument, now go be the artist. Keep pushing culture forward.” This is exactly what our department is committed to doing.

# State of the Department

*Article by: Chair Deborah Nolan*

We have been deeply involved in the creation of a data science major. Professors Adhikari, Huang, Jordan, Purdom, Stark, and Yu have been instrumental in building the curriculum and designing courses for the new major. This new degree reflects the interdisciplinary nature of data science, and our faculty have ensured the major has a unique and strong intellectual merit by taking an integrated approach to computational and inferential thinking.

Beyond this, a Division of Data Science is moving forward to become a fully established administrative unit within the University. In my regular meetings with Interim Dean David Culler, we continually revisit the need for collaboration across campus in order to achieve the goal of supporting a broad interdisciplinary research program. As statisticians, it is in our nature to collaborate with researchers from other disciplines, and I envision our department’s outlook in this regard will be instrumental to the new division’s success.

Amidst the launch of a data science undergraduate major and the creation of a new division, our statistics program continues to grow. Our classrooms are bursting at the seams— 225 students received a Bachelors degree, 23 were awarded Master’s degrees, and 7 got their Ph.D.s in 2017. We have experimented with innovative ways to teach our classes, including developing online curriculum and offering cloud computing environments for students to analyze data, perform simulations, and more broadly learn statistics. These changes have been possible in part due to the innovations of our newest faculty member Fernando Perez, the creator of Project Jupyter. His open-source interactive computing environment allows thousands of students to engage with statistics in ways never imagined.



As we continue to evolve, our faculty continue to be recognized for their stellar accomplishments. Nicholas P. Jewell recently was elected to the National Academy of Medicine and is the first biostatistician bestowed this honor. Nike Sun was awarded the Rollo Davidson Prize, a prestigious international award made to a young probabilist. Michael Mahoney, Bin Yu, Mike Jordan and Fernando Perez were awarded a TRIPODS grant from the National Science Foundation to launch a new center aimed at addressing fundamental questions in the theoretical underpinnings of data science.

As I lead the department into this next chapter I am excited to support our students, staff and faculty as we grow and innovate together. I am also grateful to our alumni and friends who support the learning and discovery that takes place in statistics.

# Data Science: opportunity or threat, tipping point or nothing new?

*Article by: Professor Michael I. Jordan*

Questions such as these have enlivened conversations in statistics departments throughout the world during the past few years, and Berkeley is no exception. Many of our faculty meetings have been devoted to these questions, and many of our faculty have participated on campus-wide committees convened to discuss the campus-level response to data science. These discussions have included educational and research initiatives, and, most strikingly, far-reaching consideration of possible structural change at the university to respond to the implications of data science.

The result of this effort has been that the department has come solidly down on the side of “opportunity” and “tipping point.” Although many of the issues present in data science are part of the long tradition of statistics, it’s also true that there are many new issues present in computing-oriented fields, such as databases and distributed systems, that are essential to the larger data science effort, and to date have not been part of every-day statistics. Also, the data science effort brings in ethical issues, economic issues and legal issues that engender new collaborations. Overall, many campus communities feel that they are part of data science in ways that they have not been part of statistics. We view this as an opportunity to realign our department within the university, bringing us closer to data-oriented activities across a wide swath of science, technology and the humanities.

It is also important to emphasize that the term “data science” is meaningful in industry, where there has been a surge in recruiting for jobs that emphasize skills in both computing and statistics. We have accordingly found ourselves in closer contact with industry, via our growing Industrial Alliance Program, via internships for our graduate students, and via individual faculty outreach, consulting and joint research projects.

Moreover, we have made major efforts to nourish data science in our undergraduate curriculum, designing and teaching several innovative new classes, and working with colleagues in other departments (most notably Computer Science) to design a new academic major in data science. Our efforts began in earnest in the fall of 2015, when we piloted a new, freshman-level class titled “Computational Thinking and Inferential Thinking: Foundations of Data Science.” This course reposes on a thoroughgoing integration of computational and statistical ideas, accompanied by a mandate for working with real-world data sets. Students learn to program (in Python) and they learn inferential principles that are based on resampling ideas and visualization. They write programs that allow them to understand inferential ideas and to deploy them in data analysis. This course has been a major success (reaching an enrollment of nearly 1000 students this fall!). We have built on the



success of the class to design other data science classes at both lower-division and upper-division levels; these courses will be part of the new major which we expect to be announced officially during the coming year.

This major is hotly anticipated by our students— they are excited by the power of combining computing and inference to solve real-world problems that are personally interesting to them, and they are clearly aware that there are exciting careers to be found in data science.

We have begun to seek outside financial support for our curriculum-building effort in data science. While the university has been morally supportive, the university is facing major financial challenges and is not able to underwrite our efforts (particularly when our efforts lead to classes of size 1000!). Given that the modern education in data science that we are providing will allow our students to prosper, and to lead, in scientific, industrial and governmental sectors, we are hoping for increased support from these sectors. We now have a new Interim Dean of the academic division whose mandate is to support data science (Interim Dean David Culler), and we are vigorously pursuing outside support in collaboration with Dean Culler. We are hoping that departmental alumni who will help us forge new contacts in this regard. If you are interested in getting involved, please contact the department.

In summary, Berkeley is leading in the effort to make data science a disciplinary reality. We believe that data science is not merely a trend; rather, it is a transformational force in society and in academia. Within statistics, it is having a major impact on our theory and practice, on the breadth and scope of our collaborations with other fields, the style in which we work, the job opportunities for our students, and the content of our courses. But data science is having this kind of impact on other fields as well, and thus we find ourselves, as never before, in the midst of a far-flung interdisciplinary dialog with major societal implications.

# JUPYTERHUB in Statistical Education at Cal

*Article by : Professor Fernando Pérez & Ryan Lovett*

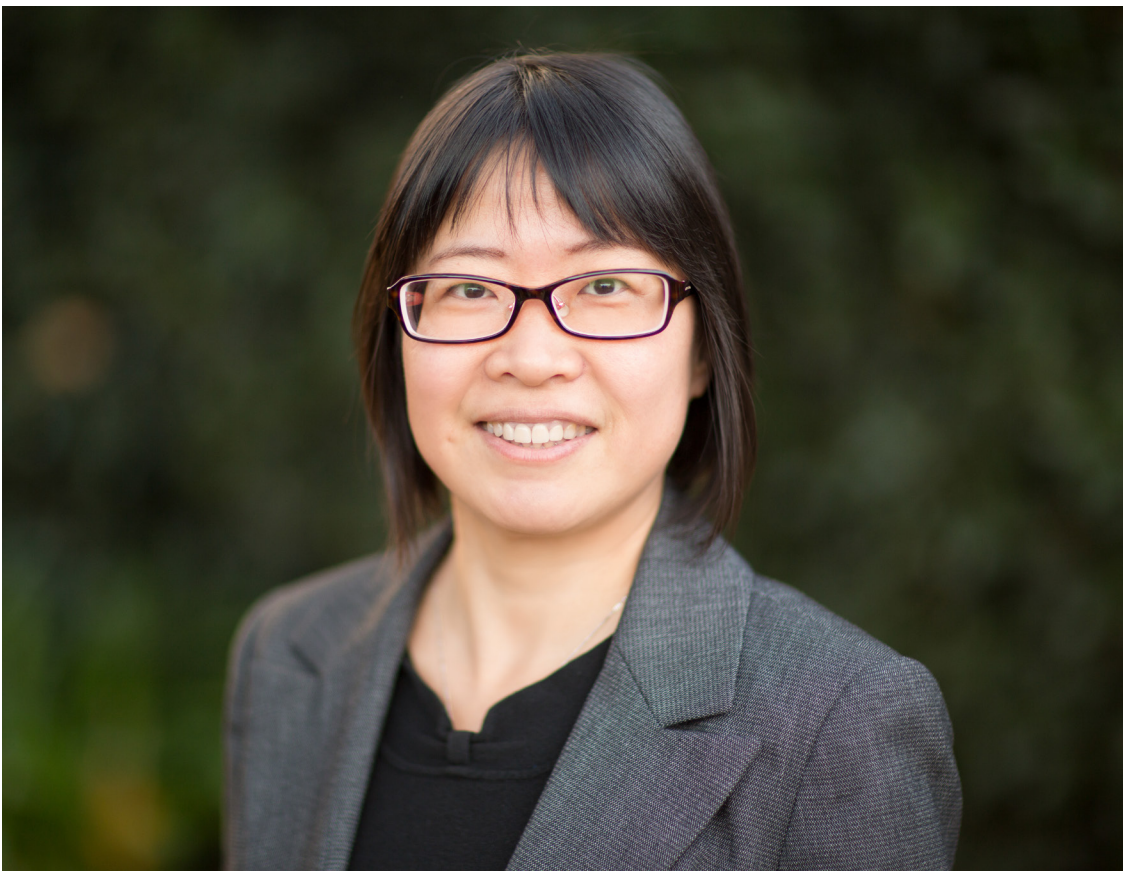
The Jupyter Notebook has become a popular teaching tool in statistics and data science courses where instructors can mix live Python or R code, explanatory text, and visualization in an easy-to-use web application. While the notebook software can be installed on each student's personal computer, the hosted, multi-user service called JupyterHub can be installed centrally on a single powerful computer or cluster, providing everyone with identical, reproducible environments. JupyterHub has been deployed for the Data Science Education Program courses for the past two years as well as for Stat 28 and 140. The Haas School of Business hosted a JupyterHub instance for Data 8 in Summer '17, and Berkeley's Research IT is experimenting with JupyterHub on part of the Savio cluster.

# The Recent Research Endeavors of Professor Haiyan Huang

*Article by: Professor Haiyan Huang & Edited by Erin Blanton*

Professor Haiyan Huang's research activities have focused on identifying, formulating and resolving important, yet not previously addressed, statistical questions from the frontiers of biology, medicine, and other scientific domains in order to assist in making novel scientific discoveries. In this research feature, a few of her research accomplishments will be highlighted.

Since the inception of RNA-Seq technology, various attempts have been made to utilize RNA-Seq data in assembling full-length mRNA isoforms de novo and estimating abundance of isoforms. However, for genes with more than a few exons, the problem tends to be challenging, and often involves identifiability issues in statistical modeling. As a key member of a research group, Professor Huang's first effort in this application was to develop a statistical method called Sparse Linear modeling of RNA-Seq data for Isoform Discovery and abundance Estimation (SLIDE) that takes exon boundaries and RNA-Seq data as input to discern the set of mRNA isoforms that are most likely to present in an RNA-Seq sample. This work appeared in the Proceedings of National Academy of



Science USA (2011). Professor Huang co-lead this project with fellow UCB Statistics Department faculty member Professor Peter Bickel and help from their joint student Jessica J. Li who was the first author of this work. Their method has made a big impact to the computational genomics community.

Developing methods for finding complex gene relationships, under careful

considerations of how genes behave across the given samples, has been a long-term research interest for Professor Huang. This is an important component in biological network analysis, and is especially challenging when the gene relationships go beyond linearity or only present locally. In the *Annals of Applied Statistics* (2015) paper, with her student Rachel Wang and their collaborators, they focused on gene group interactions and developed a new measure to infer the group relationships using appropriate partial correlations between genes, that is, the conditional dependencies between genes after removing the effects of a set of other functionally related genes.

The measure was defined based on sparse canonical correlation analysis (SCCA) coupled with repeated random partition and subsampling of the gene expression data set. By considering different subsets of genes and ways of grouping them, their new measure, which can be viewed as an aggregated estimate of partial correlations of different orders, is unique in evaluating conditional dependencies when the correct dependent sets are unknown or only partially known. Since publication, they have received many requests of source code for this approach. Furthermore, in another effort published in the *Proceedings of the National Academy of Science* (2014), their group reported two new count statistics to assess gene relationships especially for time-course data with local dependence structures (e.g., time series coupled over a sub-region of the time domain). This work involves interesting application ideas related to bivariate associations and count statistics. The asymptotic studies of the two new statistics have made nice contributions to the combinatorics literature. In addition, they recently extended this work to identify associations between expression profiles consisting of ordered samples, but the samples are not aligned and contain local association patterns possibly separated by gaps. This method has led to many new applications such as assessing functional relevance between genes from different species or from different phenotypic contexts. This work just appeared in *Bioinformatics* (2017).

Single cell RNA-Sequencing (scRNA-Seq) is becoming a technological standard in life science research.

Single cell RNA-Sequencing (scRNA-Seq) is becoming a technological standard in life science research. Focusing in on the single cell level allows researchers to investigate the meaningful and illuminating heterogeneity among cells of interest and to discover cell-based biologics, e.g., to identify cellular sub-populations and rare cell types. A key component to identifying cell subpopulations is the (dis)similarity measure used to group the cells. In the *Statistics in Biosciences* (2017) paper, Professor Huang and her student Courtney Schiffman, together with collaborators, introduced a novel measure, named SIDEseq, to assess cell-to-cell similarity using scRNA seq data. The idea was to incorporate information from all cells when evaluating the similarity between any two cells, a characteristic not commonly found in existing (dis)similarity measures. When used on a scRNA-Seq dataset generated by their collaborator Professor Sohn's lab, SIDEseq uncovered some important relationships between cells, and outperformed several popular (dis)similarity measures.

This feature has focused on some of the most recent research projects of Professor Haiyan Huang. As an applied statistician, Professor Huang has about 15 years' experience with data. She is a valuable part of the UC Berkeley Statistics Department and we hope you have enjoyed reading this snapshot of her research.

This feature has focused on some of the most recent research projects of Professor Haiyan Huang. As an applied statistician, Professor Huang has about 15 years' experience with data. She is a valuable part of the UC Berkeley Statistics Department and we hope you have enjoyed reading this snapshot of her research.



# Juliet Shaffer & the Creation of Consulting

Article by: Sara Stoudt

Before the start of the Fall 2017 semester Ph.D. student Sara Stoudt sat down with teaching professor emerita Juliet Shaffer to discuss Professor Shaffer's time in the department and learn more about how she came to create and develop the popular consulting class for the Statistics Department.

SS: What was your path to becoming a statistician?

JS: I always liked mathematics, but in those days women were considered not appropriate for mathematics. I didn't want to major in mathematics because I felt that it was too theoretical, and I wanted to do something that affected the world. I didn't know anything about statistics at that time. I was first a chemistry major in college, but switched to psychology. I still took mathematics along the way though. In graduate school at Stanford, I took mathematics and statistics classes while getting a Ph.D. in psychology. My first job after graduate school was at the University of Kansas where psychology and chemistry were the two liberal departments that would hire women. Since I had more of a mathematics and statistics background than most of the psychologists, I ended up teaching the undergraduate and graduate statistics courses. I got more and more interested in understanding the material better. When I finally got a sabbatical, I came here [Berkeley] because I was using a book by Erich Lehmann as a reference, so it seemed a good place to go. A few years later, I got an offer from Berkeley to be a lecturer and then worked up to being a senior lecturer.

SS: When you started at Berkeley, what was the department like? What changed?

JS: The department was much more theoretical then. I came in with a lot of consulting experience, and Peter Bickel proposed setting up the course.

SS: Was there a particular motivation for starting it or did it happen serendipitously?

JS: It just sort of happened. When Peter proposed it, I said yes. I fearlessly wandered into consulting. I think back, and say "fearlessly," because I didn't know that much statistics at the time, but I had a sense of the problems as well as common sense about how to treat problems and discuss them with the clients.

SS: Who participated in the consulting class when it started?

JS: The students were both MA and Ph.D. students. We first set up consultants in pairs, because I wasn't sure how many clients we would get and didn't want to plan too many consulting sessions. Students liked being together because they could rely on each other, so we kept that arrangement. I said I would only teach the course if the service was open to everybody



Photo provided by Juliet Shaffer

including town people, who came occasionally. The clients were mostly graduate students in other departments.

SS: What do you think is important to be an effective consultant?

JS: Listen hard. Check the design carefully. Ask lots of questions, don't just jump into giving advice. When you give advice don't be too complicated. Sometimes it is very hard to know what to ask. We had one client who said there were 150 subjects in his experiment. As he was leaving he said, "Oh, by the way, I don't think it matters, but these were 75 couples." When a study involved a number of subjects in several groups, the consultants sometimes didn't ask whether these were the same subjects in each group or independent groups, clearly very important for analysis. We once had someone whose position in the University was to attend classes and give advice to the faculty and graduate student instructors; she came to one of our two-student consulting sessions. She had two valuable suggestions: (1) ask the clients what they would do if they couldn't get any advice— this would help us understand what they were looking for, and (2) at the end, go over the advice you have given to be sure it is understood properly.

SS: What was the most challenging or exciting topic you saw in the class?

JS: I think we were some of the first people outside of the biological sciences that heard about DNA. Students would come in and teach us about the four letters and their problems involving sequences of the letters and how to match them. I thought, 'Wow, this is fascinating.' There was also an interesting case when a guy was studying grasshoppers. He had lots of groups of grasshoppers and was measuring how many of certain kinds of events occurred. But he couldn't tell the grasshoppers apart. We didn't know whether it was the same grasshopper doing it all or how the events were spread out, so we had to figure out how to analyze it without knowing that information. He had many interesting problems, so he kept coming back. At the end of the semester, he invited us... *continued on pg 23* for being so helpful, to a dinner he was giving in his dorm. He gave us a very fancy French menu and every course was some insect. I couldn't bring myself to go, but I hoped some of the students would go. None of them did. There was an article about it in the Daily Cal though. The thing I remember most was that one woman was running around asking everyone, "Are worms red meat? Because I don't eat red meat."

SS: Do you have any advice for students who are thinking about taking the course?

JS: Take the course! Every time I see former students, they stop me and say that the course really helped them in their careers. It is different from the rest of the classes, and you will get something out of it. In typical classes you have a topic and are given related problems. Here you get problems and you have to decide what to do, which can be challenging yet fun.

## Quotes from Students who Took Consulting:

"For many of us, the consulting class is the first time that we are really exposed to statistical problems in the wild, and is often the first chance we get to really apply what we have learned in our classes. We get to practice both statistical reasoning on real-world problems as well as our ability to communicate statistical methodology to individuals with a less technical background. It is really an eye-opening experience!"

"I learned a lot from the consulting class by understanding the clients' expectations of a statistician. A lot of things they used are unfamiliar to me which motivates me to explore those areas."

"I'd say I got three main things out of the class. I got to see how scientists actually use statistical methods, and what kinds of things they want to get out of the process. I gained a better understanding of the practical advantages and disadvantages of some common techniques like random effects, thanks to Elizabeth's wisdom. And I got the satisfaction of actually helping some fellow grad students!"

"The consulting class was the first opportunity for most students to see the gulf between statistics as taught in theoretical courses and the reality of applications pursued in the real world. The difference was stark! Everyone, apart from Neyman himself (who, legend suggests, only looked at the data after deciding precisely which test and design to use) seems to engage in data snooping and post-hoc analysis: things that have only recently become possible to do in a defensible and rigorous way. Modern theory really is important!"



# Running the Numbers: D

Article by: Karl Kumblier

Right after coming back from their spring break, 64 undergraduates gathered at the Berkeley Institute for Data Science to participate in DataFest, a weekend of data exploration and analysis put on by the American Statistical Association in partnership with the UC Berkeley Department of Statistics. Like a hackathon for statistics, DataFest provides an opportunity for teams of undergraduate students to work with big, messy, real-world data to derive and communicate insights to industry sponsors. Participants were mentored by graduate students, faculty, alumni, and industry experts as they worked around the clock to identify travel trends from a database containing hundreds of features on millions of Expedia customers.

When do people book their vacations and where do they travel to? What are the popular activities in different travel destinations? Can customers be grouped in ways that reflect purchasing behavior? How do country-level features such as GDP and national holidays relate to travel patterns? These were just a few of the



“DataFest provides an opportunity for teams of undergraduate students to work with big, messy, real-world data to derive and communicate insights to industry sponsors.”



# DataFest @ UC Berkeley

questions that students sought to answer over the course of the weekend. To investigate these questions, participants were armed with unprocessed data from Expedia, statistical/data science techniques they had learned throughout their coursework, and any other relevant data they could find. Even more impressive than their results was the fact that students formulated the motivating questions for their analyses from only a brief description of Expedia's goals and individual conversations with company representatives. At the end of the weekend, participants had 5 minutes and 3 slides to summarize their findings for a panel of judges, many of whom were alumni of the PhD program. Awards were presented to the teams for the best overall presentation, best data visualization, and best use of external data.

This is the second year that the Berkeley Statistics Department has participated, joining over 50 universities from around the world in this celebration of data, and we are looking forward to continued participation in the future. The event provides a chance for students to develop practical skills for working with real-world data, receive valuable feedback from individuals with years of experience in statistics/data science, and network with industry professionals who come to recruit promising young talent to their organizations.



*Photos: Top- Students working on data set at DataFest*

*Bottom left: Presenting results Eric Kim, Lydia Maher, Arun Ramamurthy, Alexander Lee, Patrick Chao*

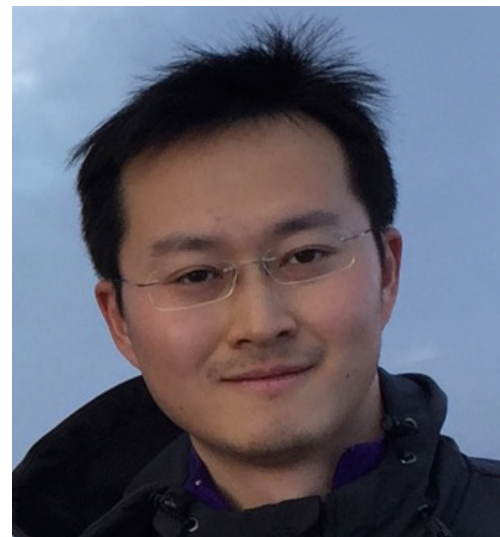
*Bottom Right: Best in Show Young A. (Irene) Ha, Brian Lee, & Minsu Kim*

# New Faculty in the Statistics Department

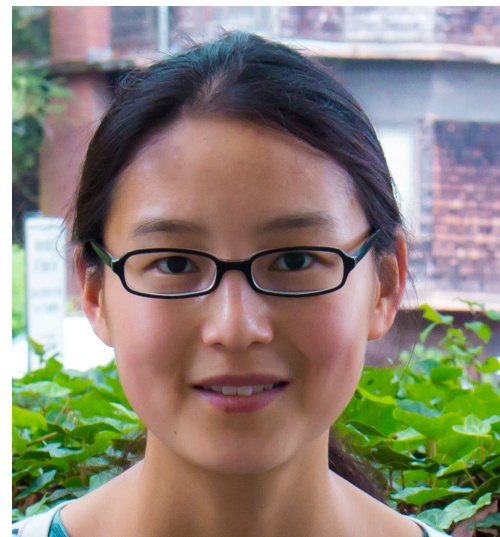
**WILL FITHIAN** became a faculty member at UC Berkeley in July 2015. He obtained his Ph.D. in statistics from Stanford in 2015 under the supervision of Trevor Hastie. During his Ph.D. he also worked closely with Jonathan Taylor and Rob Tibshirani on topics including ecological modeling, machine learning, dimensionality reduction, and scalable algorithms. His dissertation work and current primary research area is in selective inference and post-selection inference, on methods that blend exploratory and confirmatory data analysis.



**PENG DING** joined Berkeley in January 2016 after receiving his Ph.D. from the Harvard Statistics Department and working as a postdoctoral researcher in the Department of Epidemiology of the Harvard T. H. Chan School of Public Health. His research mainly focuses on causal inference from experimental and observational data. In particular, he is interested in randomization-based causal inference in experiments, using the physical randomization as a basis for statistical inference without assuming any outcome models. For observational studies, he developed sensitivity analysis techniques taking into account unmeasured confounding that could possibly overturn the results based on the observed data. His other research interests include survey sampling, experimental design, and applications of statistics to social sciences and biomedical studies.



**NIKE SUN** joined Berkeley Statistics in January 2016. She was previously at Microsoft Research New England and MIT Mathematics, under a Schramm postdoctoral scholarship. She obtained her Ph.D. from Stanford Statistics in 2014 under the supervision of Amir Dembo. Her primary research interests are in probability theory and statistical physics. She is especially interested in graphical models, and associated computational problems, on large random graphs.





**SAM PIMENTEL** became a member of the UC Berkeley Statistics Department in July 2017. He received his Ph.D. from University of Pennsylvania. His research focuses on methods for measuring causal effects in large observational datasets. He develops new designs for matched treatment-control studies, using tools from discrete optimization to balance computational and statistical demands. He is also interested in randomization inference and in related methods of sensitivity analysis that test the stability of causal inferences under unmeasured biases. He enjoys working on data with scientific collaborators, and has applied causal methods to problems in health outcomes research, education, and political science.



**FERNANDO PEREZ** became an Assistant Professor in Statistics at UC Berkeley July 2017. In addition, he is a Faculty Scientist in the Department of Data Science and Technology at Lawrence Berkeley National Laboratory. After completing a Ph.D. in particle physics at the University of Colorado at Boulder, his postdoctoral research in applied mathematics centered on the development of fast algorithms for the solution of partial differential equations in multiple dimensions. Today, his research focuses on creating tools for modern computational research and data science across domain disciplines, with an emphasis on high-level languages, interactive and literate computing, and reproducible research. He created IPython while a graduate student in 2001 and co-founded its successor, Project Jupyter. The Jupyter team collaborates openly to create the next generation of tools for human-driven computational exploration, data analysis, scientific insight and education.



He is a National Academy of Science Kavli Frontiers of Science Fellow and a Senior Fellow and founding co-investigator of the Berkeley Institute for Data Science. He is a co-founder of the NumFOCUS Foundation, and a member of the Python Software Foundation. He is the recipient of the 2012 Award for the Advancement of Free Software from the Free Software Foundation.



# Faculty Awards

## **Peter Bartlett**

Fellow of the Australian Academy of Science (2015)

## **David Brillinger**

Bernoulli-IMS Tukey Lecturer, Toronto (2016)

## **Peng Ding**

Young Investigator Award, ASA Section of Statistics in Epidemiology, Joint Statistical Meetings in Chicago, IL (2016)

## **Steve Evans**

Member, National Academy of Sciences (U.S.A.) (2016)

## **Avi Feller**

AERA Award for Outstanding Publication in "Advances in Methodology"

## **Lisa Goldberg**

Journal of Investment Management Harry M. Markowitz Special Distinction Award for "Restoring Value to Minimum Variance" (2014)

## **Adityanand Guntuboyina**

NSF Career Award (2017)

## **Haiyan Huang**

New Research Grant Award, Chau Hoi Shuen Foundation Women in Science Program (2017)

## **Michael Jordan**

IJCAI Award for Research Excellence (2016)  
David E. Rumelhart Prize (2015)  
Challis Lecture, University of Florida (2017)  
Milne Lecture, Oregon State University (2017)  
Wilks Memorial Lecture, Princeton University (2016)  
Jon Postel Lecture, University of California, Los Angeles (2016)  
Gene Brice Colloquium, Rice University (2016)  
John von Neumann Lecture, Brown University (2015)

Statistics at Berkeley

Coxeter Lecture Series, Fields Institute for Research in Mathematical Sciences (2015)  
Bahadur Memorial Lecture, University of Chicago (2015)

## **Rasmus Nielsen**

ISI Highly Cited Researcher (2014)

## **Fernando Perez**

National Academy of Sciences Kavli Frontiers of Science Fellow (2016)  
SPOT Award, Lawrence Berkeley National Laboratory (2016)

## **Elizabeth Purdom**

Hellman Family Faculty Fund Award (2013)

## **Yun S. Song**

Chan Zuckerberg Biohub Investigator Award (2017)

## **Philip B. Stark**

Leamer-Rosenthal Prize for Transparency in Social Science (2015)  
Velux/Villum Foundation Professor of Theoretical Computer Science at the IT University of Copenhagen (2016)

## **Nike Sun**

Rollo Davidson Prize (2017)

## **Martin Wainwright**

Distinguished Lecturer, Information Theory Society (2015-2016)  
David Blackwell Lecturer, Institute of Mathematical Statistics (2017)

## **Bin Yu**

Rietz Lecturer of the Institute of Mathematical Statistics (IMS) (2016)

# Commencement 2017



## **EVELYN FIX PRIZE:**

Jonathon Terhorst

## **DEPARTMENT CITATION IN PROBABILITY (Ph.D.):**

Wenpin Tang

Yumeng Zhang

## **ELIZABETH SCOTT MEMORIAL AWARD:**

Boying Gong

## **DEPARTMENT CITATION IN THE STATISTICS MASTER'S PROGRAM:**

Béatrice Girard

## **DEPARTMENT CITATION:**

Donovan Lieu

## **OUTSTANDING GRADUATE STUDENT INSTRUCTOR**

### **AWARDS:**

Amy Bray, Jordan Brooks, Johnny Hong,  
Daniel Lee, Donovan Lieu, Jordan Prosky,  
and Shamondra Shrotriya

# Updates and Memories from You, our Alumni!

## **Katerina Kechris, Ph.D. Statistics 2003**

After completing her Ph.D. under the advisement of Professor Peter Bickel, Katerina then completed a post-doc at UCSF. Next she moved to Denver for a faculty position in the Department of Biostatistics and Informatics and the University of Colorado Anschutz Medical Campus. Currently her research focuses on omics and using them to find yet unknown connections between genetics and disease. You can read more about her research using this link to the Colorado School of Public Health's website. <http://www.ucdenver.edu/academics/colleges/PublicHealth/About/news/Pages/Newsroom.aspx?&newsid=1001>

## **Fred C. Andrews, Ph.D. Statistics 1953**

Fred completed his Ph.D. in 1953 under the supervision of Erich Lehmann. He is now 93 years old, having been retired as Professor of Mathematics Emeritus, University of Oregon since January 1989. His wife, Joyce, and he are in good health considering their ages and have been married for 73 years.

## **Donald C. Wilfong (Don), BA: Mathematical Statistics 1951**

I graduated there were just three undergraduates in the Major.

By far the hardest course I ever took was Probability Theory 165 taught by Professor Michel Loève. Many people thought it was the hardest of any course in any department at the University! (It was not required for Mathematics majors.) In the class there were the three undergraduates and about 10 others going for their Master's or Ph.D.'s. Interestingly, Loève smoked a pipe during class, and just about everyone in the class took up pipe smoking. I did not like smoking, so just "wore" a pipe during class!

Another really great course I took was taught by Professor Elizabeth Scott. The course spent a great deal of time on the problem of the lady tasting tea: Can the lady tell whether the cream was put in the cup before or after the tea was poured? I wonder if that question is still being pondered at Berkeley.

Of course there were no computers at that time, so we all spent many hours punching the keys on Marchant calculators. I can still remember their rattle as they performed division!

Amazing as it may seem today, in 1951 there were no jobs for mathematicians, and no one had ever even heard of a statistician! I was fortunate to get a very low paying job in the Mathematics Research Department at Wright-Patterson Air Force Base in Dayton, Ohio. After a year I was about to be drafted into the army, but fortunately was able to receive a direct commission into the Air Force. I then had a year at UCLA in the Meteorological Department and four more years (including a year in Korea) in the Air Force. I ended up as a technical salesman for IBM.

Since retiring I have been part of a volunteer group at the San Ramon Senior Center that helps struggling students. One of the students I tutored is in college and needs help in a statistics course. I'm finally using the knowledge acquired at Berkeley!





JSM Alumni Mixer 2016 in Chicago



JSM Alumni Mixer 2017 in Baltimore



# NEWS from The Statistics Undergraduate Student Association (SUSA)

Last year, the Undergraduate Student Association (USA) experienced record-breaking growth, with almost a 100% increase of size in both our projects committee and our general membership. We made extraordinary gains in supporting underclassmen interested in statistics, in the form of our Career Exploration committee, and our organization continues to grow in its unique spirit, culture, and goals.

This fall, in continuation of the progress made last year, USA will be going through several exciting changes. For one, heartened by the rise of data science initiatives throughout campus, we have redesigned our projects committee into a larger set of consulting services, with a focus on collaborating with other student organizations, faculty, and the broader UC Berkeley community. We are also opening up a new online publication, where our members can explore statistical projects in an open forum through the guidance of mentors and editors. Finally, we are changing our name to Statistics Undergraduate Students Association (SUSA), to better reflect our close relationship with both the Statistics department and our sister club, SGSA. We greatly look forward to these new developments next semester, and hope you do too! Go bears!



Article By: Sandra Herchen & Arun Ramamurthy Photo: SUSA 2016-2017

# NEWS from The Statistics Graduate Student Association (SGSA)

With the fulfillment of many old traditions and the introduction new ones, the SGSA had a fun and productive year. Under the leadership of co-presidents Kellie Ottoboni and Rebecca Barter, we began last year with a BBQ and a beautiful hike in Tilden Park to welcome the new first year Ph.D. cohort and to show off the splendors of the Bay Area.

Throughout 2016-2017 the SGSA organized a number of new events. Jointly with the Statistics Undergraduate Student Association we put on a graduate school panel to discuss our experiences. We held our second annual Gender Issues Round-table to discuss sexual harassment, implicit bias, and impostor syndrome. Then we transitioned the Spring Women's Lunch to an all-inclusive Diversity Lunch to broaden the discussion to include other underrepresented groups. We met regularly to teach each other useful skills. The student seminar committee held several computing talks on topics including Git, Knitr, and interactive visualizations. We started a grassroots group named Practical Statistics, where we met every other week to learn what a practitioner should know about basic statistical methods in order to be an effective consultant. On the non-academic side, we continued to battle it out for the coveted title of foosball champion at our fall and spring foosball tournaments.



(Photo left to right: Yu Wang, Stephanie DeGraaf, Bryan Liu, Koulik Khamaru, Kellie Ottoboni, Dave Graham-Squire, Rebecca Barter, & Eli Ben-Michael)

The SGSA warmly welcomed the new cohort of incoming students and the two new co-presidents, Jamie Murdoch and Sara Stoudt. This year the SGSA is piloting a peer mentoring system for incoming Ph.D. students where new students are paired with more advanced graduate students based on research interests. We hope having a point person for questions and advice will help streamline the transition process into the program.

The SGSA remains committed to continuing the progress made in discussions of inclusion by continuing the Gender Issues Roundtable and Diversity Lunch. We also hope to maintain collaboration with the Undergraduate Statistics Association, to learn from one another in more student-led seminars, and to quest for eternal foosball glory.

*Article By:* Rebecca Barter, Jamie Murdoch, Kellie Ottoboni, and Sara Stoudt



# Staff Updates and Announcements

## NEW STAFF MEMBERS



**Kristine Nera**- Academic Personnel Coordinator & Executive Assistant to the Chair

Kristine joined UC Berkeley in July 2016. Prior to coming to UCB, she worked at UCLA's School of Medicine, both in the Dean's office and the Department of Medicine for 6 years. Her undergraduate degree is also from UCLA. Kristine is excited to be back in the Bay Area, where she grew up, and to be working with the department!



**Erin Blanton**- Assistant Director of Industry Relations and Master's Program Coordinator

Erin joined UC Berkeley in February 2017. She has a Master's degree from North Carolina State University in College Counseling and Student Development. Prior to starting with the Statistics Department she was the manager of graduate student services at the Academy of Art University in San Francisco for 3 years. She is happy to be a member of UCB and the Statistics Department!

## RETIREMENT: JANE MUIRHEAD



This past June the Department of Statistics celebrated the retirement of longtime colleague Jane Muirhead. During the celebration, staff and faculty from across the campus came together to wish Jane a "bon voyage" as she starts her retirement. Jane worked with the department for nearly 30 years and most recently has been dedicated to helping Ph.D. students, post-doc, and faculty make sure their grants and awards go through smoothly. Her consistency and contributions to the department mean she will be sorely missed.

# The Statistics Industry Alliance Program

The Statistics Department's Industry Alliance Program (IAP) is a mutually beneficial partnership between industry and the department. The program establishes strong connections with companies engaged in statistical research, data analysis, and machine learning. In turn the department connects its industry members to state-of-the-art research and some of the top talent in statistics. We would like to thank our members, listed below, for their continued support of our department. Their vested interest helps us provide the best resources to our students and continue to produce ground-breaking research.

## **IAP Members:**

Adobe  
Citadel  
Menta Capital  
Microsoft  
State Street

StubHub!  
TGS  
Two Sigma  
Uber  
The Voleon Group

If you think your employer may be interested in joining the IAP please visit our website, <http://statistics.berkeley.edu/industry/IAP>. You may contact the Industry Alliance Program by email at [IAP@stat.berkeley.edu](mailto:IAP@stat.berkeley.edu).

## **Berkeley Statistics Annual Research Symposium (BSTARS) 2017**

This year's BSTARS was another success. Like previous years attendees heard the latest research developments in the department, with an emphasis on possible applications to statistical problems encountered in industry. This year's conference had a very full schedule which consisted of opening statement by then chair Professor Michael I. Jordan, 25 talks & 12 posters by Ph.D. students and Post-docs about their work, multiple presentations on industrial research by our IAP members, and a keynote lecture given by faculty member Professor Jasjeet Sekhon. The day long event gave our graduate students, faculty, and industry partners an opportunity to connect, discuss research, and review the newest developments happening on-campus and in the field of statistics.

# Help Keep Berkeley Statistics on Top

This year campus has continued to set strict limitations on the budget. This is why we depend on the interest and generosity of our donors to help Berkeley Statistics remain one of the two preeminent centers for Probability and Statistics in the world. There are several ways to donate to a variety of identified needs in Berkeley Statistics, such as graduate student support, funds for the renovation of our classrooms and facilities, support to hold department seminars, and support for the student associations.

- Go online to [www.statistics.berkeley.edu/giving](http://www.statistics.berkeley.edu/giving). This will take you to a page that provides a list of established funds. The page provides links to descriptions of each fund and a link to donating online through the Give to Cal secure site.
- Use the enclosed envelope to make donations by check or credit card. Checks should be addressed to "UC Foundation." If you have an identified need or special fund that you would like to support please note on the check. If you do not identify a special fund, your donation will go to the Friends of Statistics Fund.
- Companies interested in becoming members of the Industry Alliance Program may contact Erin Blanton, Assistant Director of the IAP, for more information. 375 Evans Hall Berkeley, CA 94720-3860 [eblanton@berkeley.edu](mailto:eblanton@berkeley.edu) 510.643.0589



*Photo:*  
Graduate  
Students at  
the Annual  
Fall SGSA  
Picnic



# Special Thanks to our Recent Donors

Dr. and Mrs. Dennis J. Aigner  
Syamala & Radhakrishnan  
Aiyar  
Professor David J. Aldous  
Ms. Mckenzie S Amodei  
Ms. Kathleen E. Bailey  
Mr. Steven Barr  
Mr. Matthew Beals  
Mrs. Irene & Mr. Charles  
Beebe  
Professor Peter & Nancy  
Bickel  
Mr. Charles E. Bloomquist  
Mr. Colin & Mrs. Valerie Blyth  
Ms. Cynthia & Andrew Boyd  
Mrs. Mary Lou Breiman  
Ms. Kristina L. Bucher  
Ms. Elizabeth M. Carlton  
Ms. Jennifer Caroff  
Ms. Felicia Chavez  
Dr. Aiyou Chen & Ms.  
Haihong Sheng  
Ms. Keling Chen  
Ms. Chien-Jung Cheng  
Vicki and Tom Cline  
Ms. Margot I. Cunningham  
Michael & Melissa de la Rosa  
Malloy  
Mr. Ron J. Dean  
Mr. Peter Der  
Ms. Tianhui Diao  
Mr. Don Dudan  
Mr. Steven Hanks Dunsky  
Ms. Sharon Eberhardt & Mr.  
Perrin Meyer  
Ms. Anja Eckert  
Mr. Steven P. Ellis  
Ms. Barbara S. Feinstein &  
Mr. William J. Devore  
Francis and Gayle Fujioka  
Mr. Junyuan Gao  
Ms. Christian Geidemann  
Mr. Lawrence T. Gurley  
Mr. Donald Guthrie & Ms.  
Candace Tkachuck  
Mr. Mark Guy  
Mr. Rod Hemphill

Tien and Margaret Hou  
Mr. Steven R. Howard  
Ms. Kelly P. Huang  
Estie & Mark Hudes  
Mr. Fletcher Ibser  
Mr. Glenn R. Ingersoll  
Mr. Rafael Irizarry  
Mr. Glenn Jackson  
Barry R. and Kang Ling  
James  
Dr. Imke Janssen  
Professor Michael Jordan  
Ms. Emily & Dr. Mahmut N.  
Karayel  
Mr. Juan & Mrs. Jo-Anne  
Kelly  
Ms. Sarah M. Kelsey & Mr.  
William F. Forrest, III  
Mr. Dennis Kenig  
Ms. Genevieve L. Klein  
Mr. Michael S. Knappman  
Ms. Rachelle A. Kresch  
Ms. Joyce Kurtz  
Ms. Carrina H. Lai  
Mr. Kuan-Cheng Lai  
Mr. Charles Lakatos  
Mr. Michael S. Last  
Mr. Nathan Lawrence  
Ms. Rujia Li  
Ms. Wei Li  
Ms. Yangxi Li  
Ms. Yao Li  
Ms. Melissa Liao  
Mr. Robert Limjuco  
Ms. Shannon W. Lin  
Ms. Heather M. Llewellyn  
Ms. Po-ling Loh  
Mr. Nicholas A. Lombardo  
Mr. James P. Long  
Mr. William Low  
Ms. Nan Lu  
Ms. Shirley W. Ma  
Mr. Jon D. McAuliffe  
Mr. Wing H. Wong & Ms. Lisa  
Miao  
Mr. Douglas & Mrs. Susan  
Milikien

Mr. Luke W. Miratrix  
Ms. Carol W Morgan  
Mr. Matthew Morrissey  
Ms. Lauren Moscoe  
Ms. Emily Moshier  
Mr. Dennis P. Moy  
Mr. Christopher Muller  
Mrs. Rosemarie & Mr. Tae  
Hea Nahm  
Dr. Ezra & Mrs. Laure Nahum  
Mr. Alan Y. Nakamoto  
Ms. Richa Namballa  
Mr. Rohan J. Narayen  
Mr. Kok Lip Ng  
Mr. Allen Noren  
Ms. Elizabeth Ruth  
Obolensky  
Ms. Elizabeth C. Oconnor  
Ms. Pamela C. Ohara  
Mr. Chris Oni  
Mr. Jason J. Ouaknine  
Ms. Suzanne S. Pan &  
Professor Ching-Shui Cheng  
Mr. John Pasquin  
Mr. Kaiyin Poon  
Mr. Rafael M. Rafailov  
Ms. Carrie Reed  
Catherine B. & Daniel Renner  
Mr. Myrrhia Resneck  
Ms. Jean & Mr. Francis C.  
Riley  
Mr. Michael J. Roberts  
Ms. Liz Salk  
Professor Paul L.  
Schmidbauer  
Mr Harry V. Schultz  
Mr. Lei Shi  
Ms. Arshia Singh  
Dr. Joan & Mr. David  
Skurnick  
Professor Philip B. Stark  
Barbara L. & Charles J. Stone  
Mr. Yalun Su  
Ms. Laurie I. Tarver  
Ms. Sally J. Thomas, M.A.  
Ms. Li Tian & Mr. Li Gan  
Mr. Keith Tsukimura  
Mr. Trevor Tyler  
Mr. Eric A. Vance  
Wilhelmine Von Turk, PhD  
Mr. Mark A. Vradenburg  
Ms. Andrea T. Wang  
Mr. Andre K. Waschka  
Ms. Monique M. Webster  
Charles and Patricia West  
Mr. Andrew P. Willis

Ms. Lilia Wu  
Ms. Li-Ping Yang & Mr. Colin  
O. Wu  
Dr. Letitia Yee & Mr. Michael  
Lii  
Ronald and Anita Yee  
Ms. Helen J. Young  
Professor Bin Yu & Mr. Ke-  
Ning Shen  
Ms. Chi Zhang  
Ms. Jody Zhang  
Ms. Nan Zhang  
Dr. Ping & Mrs. Bridget  
Zhang  
Mr. Wenchang Zhang  
Mr. Yifan Zhang  
Ms. Ying Zhang & Mr. Gary  
Ge  
Dr. Hongyu Zhao  
Ms. Zirou Zhong  
Dan Zhu, M.S.  
Ms. Ying Zhu  
Mr. Zac Zimmer

University of California, Berkeley  
Statistics Newsletter  
Department of Statistics  
367 Evans Hall  
Berkeley, CA 94720-3860

NONPROFIT ORG  
U.S. Postage  
PAID  
University of  
California, Berkeley

Cover Art: New Statistics Department  
Logo by Alum Lily Li

Designed and edited by: Erin Blanton  
Supervising Editor: Laura Slakey

## **Would you like to share a story or contribute to our next newsletter?**

*Please write to us at [eblanton@berkeley.edu](mailto:eblanton@berkeley.edu)*

We would love to include more news from our alumni. You can share anything you like; send us photos, too!