

E-Tidings Newsletter

SCASA Events and News



Volume 9, Issue 5

May 2021

IN THIS ISSUE

PAGE 2	SCASA MEMBERSHIP RENEWAL
PAGE 3	POST-EVENT REPORT: ASW
PAGES 4-5	POST-EVENT REPORT: DATAFEST
PAGES 6-7	POST-EVENT REPORT: ISEF
PAGE 8	AWARD WINNER DR. CHONG HO ALEX YU
PAGE 9	R WEBINARS WITH SASSAVVY.COM
PAGE 10	WNAR SHORT COURSES
PAGE 11	THE USER! 2021 CONFERENCE
PAGE 12	JOINT STATISTICAL MEETINGS
PAGE 13	DR. NORMALCURVESAURUS, PH.D. PRESENTS

Monthly Challenge

Name the media and author of the art depicted on Page 3. The answer can be found on the last page.



"That's true. We did advertise for someone who 'works well under pressure'..."

SCASA MEMBERSHIP RENEWAL

Please Pay Your Annual Dues

All the proceeds go to a student scholarship/award fund. We have collected \$750 thus far.

We encourage everyone who hasn't paid their SCASA membership dues in 2021 to go to

<https://my.cheddarup.com/c/scasa-2021-membership-dues-collection> and pay.

If you wish to pay with a check, cheddarup.com will give you this option.

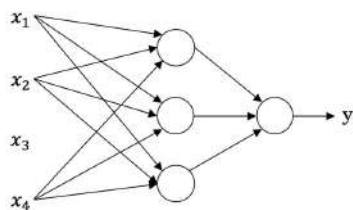
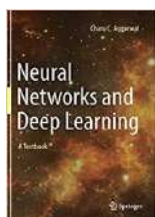
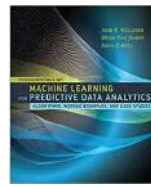
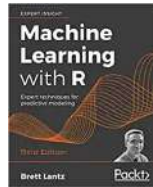
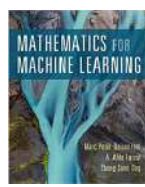
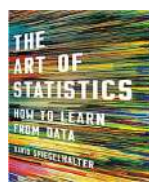
Membership is \$12/general, \$6/student and retiree.

Post-event Report: 40th Applied Statistics Workshop

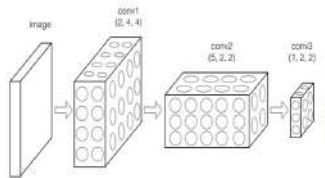
On April 24th we held a bright and cheerful celebration of the 40th anniversary of our Applied Statistics Workshop with a webinar titled "Introduction to Data Science, Machine Learning and Deep Learning (in R and Python)". Our instructors, Hui Lin (Netlify) and Ming Li (Amazon) delivered nice, perfectly structured, full of instructions and useful information lectures. The focus of the workshop was on using Databricks Community Edition (a free edition) to run big-data models in the cloud. Decision trees and various kinds of neural networks were explained and hands-on sessions were provided.

A book raffle was held at the culmination of the day as we are accustomed to by now. There were twenty nine attendees, seven of whom won excellent data science related books.

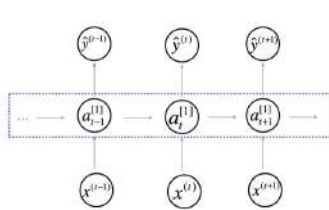
We would like to extend our gratitude to [Amgen](#), our faithful and reliable sponsor that has been sponsoring this event for many years now. We would also like to thank the [ASA](#) for allowing us to use the stimulus fund to pay for the books.



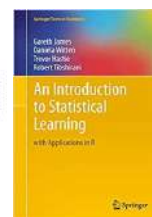
Standard NN



Convolutional NN



Recurrent NN



Post-event Report

Celebrating our 10th annual...

DATAFEST

ASA • UCLA • April 23rd - 25th, 2021



Written by Robert Gould and Linda Zanontian, UCLA

The year 2011 marked the 10th anniversary of DataFest, founded by Rob Gould in 2011. Once again, the pandemic forced us to celebrate via a virtual event, but the students rose to the challenge. We had 76 undergraduate teams compete, and teams came from UCLA, Pomona College, UC Riverside, and USC. The success of the event is due in large part to the amazing efforts of the DataFest student committee, led by Isabel Roig Penso '21 and Michelle Lee '20.

This year's virtual event matched one key component of the in-person events: students had just two days to complete the challenge, and worked around the clock from the moment of the big reveal on Friday, April 23 until the deadline on Sunday April 25. They submitted a short (5-minute) video of their findings, and a one-page summary. The short time allotted meant that students had to think really carefully about the message they wanted the judges to hear.

The data and the challenge were provided by the Rocky Mountain Poison Control Center. They asked students to provide advice for medical professionals to help them identify potential misuse of prescription drugs. The data included surveys of tens of thousands of respondents in the United States, Germany, Canada, and the UK. While the data set was perhaps smaller than in the past in terms of gigabytes, it was rich in columns: each data set from the different countries had well over 500 variables, including rich demographic variables as well as lifetime and recent use of prescription drugs.

This year the judges singled out two teams for a special "Judges Choice" awards. This award is given to exemplary work that doesn't fit neatly into the three award categories. One of these awards went to team **YYDS+** (Zhiqi Zhang, Deyi Zhu, Shiman Zhang, Yixin Chen) from UCLA, for their overall analysis of drug misuse behaviors among non-stereotypical populations, in particular healthcare professionals and college students. Another Judges Choice was awarded to **Team Sagehen** (Hannah Mandel, Emily Tomz, Adeena Liang, Chloe Sun, Ian Krupkin) from Pomona College. This team tied their numerical analysis to impressive visualizations to analyze non-medical opioid usage among Americans 18-24 years old.

The Best Use of External Data award goes to the team that uses data not provided by the organizers in order to enhance the context of the challenge and find additional nuance. The year, the prize goes to the team **No Free Lunch** (Cassandra Man, Wilson Ho, Tara Erberich, Amy Zhang, Yilin Qi) from USC, for their integration of policing data to investigate how non-medical drug use correlates with police density, and for examining the policing variables that predict drug use in the U.S. The Honorable Mention was awarded to team **BruWins** (Sivaji Turimella, Nishant Jain, Arnav Garg, Nitin Subramanian) from UCLA. BruWins examined state-level data on marijuana legalization laws and counts of drug-related arrests at the state level.

The Best Insight award went to team **K-Bandits** (Tiffany Feng, Eustina Kim, Diana Pham, Kienna Qin) from UCLA. They were awarded this prize for their multifaceted and coherent analysis of the characteristics among students, health care professionals and veterans in the U.S., that determine their use of benzodiazepines. An honorable mention in this category was awarded to team **The Stat Pack** (Robi Chatterjee, Himani Yalamaddi, Jenna Schindele, Matthew Kymn, Giorgia Nicolaou) from UCLA, for their combination of visual and statistical analyses of the relationships among age, drug use, and mental health disorders.

(Continues on the next page...)

Post-event Report

(... Continues from the previous page.)

Competition for the Best Visualization Prize is always intense, and this was proven this year by the judges' decision to award two honorable mentions. The first prize in this category went to team **Stackers** (Andrew Kan, Ben Brill, Genesis Qu, Maddy Blasingame) from UCLA. Stackers impressed the judges with an outstanding and clearly visualized analysis of cannabis legality and drug misuse. Team **We R Stats** (Ananya Sini Achan, Coco Wu, Huanfei Wang, Meirong Gu, Xiaofeng Lin) from UCLA won honorable mention for visualizing the distribution of mental health problems among three social groups, and visualizing their non-medical use of drugs. **Team R-mageddon** (Ethan Allavarpu and Dara Tan) from UCLA were awarded the honorable mention for their clear visualization of quantitative information in their analysis of drug misuse in the U.S.

The first-place winners will win a DataFest medallion with certificate and Honorable Mentions will receive a certificate. All winners will receive a one-year membership to the American Statistical Association along with the Southern California Chapter of the American Statistical Association.

DataFest is very grateful for the assistance of a large team of mentors, who held virtual consulting sessions, and judges. Even when we are virtual, we need sponsors, and so we are extremely grateful for the contributions from Lance Perry, Patrick James Crutcher who both contributed at the "Lognormal Sponsors" level. "Weibull Sponsors" included Harry Johnson, the Mary Rita Garber Living Trust, and the Southern California Chapter of the American Statistical Association. We are thankful for the support for Individual Sponsors: Ray-Bing Chen, Ph.D., Michael Tsiang, Ph.D., Kekona Kai Sorenson, Gary William Evans, Ph.D., Kayla Michelle Frisoli, Ph.D., James W. Wilson, Lauren Thompson, and Nancy Wallan. Intuit, Inc. and Microsoft Corporation provided matching donations for some contributions.

We are particularly grateful to Scott Schumacher for his gift to help launch a pooled gift, now in its second year, to endow DataFest.

It wouldn't be DataFest without data, and so we are very grateful for the hard work provided by the team at Rocky Mountain Poison Control. In particular, thank you Kari Rockhill for keeping this going over two long years and changes of personnel, and to Elise Amioka and Joshua Black for their unrelenting support and assistance.

A very special thank you to all our judges and mentors who worked tirelessly along with the participants all weekend and brought the DataFest energy to our virtual setting! The weekend would not have been the same without their great efforts!

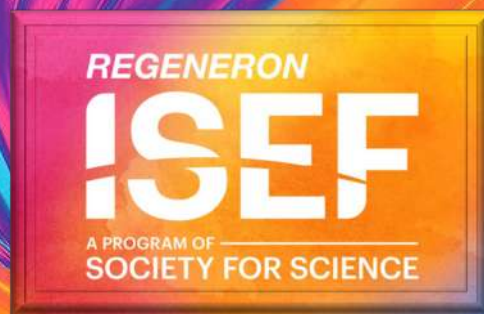
Finally, we thank Campus Program Committee and the students in the UCLA Stats



Club who volunteered for the DataFest Committee. They worked long hours and organized a fantastic event, a stellar line-up of workshops, did fantastic advertising, created the swag boxes, and kept the level of enthusiasm high. So special thanks (again) to chairs Michelle Lee and Isabel Roig Penso and their amazing team: Fontanna Yee, Stephanie Chu, Valerie Taruno, Ashley Lee, Britney Zhao, Kiara Edgel, Meichen Chen, Rebecca Xu, Rhea Singh, and Shirley Tang.

You can watch **all** of the presentations at <http://datafest.stat.ucla.edu/competition/2021-presentations/> Please take the time to see the amazing creativity and talent of our students. We congratulate all students who participated in this event during these pandemic times, and look forward to seeing everyone on campus at DataFest 2022.

Post – event Report



Written by Madeline Bauer, USC (retired), Johanna Hardin, Cal Poly Pomona, Rebecca Le, Riverside County, and Erin Chapman, Tektronix

The Regeneron International Science and Engineering Fair (ISEF) 2021 was conducted virtually from April 29, 2021 to May 21, 2021. There were 1,842 finalists with 1,430 projects across 21 scientific disciplines. The finalists represented 64 countries. Forty-seven statisticians participated in judging for the ASA Special Awards in Statistics.

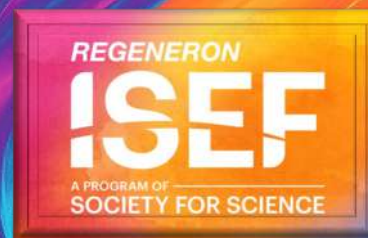
The First Award of \$1500 went to **Aditka Kendre, Cumberland Valley High School, Mechanicsburg, PA** for the Biomedical Engineering project ENBM035: "Employing Adversarial Machine Learning and Computer Audition for Smartphone-based Real-time Arrhythmia Classification in Heart Sounds." This project used ML models to detect abnormal heart sounds to classify arrhythmia. The project modeling, testing, evaluation and interpretation incorporated many of the key statistical approaches in a very innovative engineering project. They demonstrated the feasibility of implementing the ML model in a smartphone app to classify heart sounds in real time. They plan to conduct prospective clinical trials to test the smartphone app in the real world.

The Second Award of \$1,000 went to **Hannah Guan, BASIS San Antonio Shavano Campus, San Antonio, TX**, for CBIO008: a Computation Biology and Bioinformatics project – "The Genetics of Human Aging: Predicting Age and Age-Related Diseases by Deep Mining High Dimensional Biomarker Data." In this project, the Correlation Pre-Filtered Neural Network model has been developed and tested yielding the best prediction accuracy among all tested methods including Neural Networks with LASSO regularization, elastic net regularization, and the Dropout Neural Networks. Not only did Ms. Guan have a strong understanding of the technical details of the machine learning methods, but she also embraced the ideas of variability seen in data. She used testing and training data to assess the validity of her models, and she followed up the modeling part by applying statistical inference to measure the relationship between the genetic models and disease status.

The Third Award of \$500 went to **Sarah Cao, Palos Verdes Peninsula High School, Rolling Hills Estates, CA**, for a Plant Science project – PLNT016: "Analysis and Prediction of the Spread of Invasive Plant Species in Relation to the Changing Environment." In this project, the Random Forest machine learning model has been used to determine the relation between the spread of invasive plant species and various environment factors and predict future plant spread with promising accuracy. Using the WorldClim dataset, she used the model to predict the plant survival and spread rate under each environmental factor variable. The different aspects of climate change and the significance of their effects on the success of the plant type were measured. Instead of invasive species being less dependent on changes in the environment, she found that native and invasive species had very similar profiles in their dependencies on climate factors. Thus, she concluded that her hypothesis was wrong, and climate change is expected to have a very similar effect on both native and invasive species, as both plant types would need to make similar adaptations to survive. One such example is that both groups were most dependent on seasonal precipitation for their continued survival. This project also was one of six projects awarded 4th place in the Plant Sciences category.

Continues on the next page...

Post – event Report



Continued from the previous page.

The ASA Special Awards in Statistics awarded ten projects "Honorable Mention," going to finalists in the US from Albuquerque, New Mexico to Bangor, Maine, and overseas, including Lao Cai, Vietnam, Bhopal, India and Kemerovo, Russia.

The Special Awards Ceremony for ISEF 2021 can be viewed here <https://www.youtube.com/watch?v=2KDvEdehEMs>. The announcement of the ASA awards begins at 23:00. A summary of the ISEF 2021 Grand Awards is available here:

<https://www.societyforscience.org/press-release/2021-regeneron-isef-grand-awards/>

and the Special Awards summary is given here:

<https://www.societyforscience.org/press-release/2021-regeneron-isef-sao-awards/>

The project displays can be viewed at the virtual [ISEF ProjectBoard <https://projectboard.world/isef/finalist-exhibit-hall?r=4xwks>](https://projectboard.world/isef/finalist-exhibit-hall?r=4xwks) platform. But perhaps the best view of ISEF and the finalists comes from the opening ceremony:

<https://www.youtube.com/watch?v=mpBe4DvWuJo;>

Last but not least, many many thanks to the statistics judges! Forty-seven statistics judges participated in the judging for the ASA Special Awards in Statistics. These judges have always been known for their outstanding dedication and enthusiasm. This year they also demonstrated their incredible creativity and amazing flexibility under pressure! Thank you!

See the SCASA June eNewsletter for more about the joys and rewards of judging for the ASA Special Awards in Statistics!





Recently Dr. Chong Ho Alex Yu, our Vice President for Statistical Education, won the Distinguished SAS Educator Award, which recognizes an educator who has made significant contributions to advancing analytics education and preparing students for careers in analytics and data science. This prestigious award is given to an educator whose efforts and impact in the field and engagement with SAS have spanned more than five years. The applications for the competition were submitted before March 2021, and then the SAS committee selected one winner and one finalist from each group of continents. **[Dr. Yu is the winner of the American continent \(North and South America\).](#)**

https://www.sas.com/en_us/events/sas-global-forum/program/awards-academic-programs.html#f0703c75-2034-4532-9de4-6b08388df14f



Dr. Yu holds a Ph.D. in Educational Psychology (Arizona State University, ASU) with a focus on Measurement, Statistics and Methodological Studies, as well as a Ph.D. in Philosophy (ASU) with specialization in History and Philosophy of Science. Currently he is a Professor of Behavioral and Applied Science and a member of the adjunct mathematics faculty at Azusa Pacific University (APU), USA. He is also a quantitative research consultant and the committee chair of the Big Data Discovery Summit/Data Science Consortium at APU. His research interests include exploratory data analysis, data visualization, and data mining. He has published numerous book chapters and journal articles pertaining to international assessments and big data analytics. Currently he is working on a book entitled "From Traditional Statistics to Modern Data Science," which will be published by CRC Press.

<https://creative-wisdom.com/pub/pub.html>

Since 2000 he has presented numerous papers in international, national, and local SAS conferences, such as SAS Global Forum, SAS Analytics Experience, and Western Users of SAS Software Conference. His paper "An introduction to computing and interpreting Cronbach Coefficient Alpha in SAS" presented in the 26th SAS User Group International Conference is one of his most cited articles, as indicated by Google Scholar and Semantic Scholar.

<https://scholar.google.com/citations?user=mdGny3EAAAAJ&hl=en>

<https://www.semanticscholar.org/author/Chong-Ho-Yu/4557614>

He won the SAS Faculty Scholarships three times. In addition, the paper coauthored with his students "Nonlinear modeling with big data in SAS and JMP" won the best paper award in the category of data science and big data analytics at the 2017 Western Users of SAS Software Conference. Dr. Yu has been teaching SAS in the classroom since 1999. Currently he is teaching data visualization and data mining courses at APU. Because the research finding based on a single analysis with a small sample is not highly generalizable, students in his classes are required to perform triangulation (use of multiple methods), model comparison, and big data analytics.

<https://creative-wisdom.com/teaching/teaching.html>



By Sunil K. Gupta, Founder of SASSavvy.com and a long-time supporter of SCASA

SAS Savvy is happy to offer free R Programming webinars! SASSavvy.com is now an open source for all R and SAS programmers! To attend these R webinars with live Q&A and access a free R Programming e-book, please sign up. Enter your R topics for future presentations.

Please sign up to receive registration information:

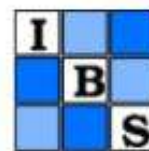
<https://lp.constantcontactpages.com/su/kwo8FtG/RWebinars>

List of Webinars

- Jun. 24, 2021** What is R and Why Should You Learn R?
- Aug. 17, 2021** Writing Your First R Program, Just the Basics
- Oct. 26, 2021** Anatomy of R Reports, Look what R can do!
- Dec. 9, 2021** R is for Pharma, R You?
- Jan. 27, 2022** R and SQL: Best of Both Worlds



Upcoming Online Short Courses

**WNAIBS**

The Western North American
Region of The International Biometric Society

Hello, **WNAR Members**:

Discount registration fees are available through **June 18** for online short courses in the 2021 UW Biostatistics Summer Institutes. We invite you to check out the 50+ online short courses offered in July 2021:

- **Statistical Genetics (SISG), July 7-23** <<https://si.biostat.washington.edu/suminst/sisg2021/modules>>
- **Statistics for Big Data (SISBID), July 12-23** <<https://si.biostat.washington.edu/suminst/sisbid2021/modules>>
- **Statistics for Clinical & Epidemiological Research (SISCER), July 7-30** <<https://si.biostat.washington.edu/suminst/siscer2021/modules>>
- **Statistics and Modeling in Infectious Diseases (SISMID), July 7-23** <<https://si.biostat.washington.edu/suminst/sismid2021/modules>>

A limited number of registration scholarships are still available to graduate students attending a U.S. college or university for short courses in **Statistical Genetics (SISG)**

<<https://si.biostat.washington.edu/suminst/sisg/scholarships>>

Learn more at si.biostat.washington.edu and email uwbiost@uw.edu if you have questions.

Best,

Bruce Weir, Director, Summer Institute in Statistical Genetics

Ali Shojaie, Director, Summer Institute in Statistics for Big Data

Kathleen Kerr, Director, Summer Institute in Statistics for Clinical and Epidemiological Research

M. Elizabeth Halloran, Director, Summer Institute in Statistics and Modeling in Infectious Diseases

5-9 JULY, 2021

useR! 2021 The R Conference



The useR! conferences are non-profit conferences organized by community volunteers for the community, supported by the R Foundation. Attendees include R developers and users who are data scientists, business intelligence specialists, analysts, statisticians from academia and industry, and students. The useR! 2021 conference will be the first R conference that is global by design, both in audience and leadership. It will take place **5-9 July 2021**. Monday, Tuesday, Thursday and Friday are the main conference days. On Wednesday, you can “take a break” from the conference at one of our tutorials, participate in one of the satellite events or simply enjoy some time away from your screen.

For the [program overview](https://user2021.r-project.org/program/overview/), please go to <https://user2021.r-project.org/program/overview/>

Early bird fees apply until May 28th! Industry \$75 • Academia \$50 • Student \$25

Tutorials To view the schedule of the tutorials, please go to <https://user2021.r-project.org/program/tutorials/> The basic price is for two tutorials. If you book only one tutorial, you will get a 50% discount. Industry \$50 • Academia \$30 • Student \$15

To register for the conference and tutorials, go to

<https://user2021.r-project.org/participation/registration/>





JSM Is ... One of the Largest Statistical Events in the World

- More than 6,500 attendees from 52 countries
- 600+ sessions, including invited, topic-contributed, contributed, and poster
- More than 1,000 student attendees
- 75+ employers hiring for more than 200 positions
- 80+ exhibitors
- More than 40 Professional Development short courses and workshops

It is also one of the broadest, with topics ranging from statistical applications to methodology and theory to the expanding boundaries of statistics, such as analytics and data science.

JSM also offers a unique opportunity for statisticians in academia, industry, and government to exchange ideas and explore opportunities for collaboration. Beginning statisticians (including current students) can learn from and interact with senior members of the profession.

Make plans now to attend the **2021 Joint Statistical Meetings**. Take advantage of **early registration** and join thousands of your colleagues from around the world **August 8-12**.

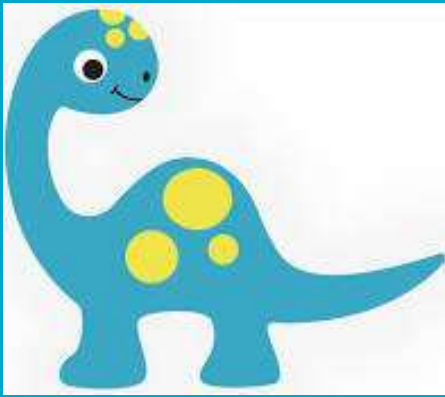
<https://ww2.amstat.org/meetings/jsm/2021/>

Early Registration for JSM 2021 Is in Full Swing Through June 15

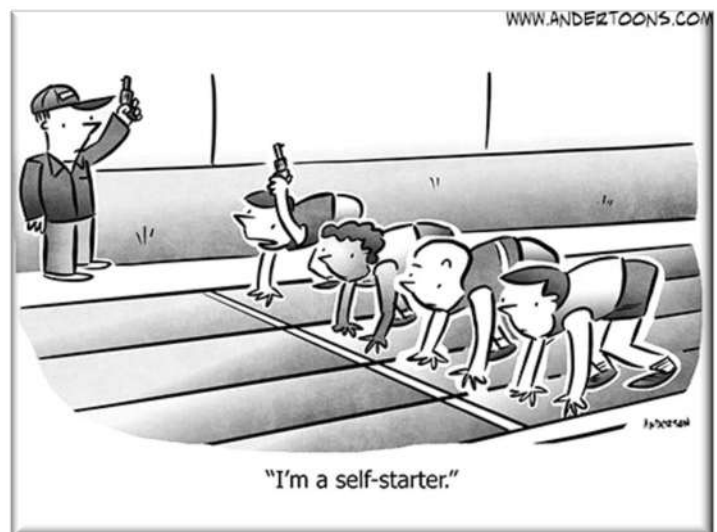
<https://ww2.amstat.org/meetings/jsm/2021/registration.cfm>



REGISTER TODAY!



Dr. Normal curvesaurus, Ph.D. presents



Answer to Monthly Challenge: Dale Chihuly's Glass Art