

Bryant University

## Bryant Digital Repository

---

Analytics Without Borders

Conferences and Events

---

3-2020

### **Analytics Without Borders Program Schedule**

Tufts University

Follow this and additional works at: <https://digitalcommons.bryant.edu/analytics>

---

[Home](#) » [Tufts Analytics Without Borders Conference \(TAWB 2020\)](#) » Schedule and Program

# Schedule and Program

FEB 28			JAHARIS LOBBY	BERKHARIS AUDITORIUM	J118	J156	J105	CAFE	
OPENING/CLOSING	8 AM	8:00	REGISTRATION					BREAKFAST	
KEYNOTE REMARKS		8:15							
PANEL DISCUSSION		8:30							
TECHNICAL WORKSHOP		8:45							
DISCUSSION WORKSHOP									
RESEARCH TALKS	9 AM	9:00		ABANI PATRA TUFTS DATA INTENSIVE STUDIES CENTER					
		9:15		BRIAN SMITH NOVARTIS PHARMACEUTICAL CORPORATION					
		9:30							
		9:45							
		10:00			COFFEE BREAK				
10 AM	10:15	CAREER PANEL FOR INDUSTRY-MINDED YOUNG PROFESSIONALS		INTRO TO GIS USING QGIS: EXPLORING ACCESS TO HEALTHY FOODS IN CAMBRIDGE MA	RESEARCH TALKS (GROUP A)				
	10:30								
	10:45								
	11:00								
	11 AM	11:15		RESEARCH TALKS (GROUP B)		TEACHING DATA ANALYTICS TO NON- ANALYTICS STUDENTS			
11:30									
11:45									
12:00									
12 PM		12:15			POSTER PRESENTATIONS & LIGHTENING TALKS			LUNCH	
	12:30								
	12:45								
	1:00								
	1 PM	1:15		DATA ANALYTICS FOR MITIGATING HEALTH DISPARITIES		IDENTIFYING YOUR AUDIENCE TO CAPTIVATE	RESEARCH TALKS (GROUP C)		
1:30									
1:45									
2:00									
2 PM		2:15		TRADE TRADEOFFS: EFFICIENCY, RESILIENCE, AND SUSTAINABILITY	COMPARISONS ACROSS STATISTICAL SOFTWARE	WHAT'S IN THE RECIPE: MODEL DEVELOPMENT AND DIAGNOSTICS	RESEARCH TALKS (GROUP D)		
	2:30								
	2:45								
	3:00								
	3 PM	3:15			COFFEE BREAK				
3:30		UNDERSTANDING BIG DATA IMPLICATIONS FOR FOOD POLICY		SPEAKING RESULTS WITHOUT WORDS DATA VISUALIZATION	BREAKING DOWN SILOS: CONDUCTING TEAM-BASED DATA ANALYSES	RESEARCH TALKS (GROUP E)			
3:45									
4:00									
4:15									
4 PM	4:30	ALFREDO MORALES REDZONE PRODUCTION SYSTEMS							
	4:45								
	5:00								
	5:15								
	5 PM	5:30		MINGFEI LI BENTLEY UNIVERSITY					
5:45									
6 PM		6:30		DINNER RECEPTION (CAFÉ)					



OPENING/CLOSING  
KEYNOTE REMARKS  
PANEL DISCUSSION  
TECHNICAL WORKSHOP  
DISCUSSION WORKSHOP  
RESEARCH TALKS

FEB 29		JAHARIS LOBBY	BERKHARIS AUDITORIUM	J118	CAFE	
8 AM	8:00	CONFERENCE CHECK-IN			BREAKFAST	
	8:15					
	8:30					
	8:45					
9 AM	9:00		RYAN SIMPSON ELIZABETH MARSH CONFERENCE CHAIRS			
	9:15		FRANK HU HARVARD TH CHAN SCHOOL OF PUBLIC HEALTH			
	9:30					
	9:45					
10 AM	10:00		COFFEE BREAK			
	10:15		PERSPECTIVES ON ENVIRONMENTAL AND ECONOMIC COSTS TO FOOD SECURITY	NETWORK SCIENCES AND COMPLEX SYSTEMS IN NUTRITION & DISEASE RESEARCH		
	10:30					
	10:45					
11 AM	11:00					
	11:15	DATA ANALYTICS IN NUTRITION AND HEALTH BUSINESS: A DATA ANALYTICS CAREER PANEL				
	11:30					
	11:45					
12 PM	12:00					
	12:15	ELENA NAUMOVA TUFTS UNIVERSITY				
	12:30	PROFESSIONAL NETWORKING LUNCHEON (CAFÉ)				
	12:45					
1 PM	1:00					
	1:15					
	1:30					
	1:45					
2 PM	2:00					
	2:15					

## Keynote Addresses

^ Brian Smith, Executive Director NIBR Biostatistics, Novartis Pharmaceutical Corporation

^ Alfredo Morales, Principal Research Scientist, Redzone Production Systems

✓ Frank Hu, Chair, Department of Nutrition, Harvard T.H. Chan School of Public Health

Dr. Frank Hu's research has focused on diet/lifestyle, metabolic, and genetic determinants of obesity, type 2 diabetes, and cardiovascular disease (CVD). His major research interests include major topics in epidemiology and prevention of cardio-metabolic diseases through diet and lifestyle. His group has conducted detailed analyses of many dietary and lifestyle factors and risk of diabetes and CVD, and the findings have contributed to current public health

recommendations and policies for prevention of chronic diseases. Dr. Hu's group has also identified novel biomarkers and gene-environment interactions in relation to risk of obesity and diabetes by integrating cutting-edge omics technologies into epidemiological studies and pioneering the Systems Epidemiology approach.

## Panel Descriptions & Panelists

- ^ Career Panel for Industry Minded Professionals
- ^ Data Analytics for Mitigating Health Disparities
- ^ Trade Tradeoffs: Efficiency, Resilience, and Sustainability
- ^ Understanding Big Data Implications for Food Policy
- ^ Perspectives on Environmental and Economic Costs to Food Security
- ^ Data Analytics in Nutrition and Health Business: A Data Analytics Career Panel

## Workshop Descriptions

- ^ An Intro to GIS using QGIS: Exploring Access to Healthy Foods in Cambridge, Ma
- ^ Teaching Data Analytics to Non-Analytics Students
- ^ Comparisons Across Statistical Software
- ^ Identifying Your Audience to Capture: Effective Communication of Data-Driven Research
- ^ What's In The Recipe: Model Development and Diagnostics
- ^ Speaking Results Without Words: Data Visualization
- ^ Breaking Down Silos: Managing and Analyzing Data As A Team
- ^ Network Sciences and Complex Systems in Nutrition-Related Research

## Oral Presentations

- ^ Group A: Nasir et al. Substance Use Disorder Outcome Predictions for Decision Support: A Holistic Data Approach
- ^ Group A: Wang et al. Sleeping Disorders and Alzheimer's Disease: Evidence for Heterogeneity
- ^ Group B: Stark. What a "Batman Graph" shows about US income: A cautionary tale of correlation, principal components, and nonlinear regression
- ^ Group B: Nasir et al. Synthetic Average Neighborhood Sampling Algorithm (SANSAs): A Neighborhood Informed Synthetic Sample Placement Approach to Improve Learning from Imbalanced Data
- ^ Group C: Ericson. Minds in Motion: Using positional data to study human behavior and cognition
- ^ Group C: Abhilash et al. Building a Predictive Framework for Analyzing Opioid Patient Toxicology Data to Improve Medication-Assisted Therapy Adherence and Relapse

- ^ Group D: Bai & Wallbaum. Dual-dynamic Retirement Income Strategy for Better Retirement Outcomes
- ^ Group D: Mentzer et al. The Role of Gender and Party on the Twitter Conversation in the 2018 U.S. Senate Elections
- ^ Group E: McGuirk. Business Ethics at the Core of Successful and Sustainable Analytics Practices

## Poster Presentations

- ^ Hsu et al. Differences in nutrient intake between individuals with and without familial longevity
- ^ Gregory & Byrd. Using Geographic Information Systems (GIS) to Determine Sites for SNAP-Ed
- ^ Chow et al. Use of advanced statistical analysis in determining death manner: lessons from 330,000 national violent death data
- ^ Hur et al. Combining unstructured social media data with structured scientific literature: a data merging problem
- ^ Deichmann et al. European Moods: Satisfaction levels from 2003 to 2016



## **Keynote Addresses**

### **Brian Smith, Executive Director NIBR Biostatistics, Novartis Pharmaceutical Corporation**

Dr. Smith is an executive director and global group head in biostatistical sciences supporting the Novartis Institute of BioMedical Research. Before joining Novartis in 2014, Dr. Smith worked at Amgen Inc. (2005-2014), Eli Lilly and Company (1996-2005), and University of Louisville Kidney Disease Program (1994-1996). Prior to his professional career, he received a Ph.D. in statistics from University of Kentucky. All positions in his career in the pharmaceutical industry have supported early clinical development. Dr. Smith plays an active role in promoting quantitative sciences in drug development. He is currently a member of the American Society of Clinical Pharmacology and Therapeutics, the International Society of Pharmacometrics, and the American Statistical Association. Dr. Smith is the current Chief Statistical Advisor of Clinical Pharmacology and Therapeutics, an Associate Editor of Statistics in Biopharmaceutical Research, and a faculty member of the American Course on Drug and Regulatory Science.

### **Alfredo Morales, Principal Research Scientist, Redzone Production Systems**

Dr. Alfredo Morales works in understanding the complex behavior of social systems by analyzing big data with artificial intelligence algorithms, networks and complexity science. He explains large scale, societal behaviors, such as social segregation and polarization, by retrieving unstructured patterns of information from large datasets resulting from human activity on Internet, mobile phones and shopping data. He works closely with world class researchers from academia and industry, including institutions like MIT, Harvard and the UN. He is a member of the New England Complex Science Institute (NECSI). In 2017 Dr. Morales was included in the list of Latinos of the Future by the journal El Planeta and in 2018 – in the list of 35 Innovators Under 35 by MIT Technology Review.

### **Frank Hu, Chair, Department of Nutrition, Harvard T.H. Chan School of Public Health**

Dr. Frank Hu's research has focused on diet/lifestyle, metabolic, and genetic determinants of obesity, type 2 diabetes, and cardiovascular disease (CVD). His major research interests include major topics in epidemiology and prevention of cardio-metabolic diseases through diet and lifestyle. His group has conducted detailed analyses of many dietary and lifestyle factors and risk of diabetes and CVD, and the findings have contributed to current public health recommendations and policies for prevention of chronic diseases. Dr. Hu's group has also identified novel biomarkers and gene-environment interactions in relation to risk of obesity and diabetes by integrating cutting-edge omics technologies into epidemiological studies and pioneering the Systems Epidemiology approach.

## **Panel Descriptions & Panelists**

### **Career Panel for Industry Minded Professionals**

#### **Description**

This panel will invite local business professionals in fields of industry and business to discuss skills important for students as they transition to become young professionals. The panel will discuss those qualities each company seeks in applicants as well as why certain data analytics skills may be of greater importance in industry than others.

#### **Panelists**

Kevin Mentzer, Assistant Professor, Bryant University

Fotios Kokkotos is a Partner and Head of Data Science Statistics at Trinity Life Sciences. Dr. Kokkotos is an accredited professional statistician and data scientist with over 25 years of experience in statistical consulting. Joining Trinity Life Sciences from PricewaterhouseCoopers in 2002, Dr. Kokkotos was able to create the statistics and subsequently the advanced analytics and data science groups, introduce Trinity to health economics and outcomes research projects and identify appropriate databases to support our client project needs. Beside his current research interests in data science and ongoing academic collaborations with many universities, Dr. Kokkotos is a board member of the American Statistical Association's accreditation committee, which reviews and approves professional statisticians around the world. Dr. Kokkotos earned a doctorate degree in mathematical statistics from American University.

After graduating from the NYU Stern School of Business in 2017 with a degree in finance and marketing, Helen Hsia began working as a data analyst for the IBM corporate marketing team. During her time with the team, she helped develop an audience quality metric and managed an in-house effort to build a B2B multi-touch attribution system. Since moving back home to Boston late last year, she has been furthering her data skills as a marketing analyst for the IBM Security team, working closely with the Threat Management portfolio, digital, and brand teams.

Bola Ajayi, Advanced Analytics

### **Data Analytics for Mitigating Health Disparities**

#### **Description**

This panel will discuss health disparities affecting racial minority groups, and how the emergence and application of big data can help mitigate such disparities through analysis of differences in disease rates, exposures and risk, socioeconomic status, health coverage and insurance. This talk will discuss why disparities might arise, giving students a better



understanding of the problems faced by different racial demographics, and big data's potential for attenuating those disparities.

### **Moderators**

Kimberly Dong, Tufts School of Medicine

Seblewongel Yigletu, Tufts University

### **Panelists**

Adam Pittman is a data Scientist at Folia Health, a startup company dedicated to helping patient and caregiver observations make their way into clinical practice. Currently, he works on integrating real world evidence and patient reported outcomes into existing healthcare data structures. Previously he worked with a Colorado syringe access program to help drive new legislation for individual and population safety in the opioid usage space. Adam holds a ScM in Biostatistics from Johns Hopkins Bloomberg School of Public Health.

Dr. Anna Orlova, a Senior Faculty, Health Informatics and Analytics Program, Tufts School of Medicine. She is also a Visiting Associate Professor in informatics at Johns Hopkins School of Medicine. Dr. Orlova's informatics interests are in the areas of the Electronic Health Records, telehealth and digital health; health data and systems standardization and interoperability; and data trust. Dr. Orlova joined the Tufts University in 2018 to launch the Health Informatics and Analytics program at Tufts. Her online teaching experience includes teaching various informatics and health IT standardization courses at Johns Hopkins, University of Massachusetts, Amherst, and Tufts.

Dr. Carlota Dao is a Scientist III in the Energy Metabolism Laboratory at the Human Nutrition Research Center on Aging. Her research interests include obesity, weight management, food culture and eating behavior, gut microbiota, and health disparities. Specifically, her projects span three interrelated areas within obesity research: 1) Studying how cultural, environmental and biological factors determine eating behavior and weight status. 2) Developing culturally relevant lifestyle interventions for weight loss, including an ongoing community-based pilot intervention in older Hispanic adults. And 3) Understanding the role that the gut microbiota plays in chronic disease risk. Her previous experience includes establishing novel data integration approaches to analyze 'big data' and gain new insights on the interaction between host biology, gut microbiota, and environmental factors.

Glory Song has been working as an Epidemiologist for the Department of Public Health since 2011. She supports surveillance and evaluation work for a number of chronic disease prevention programs within the Bureau of Community Health and Prevention, including the Massachusetts Tobacco Cessation & Prevention Program (MTCP) and Mass in Motion. She has background and interest in quasi-experimental study designs, survey instrument

development, and community-level policy evaluation. Glory received her MPH from Boston University School of Public Health.

## **Trade Tradeoffs: Efficiency, Resilience, and Sustainability**

### **Description**

Economic and trade theory guide our understanding of agricultural policy, international food prices, and foreign food aid programming and planning. Experts in this panel will discuss innovation and barriers to creating sustainable supply chains: from community-level logistics to international trade policies such as trade war tariff increases and the NAFTA renegotiation. They will also share how they are innovating sustainably, both in terms of environmental impacts, and in terms of social inclusion and equity.

### **Moderator**

Will Masters is a Professor at Tufts University, in the Friedman School of Nutrition and the Department of Economics, working on the economics of agriculture, food and nutrition. From 2006 through 2011 he edited the journal Agricultural Economics. He is an elected Fellow of the Agricultural & Applied Economics Association (AAEA). Details online at <http://sites.tufts.edu/willmasters>.

### **Panelists**

Christopher Mejía Argueta is a Research Scientist at the MIT Center for Transportation and Logistics (CTL). He develops applied research on retailing operations and food supply chains for multiple stakeholders in the Food and Retail Operations Lab (FaROL). His research focuses on improving the efficiency, flexibility of operations in multiple stakeholders, creating food access models to address the fragmented retail market and farmer's side. His research focuses on reducing undesired socioeconomic and health problems related to income disparity, food malnutrition, food waste by proposing sustainable policies, business models to help vulnerable population segments. Dr. Mejía is also the Director of the MIT SCALE network for Latin America.

Andrew Feierman is a Data Scientist working on the Trase project. Before joining SEI, he worked on global environmental policy for Dr. Angel Hsu in the Data-Driven Environmental Solutions Lab, jointly based out of Yale University and Yale-NUS in Singapore. He also has experience working with large private companies on reducing energy consumption in buildings through the Institute for Market Transformation in Washington, DC, and holds a degree from American University's School of International Service.

Ravdeep Jaidka is the Sourcing Manager at the fresh produce division of Equal Exchange, managing the banana and avocado programs sourced directly from small farmer cooperatives in Ecuador, Peru and Mexico. Ravdeep started at Equal Exchange in 2015,

after receiving her Master's degree in the Agriculture, Food and Environment Program at the Friedman School.

## **Understanding Big Data Implications for Food Policy**

### **Description**

There is no commonly accepted definition of the term big data, yet this ambiguity has not stopped policy makers from taking an interest in its collection, management and use. But how is big data relevant to nutrition and food policy? Increasingly, participants in the agricultural supply chain are collecting data from the farm gate to the plate, while questions of ownership of and access remain unanswered. Evidence-based research produces vast amounts of data about critical policy issues, which informs and justifies the actions of public, private and not-for-profit organizations, yet questions of appropriate data analysis remain. This panel will pursue the question, how does this growing body of highly detailed information – often referred to as “big data” - influence research, funding and implementation of food and nutrition policy, and what are the strengths and challenges of this approach?

### **Moderator**

*Katrina Sarson* is an Emmy Award winning television producer who is currently a Masters student at the Friedman School of Nutrition Science and Policy at Tufts University. Previously, she earned a Masters in Education at the Harvard Graduate School of Education with an emphasis on Technology in Education. Her interests are nutrition education, communication, and the ways that corporations and government agencies interact to shape food and nutrition policies. During her time at Friedman, she has worked on the Public Impact Initiative, and co-founded Friedman's Potluck Club, and Tufts Food Week. She is passionate about food, nutrition, bread baking, and engaging conversations.

### **Panelists**

*Laura Benavidez*, MBA has been the executive director of food and nutrition services of Boston Public Schools since August 2016. Laura was formerly with the Los Angeles Unified School District (LAUSD), where she was the interim co-director. She oversaw the operations and logistics for LAUSD, the second largest school district in the country with more than 560,000 students, 700 schools, 1,100 meal programs, and over 4,000 foodservices employees. Laura earned her bachelor of science in food science and technology, and master of business administration. She is currently pursuing a doctorate degree. Since starting at BPS' Food and Nutrition Services department, her focus has been to be fiscally sustainable, decrease waste, increase technology, and build the culture of the program. (Dorchester, MA)

Alana Davidson is the SNAP interagency specialist at the Massachusetts Department of Transitional Assistance. She manages the Department's SNAP projects with external agencies to address food insecurity in a more holistic way and improve SNAP access. This includes policy alignment, following legislation and regulations, research and assisting with writing the Department's publicly submitted comments for federal rule making. Prior, Davidson worked at anti-hunger nonprofit organizations on child nutrition advocacy and outreach. She holds a Master's of Science in Food Policy and Applied Nutrition from Tufts University and a Bachelors of Science in nutrition, dietetics from the University of New Hampshire.

Eileen Kennedy is a former dean of the Friedman School. Currently a professor at the School, Kennedy's research interests include assessing the health, nutrition, diet and food security impacts of policies and programs; nutrient density and diet diversity; and agriculture nutrition linkages. She is a member of the High Level Panel of Experts on Food Security and Nutrition of the UN Committee on World Food Security. Formerly a member of the UN SCN Advisory Group on Nutrition. She founded and was the first Executive Director of the USDA Center for Nutrition Policy and Promotion. She created the Healthy Eating Index which is used as a single summary measure of diet quality. She is currently a member of the World Economic Forum's Global Council on Food Security and Nutrition.

Parke Wilde (PhD, Cornell) is a food economist and professor at the Friedman School of Nutrition Science and Policy at Tufts University. Previously, he worked for USDA's Economic Research Service. At Tufts, Parke teaches graduate-level courses in statistics and U.S. food policy. His research addresses the economics of federal nutrition assistance programs. He was Director of Design for the SNAP Healthy Incentives Pilot (HIP) evaluation in Hampden County, Massachusetts. He has been a member of the Institute of Medicine's Food Forum and is on the scientific and technical advisory committee for Menus of Change, an initiative to advance the health and sustainability of the restaurant industry. He directs the USDA-funded Tufts/UConn Research Innovation and Development Grants in Economics (RIDGE) program. In March, 2018, Routledge/Earthscan released the second edition of his book, Food Policy in the United States: An Introduction.

## **Perspectives on Environmental and Economic Costs to Food Security**

### **Description**

How innovative analytic methods for understanding food security and food access in the context of environmental and economic stressors are yielding nuanced understanding of barriers to food security. This session evaluates the role data sciences can play in integrating interdisciplinary knowledge and data to understand the marriage of environmental and economic factors on topics of food security.

## **Moderator**

Dr. Meg Hartwick completed her Ph.D. work in Molecular and Evolutionary Systems Biology developing predictive models for emerging food and waterborne pathogens and received her MSc through the Tufts Cummings School of Veterinary Medicine in Conservation Medicine. She has worked as Data Scientist examining the intersection of food production and human and wildlife disease with the University of New Hampshire, Woods Hole Oceanographic Institute and Tufts University. She is currently a Data Scientist with InForMID.

## **Panelists**

Yan Bai is a doctoral student at Tufts' Friedman School of Nutrition. His work is focused on index studies on the cost of nutritious diets around the world. Before that, Yan earned the Master of International Business (MIB) at the Fletcher School. Besides advancing his knowledge of international finance and global health there, he also developed quantitative skills at Harvard Chan School. Prior to his graduate education, Yan worked as an investment analyst in the finance industry, where he obtained industry experiences in health and agriculture sectors in China, Southeast Asia, and Africa. Yan also holds a master's degree in Economics from Tufts and bachelor's degrees in Chemistry and Economics from Peking University in China.

Nicole Tichenor Blackstone is an Assistant Professor in the Division of Agriculture, Food, and Environment at the Friedman School of Nutrition Science and Policy at Tufts University. Dr. Blackstone's research focuses on developing and evaluating strategies to improve food system sustainability. Her work fuses industrial ecology, nutrition, and social science methods. To date, her research has explored the environmental and social implications of livestock agriculture, human diets, food waste management, and regional food systems. She teaches graduate courses on U.S. agriculture, environmental life cycle assessment, and corporate social responsibility in the food industry. Dr. Blackstone holds a Ph.D. and M.S. in Nutrition from Tufts University and a B.A. in Philosophy and Religious Studies from the University of Kansas.

Sean B. Cash is the Bergstrom Foundation Professor in Global Nutrition and an Associate Professor at the Friedman School of Nutrition Science and Policy at Tufts University. As an agricultural and food economist, his research focuses on how food, nutrition, and environmental interventions and policies affect both producers and consumers. He has conducted research in the areas of environmental impacts in food production, including projects on climate change and coffee and tea production, and invasive species management. Other work includes assessing the efficacy of food label and price interventions as public health and environmental tools; children's food choices in commercial and school environments; and consumer interest in food labeling of ethical attributes of food production.

Dr. Bea Rogers is Professor of Economics and Food Policy and Director of the Food Policy and Applied Nutrition Program at the Friedman School of Nutrition Science and Policy, where she has been on the faculty since 1982. Prof. Rogers has over 30 years of experience promoting evidence-based policy and programs related to food security, food consumption, and nutrition in the developing world. She has been responsible for the design and implementation of national household income, expenditure, and consumption surveys in several countries, and has conducted many smaller scale surveys of household economic and consumption behaviors. Her current work looks at effectiveness of alternative food aid products used in nutrition programs and on the sustainability of food aid program impacts in the face of insecurity and civil unrest. She is also working on a project to improve dietary data collection methods and promote the use of such data in policy-making.

Jennifer Coates, Tufts Friedman School of Nutrition Science and Policy

## **Data Analytics in Nutrition and Health Business: A Data Analytics Career Panel**

### **Description**

Nutrition and food related business and entrepreneurship requires constant product iteration, innovation, and creativity. The application of advanced technological programs can further improve the quantity, quality, and complexity of data and information available for this research and development. This panel will invite local business experts and entrepreneurs to discuss applications of data in product development and business planning for establishing healthier lifestyles and promoting more nutritious products.

### **Moderator**

Christine Kressirer is the Site Director of Tufts Launchpad BioLabs, a premier co-working facility for life science startups in Boston. Christine spent 7 years at the Forsyth Institute most recently as the Director of Core and Laboratory Services and 5 years at Arizona State University in research and laboratory coordination. She received her Ph.D. in Pharmaceutical Biology from Ludwig-Maximilians University in Munich, Germany and was a Postdoctoral Research Fellow at the Forsyth Institute and the Harvard School of Dental Medicine.

### **Panelists**

Benjamin Batorsky is the Associate Director of Data Science at MIT Sloan, where he leads data science projects for the Food Supply Chain and Analytics group. Previously he worked on the data science team at ThriveHive, where he scoped and built data products by leveraging multi-modal datasets on small businesses and their customers. In his work, he is often posed difficult business questions and is able to develop and execute a strategy for

answering them with either one-off analytic products or production-ready prototypes. He earned his PhD in Policy Analysis from the RAND Corporation, working on analytics projects in the areas of health, policy and infrastructure.

Erin Baumgartner is a local food nerd and entrepreneur. She is the CEO and founder of Family Dinner a local farmer's market delivery service. Family Dinner seeks to use data to improve the local food supply chain and eliminate waste in the system while highlighting the importance of local food through data visualization. Erin spent 11 years at MIT, most recently as the Assistant Director of the MIT Senseable City Lab, an Urban Science Lab within the Department of Urban Studies and Planning. She is also the former Director of the MIT-France Program, organizing scientific exchanges between MIT students, faculty and researchers and the French Scientific Community. Erin spent many years living and working in France and currently lives on Boxford, MA with her husband and co-founder Tim, and their toothless dog, Frank.

Dr. Svetlana Vinogradova is a Lead Data Scientist at InsideTracker, working with the Data Science team to integrate blood biomarkers and DNA data with physiological data from activity trackers to improve lifestyle recommendations and discover new patterns and optimal zones in sleep, heart rate, and blood biomarkers. Prior to Inside Tracker, Svetlana got her PhD in Bioinformatics and Mathematical Biology from Lomonosov Moscow State University and then completed a Postdoctoral training at Harvard Medical School and Dana-Farber Cancer Institute, where she worked as a bioinformatician developing statistical methods to study epigenetic mechanisms affecting gene expression. In addition to being a researcher and data scientist, Svetlana is an aspiring marathon runner and Boston marathon qualifier.

Marcia Hooper is a principal of Branch Venture Group, LLC, an angel investing group, focused on food startups, targeting food products, food technology, business services for food-related companies, ag-tech, and sustainability. She currently serves as a Senior Advisor to Bowside Capital, a private equity firm focusing in the small capitalization market. She has over 30 years of private equity and venture capital investing experience. She has served as a Director of over 30 private and publicly listed companies. She began her career at IBM in marketing. Ms. Hooper earned an MBA from the Harvard Graduate School of Business, a MA from Columbia University and an Sc.B. from Brown University.

Alfredo Morales works in understanding the complex behavior of social systems by analyzing big data with artificial intelligence algorithms, networks and complexity science. He explains large scale, societal behaviors, such as social segregation and polarization, by retrieving unstructured patterns of information from large datasets resulting from human activity on Internet, mobile phones and shopping data. In 2018 he was included in the list of 35 Innovators Under 35 by MIT Technology Review and in 2017 he was included in the list of Latinos of the Future by the journal El Planeta in Boston.

## **Workshop Descriptions**

### **An Intro to GIS using QGIS: Exploring Access to Healthy Foods in Cambridge, Ma**

#### **Presenter**

*Carolyn Talmadge* joined the Research Technology team at Tufts University in 2013 after graduating from Tufts School of Engineering with an M.S in Environmental Health. Carolyn currently works as the Senior GIS Specialist and teaches the GIS for Conservation Medicine course within the Cummings School of Veterinary Medicine. Carolyn manages the Data Labs and participates in GIS projects, classes and grants across all schools and departments. Carolyn's research interests focus on the One Health paradigm and using geospatial tools to investigate the spatial relationships between environmental health, animal health and public health issues facing us in today's world.

#### **Description**

Wouldn't it be great if ArcGIS was less expensive, easier to use, and more versatile? Never fear, QGIS is here! QGIS is a free and open-source Geographic Information System (GIS) software that allows you to create, edit, visualize, analyze and publish geospatial data on Windows, Mac, and Linux platforms.

This workshop will introduce users to introductory GIS concepts using QGIS. We will cover:

What is GIS and who is using it?

What are the different GIS softwares and why is QGIS a great option?

We will also do a hands-on activity in QGIS that involves exploring who has access to healthy foods in Cambridge, Massachusetts. This activity will teach users how to:

Add GIS data to QGIS – including shapefiles and excel data using Lat/Longs

How to symbologize (stylize) the data appropriately

How to use tools such as Select by Attributes, Select by Location and Buffers

How to create a final map composition including all necessary map elements.

### **Teaching Data Analytics to Non-Analytics Students**

#### **Presenters**

*Elena Naumova*, Tufts Friedman School of Nutrition Science and Policy

*Mingfei Li*, Bentley University

*Kevin Mentzer*, Bryant University



## **Description**

This session will provide an open forum for professors, researchers, and students to discuss techniques for teaching data analytics. We hope to identify ways to teach graduate students to solve complex problems, think critically, and effectively communicate across inter-generational, trans-disciplinary research terms. We will discuss a data-intensive, project-based learning approach that emphasize collaborative learning to design, evaluate, and disseminate research in team environments. In particular, we will discuss ways for students to serve as both leads and reviewers of their own and their peers' works.

## **Comparisons Across Statistical Software**

### **Presenter**

*Kyle Monahan*, Tufts University

## **Description**

This aim of this workshop is to teach students how to run, view, and extract various summary statistics in R software, focusing on some of the unique capabilities of this software package. Topics include importing external data (R/Excel/XPT/DTA files), generating summary tables, graphical exploratory data analysis, correlations, performing simple linear and log-linear regressions, regression diagnostics, and extracting statistics for use in scientific paper writing. This workshop will demonstrate a data workflow process and the packages used for completing statistical analyses in data science research. The workshop will provide hands-on experience using practice datasets. This practical session will primarily use R, though strengths and limitations will be discussed across other statistical packages. Learning resources for each statistical package will also be discussed so participants can expand their skills outside of the workshop.

## **Identifying Your Audience to Capture: Effective Communication of Data-Driven Research**

### **Presenter**

*Laurie LaRusso*, MS, ELS, has been writing about health and medicine for a variety of audiences and in a variety of formats for more than 20 years. Her publications and presentations run the gamut from clinical research and continuing medical education to consumer health and patient education. Her focus is crafting text and graphics to tell data stories. She is a past-president of the American Medical Writers Association—New England Chapter and a winner of the chapter's Will Solimene Award for Excellence in Medical Communication. She holds a Master's degree in Health Communication from the Tufts University Graduate Programs in Public Health; certification as an Editor, Life Sciences from the Board of Editors in the Life Sciences; and an adjunct faculty

appointment in the Nutrition Interventions, Communication, and Behavior Change program at Tufts University Friedman School of Nutrition Science and Policy.

### **Description**

This workshop offers best practices for communicating data sciences and analytics results to various audiences. Topics include: considering your intended audience; choosing the best format for the selected audience; extracting and highlighting the essential information to tell your data story; and utilizing visual communication to engage your audience. Participants will learn how to break down and disseminate important data and results for science and nonscience audiences.

### **What's In The Recipe: Model Development and Diagnostics**

#### **Presenter**

Ken Chui, Tufts School of Medicine

#### **Description**

Starting an analysis with a clear motive can enhance our efficiency and avoid modeling pitfalls. The aim of this workshop is to provide a survey of different modeling motives, including explanatory, descriptive, and predictive models, and how our chosen motive would affect our procedure. We will facilitate the understanding by demonstrating the different procedures on the same data set. After the workshop, attendees will be able to better identify the core motive of an analysis and comment on the use of modeling procedure. Pre-requisite: basic knowledge in hypothesis testing and linear regression.

### **Speaking Results Without Words: Data Visualization**

#### **Presenter**

Tania Alarcon Falconi, Environmental Health and Engineering

#### **Description**

This workshop aims to demonstrate techniques for effective data visualization using clear, detailed graphics. Topics covered include data visualization selection; when and when not to use graphics; proper fitting of data results to chart types; and stylistic choices to highlight important information. Participants will learn how to create and critique data visualizations.

### **Breaking Down Silos: Managing and Analyzing Data As A Team**

#### **Presenters**

Ye Shen, Tufts Food Aid Quality Review

Ilana Cliffer, Tufts Food Aid Quality Review

Devika Suri, Tufts Food Aid Quality Review

Breanne Langlois, Tufts Food Aid Quality Review

### **Description**

This workshop discusses the process of data management and analysis through the experience of the Food Aid Quality Review project, which conducted 3 large-scale field trials in Malawi, Burkina Faso, and Sierra Leone. Challenges, strengths, and lessons learned will be discussed. The session aims to provide participants with an understanding of how to work effectively as a team to conduct data management and analysis in complex field settings

### **Network Sciences and Complex Systems in Nutrition-Related Research**

#### **Presenter**

Sam Scarpino, Network Science Institute, Northeastern University

#### **Description**

This workshop aims to introduce complex systems and network analyses for application in nutrition science research. This includes tracking infectious disease outbreaks, analyzing spatial correlations of famine in complex emergencies, and modeling food systems. Network modeling will be presented conceptually and supported using examples of research from a broad audience.

#### **Oral Presentations**

#### **Group A: Nasir et al. Substance Use Disorder Outcome Predictions for Decision Support: A Holistic Data Approach**

Substance use disorder treatment has been an increasingly important issue in the last few decades, and with the recent opioid epidemic, it is getting more critical to better understand and tackle the problem. Each day, 130 people in the US die due to complications from substance use. In this work, we propose a novel data analytics approach to devise decision support tools for care providers using machine-learning models. Using the models, we also highlight potentially important factors and their dynamic relationships for the various outcomes. We also show that our model can provide useful information to augment decision making processes.

### **Group A: Wang et al. Sleeping Disorders and Alzheimer's Disease: Evidence for Heterogeneity**

Disturbed sleep has been well-known as a typical symptom of Alzheimer's disease (AD), showing a high prevalence among AD patients. However, emerging evidence indicates that disturbed sleep may also contribute to the risk of Alzheimer's. Experimental studies suggest that the sleep-wake cycle directly influences the levels of toxic proteins related to AD in the brain. Poor sleep quality and sleep deprivation encourage the accumulation of and impede the cleavage of toxic proteins linked to AD. Keeping up with the previous studies, the purpose of the present study is twofold: 1) to understand the relationship between sleep disorder and risk of late-onset of AD based on a larger sample size than in past work; 2) to investigate whether a particular sub-population of sleep disorders is related to a higher risk of late-onset of AD. Applying a Bayesian survival analysis (i.e., Bayesian approach to Cox-Gompertz model), this large-scale retrospective longitudinal study reveals that patients who suffer from sleep disorders (leaving out hypersomnia/insomnia) display a higher risk of late-onset of Alzheimer's disease after adjusting for other risk factors for AD including gender, ApoE genotype, and clinician-assessed health conditions. Our results suggest the existence of heterogeneity among sleep disorders in terms of how they affect the development of late-onset of AD.

### **Group B: Stark. What a “Batman Graph” shows about US income: A cautionary tale of correlation, principal components, and nonlinear regression**

This mini-case-study provides a striking example of why we need to graph the relationships we're analyzing. Via principal components analysis, we can summarize many US Census income indicators into a Wealth index and a Poverty index, to describe each of 33,000 ZIP codes. Simple correlation indicates a weak negative linear relationship between the two indices. A plausible next step is to check for nonlinear fits, but regression with quadratic and cubic terms hardly outperforms correlation. Visualization, however, reveals a complex and puzzling relationship that clearly requires more creative and resourceful analytic approaches.

### **Group B: Nasir et al. Synthetic Average Neighborhood Sampling Algorithm (SANSA): A Neighborhood Informed Synthetic Sample Placement Approach to Improve Learning from Imbalanced Data**

Machine-learning classification models are increasingly being used in both real-world applications as well as in academic literature. However, many real-world phenomenon happen much less often, and thus are more interesting and in many cases much more high-stakes to predict. In this work, we propose a new synthetic data generation algorithm that uses a novel “placement” parameter that can be tuned to adapt to the each datasets unique manifestation of the imbalance. SANSA also defines a novel modular framework to

rank, generate and scale the new samples, which can be used in the future with other functions to propose better methods.

**Group C: Ericson. Minds in Motion: Using positional data to study human behavior and cognition**

This talk provides an introduction to three analytical approaches for using positional data to gain insight into human behavior and cognition: (1) mouse tracking, (2) motion tracking, and (3) space syntax. Each approach is illustrated through brief presentations of ongoing research projects in human-computer interaction (HCI), spatial cognition, and environmental design. The first project uses mouse-tracking to examine the spatiotemporal dynamics of human decision-making during website use. The second project uses virtual reality (VR) motion-tracking data and directional (or circular) statistics to investigate human wayfinding and the geometry of our “cognitive maps.” The third project explores the limits of space syntax, a popular set of computational tools for predicting pedestrian walking patterns in urban and architectural environments. For each analytical approach, resources—including freely available software products and useful R packages—are noted for researchers interested in exploring the power of positional data in their own work.

**Group C: Abhilash et al. Building a Predictive Framework for Analyzing Opioid Patient Toxicology Data to Improve Medication-Assisted Therapy Adherence and Relapse**

In the United States, over 2 million people are diagnosed with OUD, resulting in costs exceeding \$500 billion each year. Pharmacological intervention by medication-assisted treatment (MAT) is most effective, based on minimizing withdrawal severity, reducing relapse, and improving retention in treatment. Despite demonstrated MAT effectiveness, there are differences in individual responses due to high dropout, non-adherence, and relapse. We have analyzed data from clinical toxicology tests over 18 months from 71 providers listing suboxone as the MAT on order for approximately 1,500 patients. For these patients, analysis of clinical toxicology results (over 512,000 tests) enabled determination of frequency of testing relative to MAT positives and unexpected positives of common drugs of abuse (including non-MAT opiates, amphetamine/methamphetamine, and benzodiazepines). Descriptive statistics demonstrate that testing frequency is related to adherence (measured by consistent appearance of MAT) and relapse (measured by unexpected positives). The analysis also found that more frequent testing showed decreased adherence and higher indicators of relapse with all drugs except benzodiazepines. In addition, significant associations between testing frequency, unexpected positives, and simulated adherence (i.e. dropping a drug directly into the urine) were observed among patients in a 6-month sample of the dataset. Finally, we identified geographically-distributed patterns of drugs showing unexpected positives that were consistent with public health data on drug use. Our ongoing work includes predictive analytics to identify patients at risk for adverse adherence and relapse outcomes based on testing history, drug metabolism, and analysis of unexpected positive rates—in addition to

using data to distinguish between metabolism and drug impurities and determine the parent drug that was ingested. We propose that this predictive method can improve adherence and retention of MAT patients by aiding evidence-based treatment determination.

#### **Group D: Bai & Wallbaum. Dual-dynamic Retirement Income Strategy for Better Retirement Outcomes**

As a response to the persistent low expected return market environment, the present study proposes a dual-dynamic retirement management strategy aiming at improving different retirement outcomes in the retirement portfolio decumulation context. Utilizing a Monte Carlo simulation approach, the present study reveals that the newly proposed dual-dynamic retirement strategy embedded with a target volatility asset management component could improve the survival rate of a retirement portfolio and provide more sustainable retirement coverage for retirees over the two hypothetical retirement spans – a twenty-year retirement scheme and a twenty-five-year retirement scheme – examined in this study. Therefore, benefiting from the target volatility mechanism, the dual-dynamic retirement solution that adjusts the portfolio asset allocation and the portfolio annual withdrawal rate simultaneously could be a new option for current retirement markets.

#### **Group D: Mentzer et al. The Role of Gender and Party on the Twitter Conversation in the 2018 U.S. Senate Elections**

This study examines the impact gender and party had on the Twittersphere conversation that occurred leading up to the 2018 U.S. Senate Election. We consider gender and party of not only the candidate but also the Twitter user. This allows us to dive into who is driving the conversation on Twitter. We propose a novel technique, using sentiment analysis, to measure affective polarizations.

#### **Group E: McGuirk. Business Ethics at the Core of Successful and Sustainable Analytics Practices**

The International Institute for Analytics has included 'ethics of analytics' as a top five business imperative in both their 2019 and 2020 Analytics Predictions and Priorities list. Interestingly, this heightened need to focus on ethical practices coincides with a separate industry and business priority to ramp up efforts to identify new ways to collect consumer data and apply analytics and data science techniques on this data to gain a competitive advantage. In fact, in a recent study completed by Forrester Research, they found that over 50% of the companies surveyed are planning to launch initiatives to expand their ability to source external data and that many of these firms will appoint 'data hunters' to lead these initiatives. In this session we will dive into this critical moment in time for the analytics and data science communities. A time when an increased reliance on these practices will intensify the need to incorporate sound judgement and ethical business practices across all facets of analytics operations. Unfortunately, there have been too many recent cases of

companies committing transgressions that quickly erode consumer trust in their ability to properly manage data and deploy ethical analytics practices. In 2018, Facebook enabled Cambridge Analytica to use millions of members' personal data without their consent for targeted political advertising and in 2019 Goldman Sachs has been under fire for allowing blatant gender bias in algorithms used to establish credit limits for Apple Card customers.. This session will explore the important role educational institutions can play helping to inspire and empower students to use ethical and socially responsible data collection and analytic practices. It will also explore how industry can utilize cross-functional data governance teams to ensure diverse and empathetic perspectives are considered when deciding what data should be collected, analyzed, and used to support insight-driven decision-making.

## **Poster Presentations**

### **Hsu et al. Differences in nutrient intake between individuals with and without familial longevity**

The contribution of diet upon the ability to reach extreme ages remains unclear, in part, because centenarians (people over the age of 100) may have markedly changed dietary patterns at the end of life. Studying centenarian offspring, who are predisposed to longer and healthier lives, allows for the investigation of diet earlier in the life course when it is more likely to have an effect on longevity. The primary objective of the study was to assess the difference in nutrient intake between centenarian offspring and a referent cohort in the New England Centenarian Study. Semi-quantitative food frequency questionnaire data were collected on 280 centenarian offspring and 129 referent participants without familial longevity (mean age 72.6 years). The data were converted to 103 nutrient measurements. Wilcoxon rank sum test and generalized linear regression were used to evaluate the association between each nutrient measurement and cohort (offspring vs. referent). Principal component analysis identified eight components thus a corrected p-value of  $< 0.00625$  was considered to be statistically significant. After adjustment of age, sex, and total caloric intake, centenarian offspring had higher intakes of dietary sources (i.e., not including supplements) of iron ( $p = 0.0060$ ), niacin ( $p = 0.0016$ ), riboflavin ( $p = 0.0016$ ) and zinc ( $p = 0.0026$ ) than referents. In conclusion, centenarian offspring have a higher reported intake of some vitamins and minerals, particularly those associated with energy metabolism and oxidant levels, in comparison with individuals without familial longevity. In addition to the genetic contributions to achieving extreme ages, intake of iron, niacin, riboflavin, and zinc may contribute to the longer life and health spans of offspring of centenarians. Future research should investigate how intake levels of these nutrients by centenarian offspring compare with nationally representative data and recommended daily allowances as well as metabolite measurements from blood.

## **Gregory & Byrd. Using Geographic Information Systems (GIS) to Determine Sites for SNAP-Ed**

Mississippi State University Extension Office of Nutrition Education (ONE) is utilizing geographic information systems (GIS) technology to determine site location eligibility for Supplemental Nutrition Assistance Program Education (SNAP-Ed) programming among low-resource audiences. Methods to determine eligibility for SNAP-Ed include counties determined by the USDA to be StrikeForce, in persistent poverty, or child persistent poverty. Any county that meet one of these criteria received approval for programming at all locations within the county. Counties that do not meet any of the three criteria could use approved schools (public schools whose enrollment was 50% or more free/reduced lunch) and locations allowed by SNAP-Ed guidance (SNAP offices, WIC offices, food banks/pantries, county health departments, etc.) for programming. It was proposed to use GIS technology (GIS Online) to determine eligibility for sites that did not meet any of the above criteria. Shapefiles were created based on the following: areas within a one-mile radius of an approved school and areas within a one-mile radius of a census tract where 50% or more of the population's income was below 185% of the federal poverty level. The two shapefiles were added to an existing map layer of Mississippi counties. Sites that intersect with either of these areas are considered eligible for SNAP-Ed programming. Site locations can be validated in two ways: using a search engine for a single site or uploading addresses to be geocoded for multiple sites. Using these additional criteria to determine eligibility has increased the number of sites where ONE employees can conduct SNAP-Ed. Using GIS technology can help ONE identify eligible sites that are not currently being served. Future plans include adding additional criteria for eligibility and creating an interactive mapping tool ONE employees can use to determine eligibility without having to submit requests to the state office.

## **Chow et al. Use of advanced statistical analysis in determining death manner: lessons from 330,000 national violent death data**

Suicide is the tenth leading cause of death in the US, over 48,000 Americans died by suicide in the year 2018. The number might still be underreported, since identifying the intent of death is difficult in the absence of a suicide note. Identifying intent becomes problematic specifically for opioid overdose cases given the drug's abilities to cause respiratory depression. The ability to accurately identify and report on the prevalence of suicide has important implications for policy development in suicide prevention. The NVDRS is a state-based surveillance system funded by the Centers for Disease Control and Prevention (CDC) to collect data on violent deaths from participating states. Currently CDC allows use of restricted access data to researchers for a preapproved analysis. In this study, we analyzed ~330,000 violent death data from 37 participating US states for 15 years (2003-2017). Using NVDRS data, we aim to identify and validate predictors of suicide in the opioid overdose cases through a combination of qualitative and quantitative analyses. The



quantitative analysis begins with results from contingency tables examining the association between opiates and the manner of death coded by CDC staff. Next, we will perform a logit model predicting the probability of suicides and homicides resulting from opiates indicated as a causal factor controlling for socio-demographic factors. The qualitative method relies on the narratives generated from law enforcement and medical examiner reports. We identified a list of word/themes that might be distinctly present in suicide compared to homicide. This pattern/theme can later be used to develop a methodology for natural language processing and statistical text analysis to investigate the differential use of language (e.g., thematic content, syntax, sentiment, etc.) across death narratives and construct a measure of descriptive similarity between narratives within and across groups of decedents.

### **Hur et al. Combining unstructured social media data with structured scientific literature: a data merging problem**

Marijuana legalization efforts have been increasingly successful in the United States. There are 33 states (plus the District of Columbia) that have approved medical marijuana in recent years. With the rising popularity of medical marijuana in a variety of conditions, we aimed to perform a systematic review of clinical studies on biological indicators of cannabis dosing and administration in pain management and opioid withdrawal. A comprehensive systematic search strategy was applied to identify relevant studies from medical literature databases. All identified documents were screened through a three-stage process. Starting with 338 papers, we ended up with 32 relevant studies. Due to a limited number of scientific publications in this specific field, we optimized and implemented an automated data extraction of clinical trial data from [clinicaltrials.gov](https://clinicaltrials.gov). After multiple iterations of filtering due to irrelevant conditions and interventions, 35 trials were finalized for the analysis. Among the 35 trials, 7 are opioid sparing trials and 28 trials use cannabis in pain management. Analysis of the scientific literature and clinical trials indicated that common uses of cannabis products in a variety of indications do not match with experimentally-validated supporting evidence, leading us to compile social media content relevant to cannabis use in pain management and opioid withdrawal. We are currently performing ontology analysis to discover the origin of the most prevailing themes in common uses of cannabis products. The end goal is to identify sources of social perception on use of cannabis in a variety of indications – ultimately culminating in a comparative analysis between scientific evidence and experience-based evidence in pain management by cannabis.

### **Deichmann et al. European Moods: Satisfaction levels from 2003 to 2016**

As the referendum-mandated departure of the United Kingdom (“Brexit”) from the European Union (EU) continues to unfold, international discourse surrounds the tradeoffs of EU membership, and whether European citizens perceive the benefits to be worth the costs. The preamble to the 1957 Treaty of Rome calls for “constant improvement of the

living and working conditions” of member state citizens as well as collective action to reduce “differences existing between the various regions and the backwardness of the less favored regions.” This project will employ survey responses from all four European Quality of Life Survey (EQLS) iterations (2003, 2008, 2012, and 2016) in order to examine the extent to which enlargement helps meet the EU objective of improving living standards and the overall quality of life across the continent, with particular reference to the post-Communist New Member States (NMS) that joined the EU since 2004. The data set includes forty response variables across nine dimensions for twenty-eight EU member states, along with eight non-member states. Insights are captured through the systematic comparison of self-reported perceptions pooled at the country level before and after accession, as well as between member states and non-member states.