September 2024 • Issue #567

AMSTATNEWS

The Membership Magazine of the American Statistical Association • *http://magazine.amstat.org*

Special Issue on Data Visualization

Featuring:

Catching Up with Nathan Yau: 14 Years Later

Data Exploration Through Time

Understanding How Viewers Interpret and Interact with Statistical Charts

DIONNE PRICE
PUBLIC LECTURE SERIES

Honoring the Legacy of Dionne Price and Her Commitment to Public Good

In honor of 2023 ASA President Dionne Price, the ASA established the Dionne Price Public Lecture Series.

Dionne chose the theme "One Community: Informing Decisions and Driving Discovery" for the Joint Statistical Meetings because of her deep commitment to working for the public good.

The ASA launched a $75,000 endowment campaign to fund the lecture series and ensure Dionne's name remains synonymous with promoting the practice and profession of statistics for generations to come.

The Lecture Series Aims to:

- Highlight the impact of statistics and data science on society, the sciences, and the public good
- Inspire future statisticians and data scientists
- Foster interdisciplinary discussion
- Showcase the pivotal contributions of early-career professionals to advances in science and their positive impact on society

Visit *bit.ly/3zBjsOc* or scan the QR code to donate today! Your support will honor Dionne and make this series possible.

Support the Lecture Series

Your support funds annual lectures, travel, venue/livestreaming costs, and lecturer honorariums.

Together, we will inform decisions and drive discovery through the Dionne Price Public Lecture Series.

AMSTATNEWS

SEPTEMBER 2024 • ISSUE #567

Amstat News welcomes news items and letters from readers on matters of interest to the association and the profession. Address correspondence to Managing Editor, *Amstat News,* American Statistical Association, 732 North Washington Street, Alexandria VA 22314-1943 USA, or email *amstat@amstat.org*. Items must be received by the first day of the preceding month to ensure appearance in the next issue (for example, June 1 for the July issue). Material can be sent as a Microsoft Word document, PDF, or within an email. Articles will be edited for space. Accompanying artwork will be accepted in graphics file formats only (.jpg, etc.), minimum 300 dpi. No material in WordPerfect will be accepted.

Amstat News (ISSN 0163-9617) is published monthly by the American Statistical Association, 732 North Washington Street, Alexandria VA 22314-1943 USA. **Periodicals postage paid** at Alexandria, Virginia, and additional mailing offices. POSTMASTER: Send address changes to *Amstat News,* 732 North Washington Street, Alexandria VA 22314-1943 USA. Send Canadian address changes to APC, PO Box 503, RPO West Beaver Creek, Rich Hill, ON L4B 4R6. *Amstat News* is the member publication of the ASA. For annual membership rates, see *www.amstat.org/join* or contact ASA Member Services at (888) 231-3473.

American Statistical Association
732 North Washington Street
Alexandria, VA 22314–1943 USA
(703) 684–1221
ASA GENERAL: *asainfo@amstat.org*
ADDRESS CHANGES: *addresschange@amstat.org*
AMSTAT EDITORIAL: *amstat@amstat.org*
ADVERTISING: *advertise@amstat.org*
WEBSITE: *http://magazine.amstat.org*

The American Statistical Association is the world's largest community of statisticians. The ASA supports excellence in the development, application, and dissemination of statistical science through meetings, publications, membership services, education, accreditation, and advocacy. Our members serve in industry, government, and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare.

member news

SPECIAL ISSUE ON DATA VISUALIZATION

Call for Nominations: 2025 Ethel Newbold Prize

Members of the Bernoulli Society for Mathematical Statistics and Probability Newbold Prize Committee invite nominations for the 2025 Ethel Newbold Prize. Established in 2014 and supported by Wiley, this biennial prize honors the significant contributions of women to the statistics field. However, the award recognizes excellence in statistics without regard to the gender of the recipient.

The prize will be awarded to an outstanding early or mid-career scientist whose work demonstrates excellence in mathematical statistics or research that links developments in a substantive field to new advances in statistics. It consists of €2,500 and an award certificate. The recipient will also be invited to present a talk at the next Bernoulli-IMS World Congress, a Bernoulli-sponsored major conference, or the ISI World Statistics Congress.

Send nominations and inquiries to Adrian Röllin at *adrian.roellin@nus.edu.sg*. Nominations should include a letter outlining the nominee's achievements and contributions and a recent curriculum vitae. The deadline for submissions is November 30, and the winner will be announced in early 2025. For more information, visit *www.bernoullisociety.org/index.php/prizes?id=207*.

SPECIAL ISSUE ON DATA VISUALIZATION

Connect with us on LinkedIn
www.linkedin.com/company/american-statistical-association---asa

Join the ASA Community
http://community.amstat.org

Like us on Facebook
www.facebook.com/AmstatNews

Follow us on Instagram
www.instagram.com/AmstatNews

Subscribe to our YouTube channel
www.youtube.com/user/AmstatVideos

Visit *Amstat News* online

My JSM Story: A Journey of Collaboration and Growth

Madhumita (Bonnie) Ghosh-Dastidar

As I traveled to this year's Joint Statistical Meetings in Portland, Oregon, I could not help reflecting on my journey—a journey marked by countless hours of anticipation, preparation, and collaboration. The meetings, known to most as simply JSM, is the world's largest gathering of statisticians and data scientists. It is more than just an event to many of us—it is an opportunity to celebrate shared passions; shape the future of our field; and network, learn, and grow!

This year, my JSM experience began when I gaveled the start of the ASA Board Meeting on Friday morning. We had a full agenda including a financial report, reports from the Council of Sections and Council of Chapters representatives, and updates on ongoing activities. Sarah Cumbers, chief executive of the Royal Statistical Society and a guest at our meeting, shared many insights and—most importantly—reminded me that our opportunities and issues are global.

We were joined by ASA Past President Kathy Ensor and Dave Hunter, who provided updates about the ASA's role in the accreditation body CSAB. Our participation as a member society has ensured the ASA has a strong voice in the development and implementation of data science program accreditation.

We also had a sneak peek at the episodes planned for JSM TV, a new way to promote the very latest in everything statistics. All the episodes are available on YouTube at *https://tinyurl.com/3m39jt96.*

One of the most rewarding aspects of serving as ASA president is the opportunity to collaborate with so many smart and passionate individuals. After the packed agenda and productive deliberations of the ASA Board meeting, I felt immediately reenergized by Sunday's First-Time Attendee Orientation and Reception. I was inspired by the energy and enthusiasm of the students and early-career attendees. Their passion and fresh perspectives are a reminder of why we do this work. Reflecting on my first JSM, the colleagues I met have remained important to my personal and professional development. I know the same will be true for the JSM 2024 first-timers.

On Monday, I had the pleasure of representing the ASA by recording a segment for JSM TV and moderating the President's Invited Speaker presentation by Jason Matheny, with whom I am fortunate to work. As president and chief executive officer of RAND, Jason provides visionary leadership that blends an intense hope for the future with recognizing its threats and challenges. His talk served as a powerful reminder to statistical practitioners of the resilience and determination that drive our work. I think you will agree with his message: "(Our) profession has saved the world several times over." This year, I saw many ways we are building on these past efforts and successes.

On Tuesday morning, I valued meeting participants of the Diversity Mentoring Program, organized by a dedicated organizing committee led by Prince Allotey and Stephanie Tillman. They discussed topics such as strategies and skills to succeed in the job market and career paths in statistics and related fields.

It was also incredible to attend the Committee of Presidents of Statistical Societies annual meeting that day with presidents of other statistical

View all the episodes of JSM TV at *https://tinyurl.com/3m39jt96.*

Download a PDF copy of the awards book at *https://tinyurl.com/48erxmbs*.

societies. There are so many exciting initiatives being organized nationally and internationally across these societies. Both gatherings reinforced my belief in the power of mentoring, collaboration, and community.

Tuesday evening, I delivered my presidential address to our association and hosted the awards ceremony. I felt a mixture of trepidation and excitement. The first row was filled with family, colleagues, and friends, but there were also many new faces. I spoke about my commitment to evidence-based policy and the challenges faced, progress made, and opportunities that lay ahead. Following the talk, it was an honor to recognize the award recipients and 2024 ASA fellows. If you have not read the citations, I encourage you to look at the awards book (download from *https://tinyurl.com/48erxmbs*). The citations provide evidence of the impact we are making.

On Wednesday, I attended the Committee of Presidents of Statistical Societies award ceremony and was filled with pride for what the recipients had accomplished. The talk by Rob Tibshirani was inspiring and forward-thinking.

I ended my day by attending a reception honoring the JSM Program Committee, ably led by program chair Debashis Ghosh. The work of this group and countless others ensures JSM is a chance to come together, learn from each other, and build on our collective knowledge and experience. In my keynote, I highlighted three of my superheroes. The JSM Program Committee is made up of similar superheroes!

It was a rewarding experience to have the opportunity to talk with many new colleagues throughout JSM, and I am continuing to enjoy reading the many reflections being shared on LinkedIn. Searching for #JSM2024 and reading the posts illustrates the breadth and depth of the presentations at JSM. I cannot resist highlighting the following few:

- "Ever been to a buffet so massive you have no idea where to start? That's what it feels like attending a conference with hundreds of parallel sessions!"
- "#JSM2024 gave me the amazing opportunity to meet one of the greatest Statisticians of the era, Prof. Robert Tibshirani, who proposed the LASSO Regression method which made a huge impact to the field of Statistics. He is a renowned Statistician who has more than 500,000 citations."
- "Finally, I got to organize my first JSM contributed session on EEG data!"
- "#JSM2024 highlight: panel discussion focused on the evolving intersection of statistics & artificial intelligence"
- "With > 5k attendees, it was filled with exciting new research in single-cell and spatial 'omics, survival analyses, tensors, LLMs, and a ton of other exciting areas. It was great meeting a ton of incredible statisticians behind this research, and plus, I got to participate in a JSM basketball game!"
- "As an undergraduate student, I had such a great, fun (and hard) time studying Christian Robert's 'Monte Carlo Statistical Methods.' On Tuesday, I was incredibly lucky to be presenting a #MCMC method of mine in front of him, at American Statistical Association – ASA. Sometimes dreams do come true :-)."
- "Now I see why they call JSM the grand annual catch-up. Portland was such a sweet city! Excited for JSM 2025!"

As JSM 2024 ended, I felt a sense of fulfillment and pride. Also, I felt inspired—by the people I met, the ideas I encountered, and the possibilities I see ahead. The ideas and insights we shared are just the beginning—now it is up to us to take them forward, implement them in our work, and continue pushing the boundaries of our field. This has further reinforced my belief in the power of community and collaboration and reminded me we can drive positive change and have impact through the power of statistics!

M. Ghosh Dastidar

New Lecture Series Honors **Legacy of Dionne Price**

Ron Wasserstein, ASA Executive Director, and Donna LaLonde, ASA Associate Executive Director

"We have much to celebrate in April and beyond as we continue to inform decisions and drive discoveries."

Dionne Price, President's Corner, *Amstat News*, April 2023

This sentence is how Dionne Price concluded her April 2023 President's Corner column. She wrote the column in recognition of Mathematics and Statistics Awareness Month, but it was a guiding principle of her presidential year and throughout her career. Price passed away in February 2024. As we considered how to honor her legacy, we were inspired by those words.

Starting in 2025, the Dionne Price Public Lecture will be a celebration of our profession, which informs decisions and drives discoveries, and will advance the mission of promoting the practice and profession of statistics. The annual lectures will be held at locations throughout the United States and livestreamed for broader impact and engagement.

When Price introduced her focus for her year as president of the American Statistical Association, she wrote, "I have been employed by the US Food and Drug Administration since beginning my career as a statistician, and the reason for the longevity is simply my belief in the FDA's mission. Similarly, I believe in the mission of the ASA, which is promoting the practice and profession of statistics. Thus, 2023 will be a mission-driven year informed by ideas shared by our community."

In the April 2023 President's Corner, she also wrote, "The *Journal of the American Statistical Association* was established in 1888 and is considered the premier journal of statistical science. The breadth of our impact is highlighted by the most-read articles within the last 12 months. The research presented includes a machine learning approach to measuring housing vitality, an exploration of A/B testing and causal effects evaluation, and an investigation of causal inference for social network data."

Writing for the members of the ASA, she highlighted the profession's crucial role in shaping scientific research, informing public policy, and driving societal progress. These roles can be emphasized through a public lecture dedicated to explaining the impact of statistics and offering examples of statistical insights that have influenced policy decisions and scientific breakthroughs.

In another President's Corner, Price shared this story: "I fell in love with biostatistics during a summer internship as an undergraduate. I was an applied mathematics major searching for a direction, and an internship at the National Institutes of Health ignited my passion for statistics. The rest, as we say, 'is history.' My passion for the 'practice and profession of statistics' has grown stronger because I am fortunate to work with a talented group of statisticians and other professionals and able to serve our community."

She showed her gratitude for her mentors by being a mentor and champion of early-career statisticians and data scientists. She was equally passionate about public service. To honor this commitment, the lecture will be given by a "rising star" contributing to public welfare through statistical work.

Price understood there are many misconceptions about statistics. For many, it is viewed as a dry and complex subject. Throughout her career, she devoted time and energy to countering these misconceptions, especially with students. She inspired young people to choose educational paths that prepared them for roles as statisticians and data scientists. The public lecture series will continue this work by presenting concepts in an engaging and accessible manner.

Price spent her professional career at the FDA. At the time of her passing, she was the deputy director of the Office of Biostatistics in the Office of Translational Sciences of the Center for Drug Evaluation and Research. Of her work, she wrote, "My greatest accomplishment is the daily knowledge that my statistical leadership and expertise positively contribute to ensuring the efficacy and safety of drugs

Scan to donate to the lecture series or visit *https://secure.qgiv.com/for/dionnepricepubliclectureseries.*

and therapeutic biologics for the public. As a statistician, I make a difference in the lives of others. This is an honor and a joy!"

In her work, she was committed to promoting collaboration between statisticians and experts in other domains. She is remembered for her ability to bring together diverse perspectives, foster understanding of complex issues, and encourage innovative solutions. A public lecture shows the power of collaboration, a hallmark of Price's approach to science and policy.

In summary, a public lecture series dedicated to the impact of statistics on science and society holds significant value. By enhancing public understanding of statistical principles, informing policy and decision-making, inspiring future generations, and promoting interdisciplinary collaboration, such a series can contribute to a more informed, engaged, and innovative society. As we continue to navigate an increasingly complex and data-rich world, fostering a deeper appreciation for statistics is more important than ever. Through public lectures, we can illuminate the vital role of statistics in shaping our understanding of the world and driving progress in myriad domains.

The Dionne Price Public Lecture series is supported by donations, which are used to pay for expenses associated with an annual lecture, including an honorarium and travel expenses for the lecturer, the venue, and livestreaming. Price chose the theme *One Community: Informing Decisions and Driving Discovery* for JSM 2023 because of her deep commitment to working for the public good. This public lecture series will honor her by continuing this work. ■

Application Open for Journal Scholars Exchange Group Volunteers

The ASA and American Association for Public Opinion Research collaborated to create a program that prepares the next generation of researchers for the professional and academic world. Through a structured two-month program, students will receive mentoring, engage in professional development, and learn presentation best practices.

Applying and being selected for this working group requires a commitment of one year, starting in late summer, for the program to launch in fall 2024. Volunteers will do the following:

- Participate in planning sessions to define program objectives and themes for upcoming quarters
- Research and recommend potential speakers, topics, and activities for professional development workshops
- Help organize networking events and facilitate peer feedback sessions during club meetings
- Serve as ambassadors for the Journal Club program and promote student engagement and participation

All applicants will receive a response once the selection period has expired, regardless of decision. For questions, email Ryan Green at *rgreen@aapor.org* and Kristin Mohebbi at *kristin@amstat.org*.

Apply at *https://bit.ly/3YBYdpW.*

Pilot Program to Allow Exception to One Session Policy at JSM 2025

Elizabeth Mannshardt, ASA Membership Council Vice Chair

JSM program rules restricting speakers to one session may result in speakers choosing between speaking about technical work and professional development. These restrictions may affect the speaker and ASA community, as the community misses out on insights, expertise, and opportunities for increased diversity across sessions.

Members of the American Statistical Association Membership Council and Joint Statistical Meetings Policy Committee—who work to increase diversity across technical and professional development sessions—created a pilot program that allows exceptions to the one session policy for JSM 2025. The program aims to increase participation across both types of sessions and evaluate the impact of speaker restriction on the success of professional development and technical sessions.

For details or to apply for a JSM 2025 professional development session speaker exception, visit the ASA Community website at *https://bit.ly/46CPqGe*.

JEDI CORNER

Statistics, Science Organizations Foster DEI in Wider Community

The Justice, Equity, Diversity, and Inclusion (JEDI) Outreach Group Corner is a regular component of Amstat News *in which statisticians write and educate our community about JEDI-related matters. If you have an idea or article for the column, email the JEDI Corner manager at* jedicorner@datascijedi.org.

The statistical community embraces a wide variety of interests and includes many organizations that express a strong commitment to the principles and actions of justice, equity, diversity, and inclusion. This month, JEDI Corner looks at some of these groups—their organizational parameters and missions—and highlights their DEI initiatives.

The Statistics Society of Canada/Société Statistique du Canada has long been a DEI leader. It's even in their name and throughout their work, as they serve the science and practice of statistics and data science in Canada in French and English. Their Committee on Equity Diversity and Inclusion carries out the mission to support these principles and actions in their work as statisticians and data scientists in service to society.

The International Biometric Society has separate regional organizations in North America. The Eastern North American Region serves roughly anywhere east of about Denver, Colorado, and Regina, Saskatchewan. Their diversity liaison promotes diversity and inclusion, especially through meetings and conferences. The Western North American Region formed a JEDI Committee in 2021. Its activities include an annual diversity workshop; advising WNAR on equity, diversity, and inclusion matters; and coordinating travel awards for Indigenous students.

Justice, equity, diversity, and inclusion is championed by leading statistical institutes, as well. Members of the National Institute of Statistical Sciences have written a multifaceted statement on diversity, equity, and inclusion stating their commitment to areas including nondiscrimination, collaboration, and accountability. This statement can serve as a guide for organizations developing JEDI initiatives.

JEDI principles and actions also guide the mission of the Institute of Mathematical Statistics through its Committee on Equality and Diversity.

Because justice, equity, diversity, and inclusion are important in all sciences, the American Association for the Advancement of Science works to foster these principles across the scientific landscape. It plays a significant role in bringing people together across disciplines and areas of interest. Their annual DEI report thoroughly explores DEI issues and progress in the sciences and includes useful data and infographics.

These are just a few of the many organizations sharing a common mission to foster justice, equity, diversity, and inclusion in statistics and data science and serving as a resource for anyone seeking to promote JEDI principals and actions. Reaching out to one or more can support your own JEDI mission. ■

David Corliss is the principal data scientist at Grafham Analytics and founder of Peace-Work.

Connect and Learn at StatFest 2024

StatFest is a free, one-day conference taking place September 21 at the Columbia University School of Social Work. The aim is to encourage undergraduate students from historically underrepresented backgrounds to consider careers and graduate studies in the statistical sciences. Students can expect the following:

- Hear from graduate students and professionals across a broad range of careers in the statistical sciences
- Gain tips for networking, finding summer programs, and applying to graduate school
- Meet with representatives from graduate programs, summer internship programs, and potential employers

For more information and to register, visit *https://bit.ly/4dgNCFc*.

My ASA Story:

Jonathan Auerbach

Assistant Professor

My ASA journey began more than a decade ago. I had recently graduated from Cornell University with a degree in economics and was working for the New York City legislature—first as an aide to the council member representing Harlem and then as an analyst for the speaker in a unit responsible for data analysis and statistical modeling, among other tasks.

My experience in local government was formative. The Great Recession was over, but governments were still picking up the pieces. I gained an appreciation for the power and limitations of government, which would shape my research interests and future participation in ASA activities.

I attended the Eric and Wendy Schmidt Data Science for Social Good program at The University of Chicago in 2013. Following the program, several of us were interviewed for *Amstat News*, which was when I first became an ASA member.

I have since taken advantage of many ASA opportunities. To name a few, I was the 2016 ASA/STATS.org Public Understanding of Statistics Fellow and the 2021–2022 ASA Science Policy fellow. I won the *Significance* Young Statisticians Writing Competition in 2014, received the ASA Wray Jackson Smith Scholarship in 2015, and won the professional category of the ASA Government Statistics Section Data Challenge in 2016 with Chris Eshleman and Rob Trangucci.

These opportunities provided recognition far beyond the ASA community. For example, my work for the *Significance* competition was covered by *The New York Times,* and our work for the data challenge was covered by *Governing Magazine.*

The ASA has provided many meaningful volunteer opportunities. For instance, I was a judge for the ASA Government Statistics Section Data Challenge in 2017, the ASA Fall Data Challenge in 2020, and the ASA/Washington Statistical Society Data Visualization Poster Competition in 2022. For these and similar efforts at George Mason University, I was an inaugural recipient of the Trailblazer Award from the George Mason University College of Engineering and Computing Office of Diversity, Outreach, and Inclusive Learning in 2022, and I received the United Bank Presidential Medal for Faculty Excellence in Diversity and Inclusion in 2023.

I earned my PhD in statistics from Columbia University in 2020, where Shaw-Hwa Lo and Andrew Gelman advised me. I spent my final year as a Graduate School of Arts and Sciences Teaching Scholar, creating and teaching the course Statistics for Activists.

I then served as the 2020–2021 ASA Science Policy Fellow under the guidance of ASA Director of Science Policy Steve Pierson. Among many incredible experiences, I conducted research on US Census operations and voting by mail (see "What Would Happen if the Deadline for the 2020 Census Data Collection Operation Changed? Estimates of Apportionment of the House of Representatives and Distribution of Federal Medicaid Funding Under Different Deadlines" and "Does Voting by Mail Increase Fraud? Estimating the Change in Reported Voter Fraud when States Switch to Elections by Mail") and helped draft amicus briefs for consideration by the Supreme Court.

I also participated in an independent review of the 2020 Census with Paul Biemer and Joe Salvo, overseen by the ASA Census Quality Indicators Taskforce. This and related work led to a series of publications (i.e., "What Protects the Autonomy of the Federal Statistical Agencies? An Assessment of the Procedures in Place to Protect the Independence and Objectivity of Official US Statistics," "Safeguarding Facts in an Era of Disinformation: The Case for Independently Monitoring the US Statistical System," and "Bolstering Education Statistics to Serve the Nation") and interviews (i.e., *SciLine, HuffPost,* and a Reddit "Ask Me Anything").

In 2021, I joined the department of statistics at George Mason University, where I continue to work with the ASA

Office of Science Policy. We just released a report on the state of the federal statistical system titled *The Nation's Data at Risk: Meeting America's Information Needs for the 21st Century*. I learned an incredible amount from coauthors Steve Pierson, Claire Bowen, Connie Citro, Nancy Potok, and Zachary Seeskin.

The report follows a century-old tradition in which ASA researchers independently evaluate the federal statistical system (e.g., *Government Statistics: A Report of the Committee on Government Statistics and Information Services Sponsored by the American Statistical Association and the Social Science Research Council*). Our report is unique in that we propose to monitor the system proactively. In other words, just as civil engineers monitor America's physical infrastructure, we aim to monitor America's data infrastructure, focusing on the federal statistical system.

I am the current president of the Washington Statistical Society, the largest ASA chapter. Now is a particularly exciting time to be a member as we prepare for our 100th anniversary. I am grateful to be part of such a welcoming and distinguished group of scholars. While the chapter has several important responsibilities, I am particularly interested in engaging the many small groups of statisticians who work throughout the federal statistical system, as well as K–12 teachers in the area.

Reflection has given me the opportunity to remind myself of how lucky I am to belong to an incredible community. Like most, I am indebted to a lengthy line of mentors and supporters throughout my journey, including Ray Majewski, Leslie Hirsch, Shaw-Hwa Lo, Tian Zheng, Andrew Gelman, Michael Sobel, Steve Pierson, Ron Wasserstein, Jiayang Sun, Bill Rosenberger, and Anand Vidyashankar. ■

New Member Spotlight
IAN LANE

This month, we welcome Ian Lane, who answered the following questions so we could get to know him better:

How did you become interested in statistics and/or data science?
I am a nurse and, as I realized how important it was to read the scientific literature in health care to keep abreast of new knowledge, I began to see people sort of skim or jump over the methodology sections entirely. I tried to read them but could never grasp what I was seeing, so became determined to familiarize myself with the language of research design and then analysis. My first foray into statistics was before graduate school, trying to make sense of these otherwise incomprehensible scientific methods sections of research articles. Once I took my first of many statistics courses, I was hooked.

What do you hope understanding statistics and/or data science helps you accomplish?
I am primarily interested in the application of statistics and data science methods as they pertain to my own and others' research within nursing science. We have more than 3 million nurses in the US alone (roughly six times the number of physicians), and 1% have a doctoral degree and conduct some form of research or lead quality improvement initiatives. Also, virtually all US medical care is administered through nurses. So how much of 'biomedical' outcomes are actually nursing outcomes, which are never studied from a nursing perspective, remains to be seen. Yet nurse scientists do not feel comfortable, generally, with quantitative methods. As a unique scientist in nursing with expertise in epidemiologic methods, my goal is to use statistical methods and data science knowledge to arm nurse researchers with the tools to advance the knowledge of health care, specifically in nursing.

Is there a particular group of statisticians you would like to reach out to you?
I would be open to contact from any statistician, frankly. That said, given my interests, I feel anyone using biostatistical methods to study health care or epidemiologists developing statistical methods for epidemiological questions would be right up my alley.

What is your favorite hobby?
My husband and I enjoy working out together consistently and doing outdoor activities such as hiking or cycling.

MORE ONLINE
View the complete interview and full list of new members at *https://magazine.amstat.org/blog/category/a-statisticians-life/new-member-spotlight.*

Catching Up with **Nathan Yau:** 14 Years Later

Nathan Yau

https://flowingdata.com

All graphs in this article are courtesy of FlowingData

In 2010, we interviewed Nathan Yau and asked him about his then fairly new data visualization website, FlowingData. Since then, he has become an expert in data visualization, earned his PhD, written a couple of books (*https://flowingdata.com/books*), and expanded his website to include membership, newsletters, and tutorials. We recently caught up with him to see what he has learned since 2010 and what he can teach statisticians and data scientists about data visualization and information design.

What has been the most interesting or challenging part of maintaining FlowingData?

When we talked last, FlowingData was a side project I worked on for fun. Since graduating, FlowingData has become my full-time job, so the site has evolved in its 17 years. The site is 100% member-supported now, I publish a lot more of my own data projects, and I wrote books based on my work.

Recently, you published the second edition of your book, *Visualize This*. You wrote on Twitter/X, "The second edition, out in June, is still that step-by-step guide. Concrete practical examples. A variety of tools. A data-centric approach. But every bit has been updated. It turns out a lot can change." Can you give us some specific examples of what has changed from your first edition?

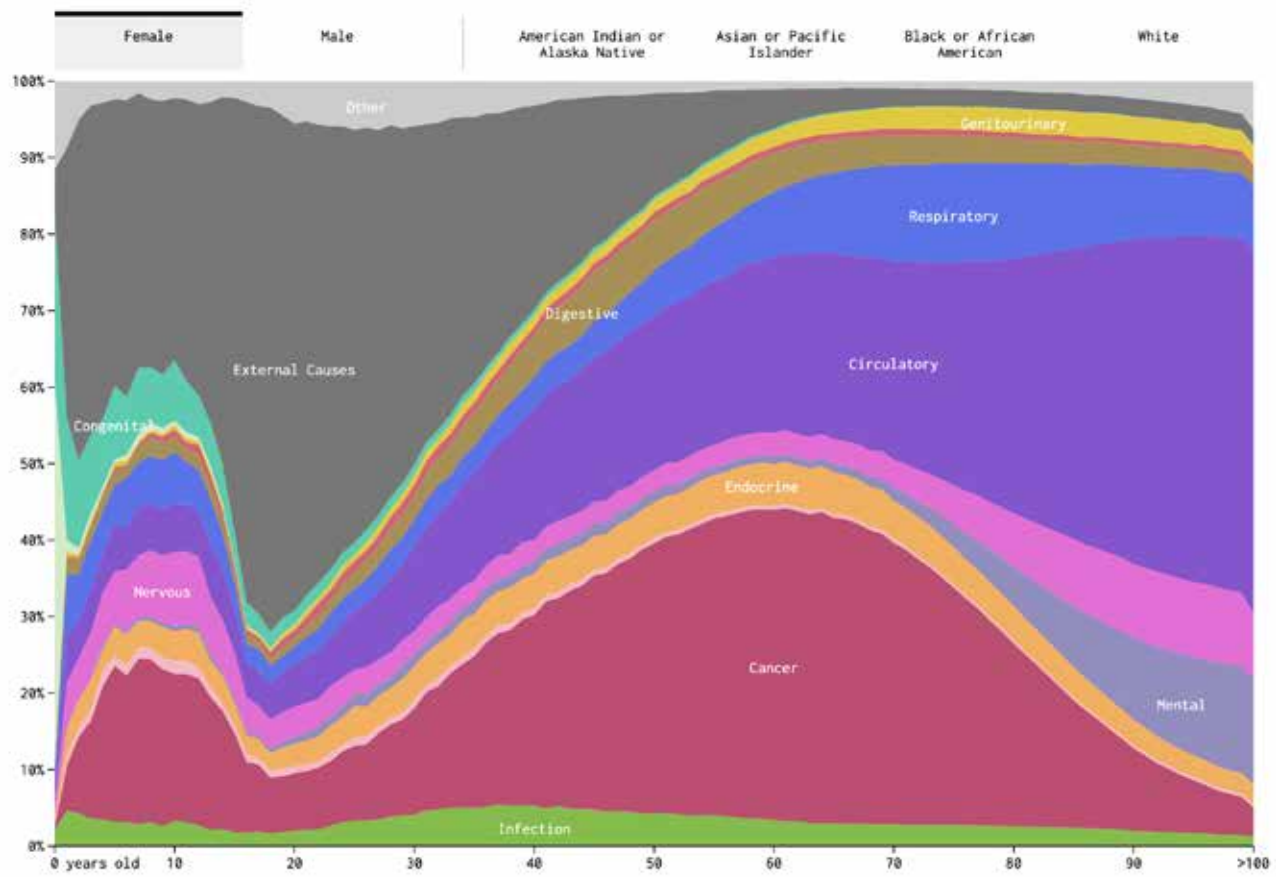
When I wrote the first edition of *Visualize This*, people approached visualization from a narrower point of view. The focus was on analysis, which meant charts were always optimized for perceptual efficiency. With a more flexible approach aimed at data communication, visualization has branched out to games, art, entertainment, and social media, which can mix with statistics, data science, and information design.

The tools are also different. For example, we used primarily Flash to animate data and make it interactive on the web a decade ago, but Flash was deprecated, and now there's a wide array of JavaScript libraries and web frameworks that work better on various devices. *Visualize This* covers the mix of tools so you can decide what works best for you.

Can you share an example of a successful data visualization project you have worked on? In your opinion, what do you think contributed to its success?

The goal with most of my projects (*https://flowingdata.com/category/projects*) is to help people relate to a data set, which encourages further exploration and a better understanding of the data overall.

For example, I made a series on mortality (*https://flowingdata.com/mortality*) that lets you enter demographic information, and the charts show or simulate data based on what you enter.



This chart shows how cause of death varies across sex and race, based on mortality data from 2005–2014. Visit the FlowingData website at *https://flowingdata.com/2016/01/05/causes-of-death* to select a group and see the changes. Select causes to see them individually.

The project worked well because people could see themselves more easily than if I were to just show life expectancy curves. The animations also help show how the data plays out and how a distribution builds.

Why is knowing how to visualize data so important?

It can be a challenge to explain data to a general audience, but people are often better equipped and more willing to read charts. So when you want to communicate data and findings to people who don't work with data regularly, visualization is the best way to do it. When designed correctly, you can potentially expand your reach.

With all the data visualization tools and platforms out there, how do you know which tools to use? Are there any you recommend for beginners?

I recommend people work with what they know until it doesn't do what they want anymore. That could be Excel. That could be R. That could be JavaScript. Figure out what you want to make, and then find the tools that help you do it.

My own toolset centers around R for analysis and static graphics, JavaScript for interaction and animation, and Adobe Illustrator for editing.

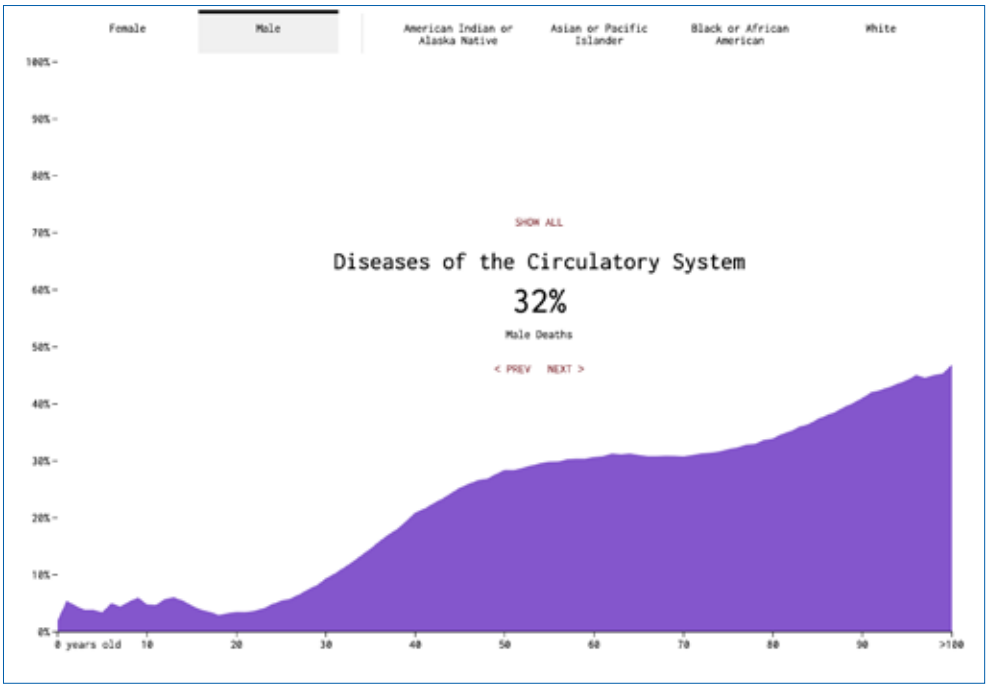
How do you stay up to date on emerging trends and best practices in data visualization?

My favorite part of FlowingData work is analyzing and visualizing data sets myself, which forces me to learn new tools and approaches for visualizing data. Also, like last time, I subscribe to feeds, sites, and newsletters.

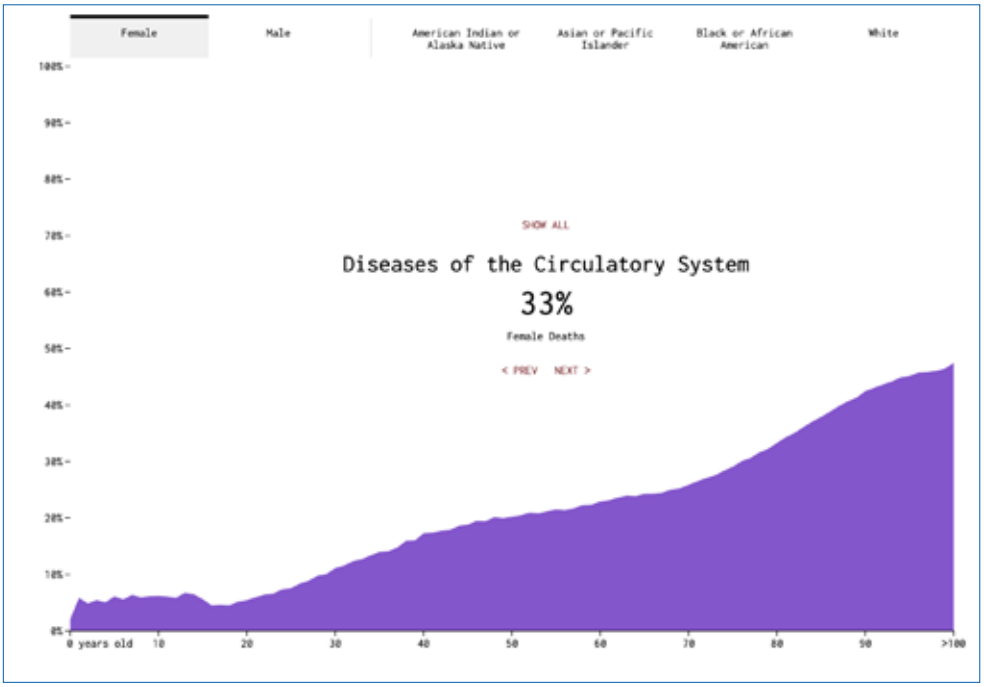
What are rules a statistician should follow to make their charts better? What are common mistakes to avoid?

I talk about chart rules sometimes (*https://flowingdata.com/tag/rules*), and it's usually in the context of treating rules more

Visit the FlowingData website at *https://flowingdata.com/2016/01/05/causes-of-death* to select a group and see the changes.



This chart shows the percentage of males who died from diseases of the circulatory system. The chart is based on mortality data from 2005–2014.



This chart shows the percentage of females who died from diseases of the circulatory system. The chart is based on mortality data from 2005–2014.

Choose visualization methods that help readers relate.

like suggestions. My main suggestion for statisticians who want to make great charts is to treat visualization as a way to explain what your data is about. Don't assume that if you make a chart the data instantly shows something interesting. Annotate and highlight. Choose visualization methods that help readers relate.

How do you see data visualization changing in the next five years?

The process of making charts will likely grow easier—maybe with AI-based tools and maybe with more traditional point-and-click tools. Hopefully with less time spent on implementation, there will be more time for analysis, clarity, and fun with data.

Is there a new skill you are currently learning?

I've grown more curious recently about how we can use the playfulness of games to encourage data exploration and understanding. So I'm leaning harder into animation and interaction.

What do you do when you are not charting data?

I cook a lot. I get a lot of satisfaction out of grabbing raw ingredients from the refrigerator and turning them into a tasty meal for the family. ■

Data Art: Where Statistics Meet Creativity

Kim Gilliam, ASA Marketing Project Manager

Gregory J. Matthews

Gregory J. Matthews, statistics professor and director of the Loyola University Chicago Center for Data Science and Consulting, knows he is an unlikely creative. Over the last 15 years, his path has led him to the intersection of mathematical precision and artistic expression, where he is thriving.

Matthews admits he was a self-described art skeptic when he met his wife, who holds a BFA, almost 20 years ago. She has helped shape his perspective over time and encouraged his interest in experimenting with data art, which began shortly before their move to the Windy City in 2014. Calling Chicago home gave the couple more access to world-class art museums, galleries, art fairs, and street art, all of which continue to spur Matthews' growth as an artist himself.

"Today, I view art very broadly," says Matthews. "Simply, art is created with intention, and that intention is to invoke a reaction from the viewer.

"For example, consider a stop sign on the street—it's purely functional. But place that same sign in a museum and label it as art, and its context changes. It becomes art through intention. You don't have to like it or consider it good art, but someone deliberately put it there. The key is to try to understand their reasoning. You can then decide if it's garbage or not, but the act of consideration is what matters."

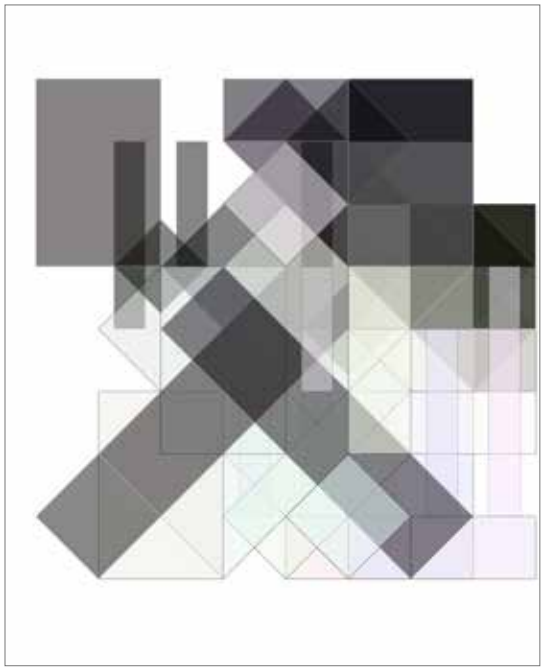
Data Art and Data Visualization—What's the Difference?

Data visualization and data art are both graphical presentations of data, but they have different objectives, or rather different intentions. Matthews values and enjoys both, distilling his philosophical distinction between the two disciplines by explaining, "Data art provokes a question. Data viz answers a question." He adds, "It sounds super corny, but this sums it up perfectly. Put it on my gravestone!"

His curiosity guides both his academic pursuits and his interest in art; it was only natural his love for statistics and the need to visualize data would converge with his desire to ask questions rather than always providing answers. Matthews has often drawn on personal information for his data art, creating abstract images from personal Google locations and Fitbit health data, turning his daily movements into colorful geometric patterns.

Beyond being visually engaging, those particular pieces are also meant to highlight a point about data privacy and the role of technology in our lives. Matthews says, "While we may monitor our personal information, have we considered the extent to which major tech companies like Google and Fitbit are amassing vast quantities of our sensitive data? What are our thoughts on this practice? Is it something we should accept?"

Clown (2019)



Chess Immortal Game (2017)

In recent years, Matthews' work has expanded to include Google Images search data sets, and he creates art using the resulting images as input for statistical models. For example, when Matthews set out to create a new piece of data art in 2019, he searched the word "clown" on Google Images and selected the top results to create a data set to train a statistical model.

"So, imagine I'm building a model where the target is the color of a pixel, and all my predictors are every other color pixel in the image," says Matthews. "One could easily build a model that will accurately predict the pixel color very well. But I find it more interesting to use 'bad' statistical models (ones that underfit the data)—rather than simply averaging these images—to produce more visually intriguing results, with greater contrast and appeal than simple pixel averaging."

Matthews has also tackled broader social issues through his art using US Census data to create a series of black-and-white visualizations of race and income data based on ZIP codes to reveal socioeconomic patterns.

Showcasing Data Art

Matthews is the mastermind behind the JSM Data Art Show, which debuted at JSM 2016 in Chicago. His mission is to showcase data-inspired artwork within the statistical community. Although it began as a no-frills affair in an expo center, he wants to create a more gallery-like atmosphere with proper displays and an opening reception. Ambiance aside, enthusiasm was high for the art show for several years—until COVID-19 forced JSM online.

As we return to gathering together once again, Matthews remains committed to the show and encouraging more statisticians to create and submit data art. He emphasizes that the barrier to entry for creating data art is lower than many might think, urging people to start creating without worrying about expertise or even initial outcomes. "You want to be an artist? Just say you're an artist and start making art—and sign up for the JSM Data Art Show!"

To revive the show, Matthews proposes a multi-faceted approach:

- Begin promotion efforts earlier, starting in October.
- Broaden participation by emphasizing the inclusive nature of the event.
- Recruit volunteers for organizational and promotional tasks.
- Maintain a consistent effort to build year-over-year momentum.

With his eye on JSM 2025 in Nashville, he may get his dedicated JSM gallery yet.

Uncontrolled Variables

Among his diverse portfolio of artistic outlets, Matthews produces and performs in a monthly show—a unique mix of science and comedy called "Uncontrolled Variables"—at the Lincoln Lodge in Chicago.

"So, when someone moves to Chicago, they're expected to take an improv class. It's a rule in this city—so I took a year of improv classes at Second City and met a bunch of people," says Matthews. "When COVID hit, I started doing online open mics and then in-person mics

when restrictions were lifted. I basically stumbled upon the folks producing 'Uncontrolled Variables' and wanted to know how I could help with the show. They said, 'We need scientists,' and I was in."

"'Uncontrolled Variables' has a formula that works. The show is incredible, and hilarity ensues," says Matthews. The format pairs scientists with comedians: Comedians perform short stand-up sets and then attempt to present scientific slides they've never seen before. Next, the scientist who prepared the slides is invited to re-present the slides, explaining their actual research. Matthews contributes by delivering comedic "guest lectures" featuring statistical analyses related to the show's scientific theme.

"I can't believe I run a monthly science and comedy show in a major American city at an actual comedy theater," says Matthews. "And it's not just my mom or some friends in the audience—it's often 70 to 100 who show up through word of mouth. It's an incredible thing."

Statsinthewild.com

During graduate school, long before any thoughts about making art or getting on stage, Matthews started blogging about statistics and its applications in the real world. Today, he maintains the website—called Stats in the Wild (*https://statsinthewild.com*)—which houses his blog, his art, a library of his comedy YouTube videos, and publications.

For those interested in any type of niche topic, he recommends starting a blog or website, which he says can be an excellent way to showcase your work and connect with like-minded individuals.

"Blogging hones your writing, too, which is a really important skill that isn't taught at all in statistics. As a STEM PhD student, you are immersed in math and science, but I believe it's often assumed you can write. That's not the case, and writing is hard."

Beyond His Artistic Endeavors

Matthews is deeply involved in academic and community work. As the director of the Loyola Center for Data Science and Consulting, he focuses on applying data science to assist those with analytics needs both internally and externally, bridging the gap between statistical expertise and community needs. "I've wanted to lead a consulting center from day one," admits Matthews, "but it took quite a while to make it a reality.

"Nonprofits with limited resources are where we can make a difference by helping them with their research or data science needs. We've worked with farmers markets who have some data, but they're making pie charts in Excel—they need our help," says Matthews. "We've assisted the Cook County Community Survey, which is run by several professors at Loyola University Chicago, by developing a dashboard to disseminate their survey results to the public. Getting that off the ground was a big win for us this year."

The center is also collaborating with Beat the Streets, a nonprofit organization that uses wrestling as a tool to engage kids in underserved communities and support them athletically and academically. "We're able to help them with their data analytics needs like data cleaning, data analysis, and data visualization for reporting purposes," says Matthews.

"Our center-affiliated students get to work on these projects and develop real-world experience while still in school—which is great for everyone involved! We're slowly trying to build up the center and increase its presence in the Chicago area. We've applied for grants and are raising money to pay students and faculty for their work so we can expand capacity, and it's been awesome."

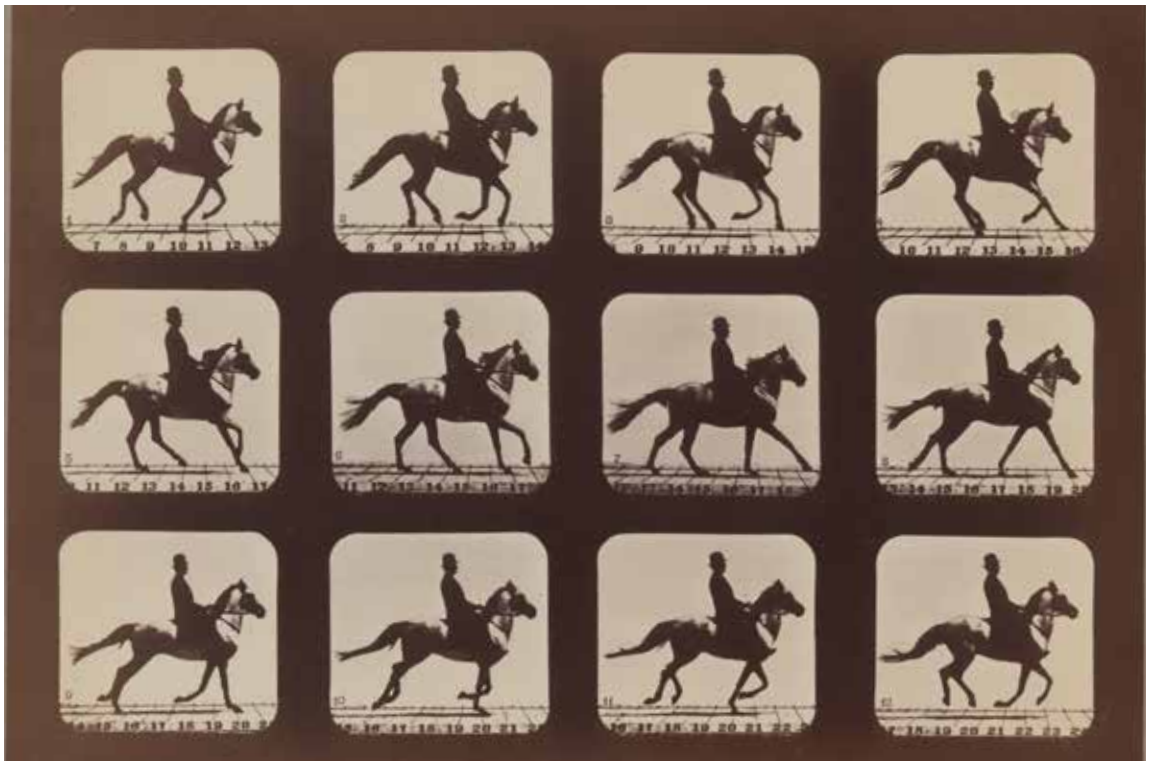
> You want to be an artist? Just say you're an artist and start making art—and sign up for the JSM Data Art Show!

Dream-Come-True Scenarios

Matthews is a husband, father, professor, statistician, artist, Kaggle champion, Big Data Bowl finalist, "openWAR" author, comedian, and improviser. If it sounds like this collection of unlikely intersections makes for a full life, it's safe to say Matthews would agree, but he wouldn't want it any other way. He says, "I'm fully aware that I'm living my dream life." ■

Data Exploration Through Time

Catherine Durso, University of Denver



Eadweard Muybridge, 1881 print of successive positions of galloping horse (public domain: *https://public.work*)

Catherine Durso is a consulting statistician for research at the University of Denver and a teaching professor in the computer science department. She can happily putter for hours on a data visualization.

Technological advances make time increasingly available as a dimension for data visualization. Historic developments in visualization, both animated and static, give a framework for thinking about current data animation. As described in *A History of Data Visualization and Graphic Communication* by Michael Friendly and Howard Wainer, the technology of visualization of time and space has evolved from sequences of still photographs through readily available tools for interactive data animation.

The photographs of successive positions of a galloping horse by Eadweard Muybridge from the 1870s correspond to arrays of two-dimensional data visualizations ordered by another variable, such as time. Muybridge used a device of his own invention to display such images in rapid succession, giving the viewer the illusion of motion.

Motion pictures use time directly to represent a scene over time. Computer graphics enable the use of playback time to represent time in a data visualization. This gives the visualization designer access to an additional dimension and the viewer's ability to detect motion and change over time. The frames can draw on the full range of techniques for static visualization.

Thinking of motion pictures as data representation raises some of the complexities that

View percent population change from 1991–2010.

Basic

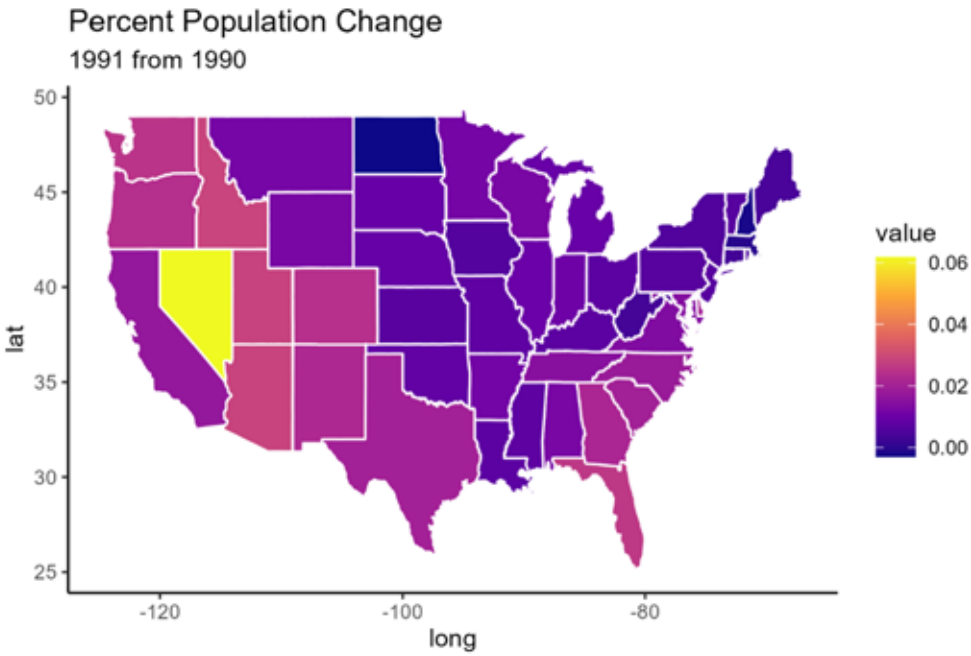


Figure 1: A choropleth of the USA showing percent population change from 1990–1991 by state

persist in computer graphics. The key feature from motion pictures that carries over to data animation is that the playback of the multiple images captured at different time points gives the viewer direct experience of the time dimension of the scene. The hue and intensity at space locations in a single frame of a multi-frame motion picture represent the projection of the value of light from a scene into two dimensions. Details of this relationship would involve a dive into the study of optics. Frame rate, shutter speed, ISO, and playback speed will also affect the information conveyed by viewing the motion picture.

Similarly, a data animation must address the assignment of data to separate frames, the representation of the data within each frame, and the transition from frame to frame. Periodicity in the scene or data may interact informatively or confusingly with the choice of frames, as in the changes in apparent rotation of wagon wheels in movies as the wagon changes speed, changing the rotation time of the wheels. In data animation, one might choose frames to emphasize or suppress a shorter-term cycle in favor of a longer one.

Moving into more abstract visualizations, the role of a time dimension is particularly clear for data sets having multiple cases each with two space coordinates (or projections of three space variables or fully three-dimensional spatial data with access to virtual reality technology) and a time coordinate. By analogy with motion pictures, each frame represents the spatial information for a specific period. As the frames are displayed in succession, the time at which a frame is displayed represents the time coordinate for the cases in that period.

Animating a time sequence of choropleths for the same region and coloring principle illustrates this approach. Figure 1 shows a map of the United States with states colored by their annual percent population gain between 1990 and 1991. A map animation of such maps for 1990–2010 gives an overview of population trends over time, showing periods and regions of fastest growth, for example in the basic choropleth animation at *https://cs.du.edu/~cdurso/data_animation/prev.diffs_1990_2010_amstat.gif*. The animation has other lessons. The colors don't vary smoothly. One abrupt transition is eloquent; note the sharp drop in population in Louisiana in 2006, the year following Hurricane Katrina.

Overall, though, visual interpretability might be improved by mathematically smoothing or interpolating data to provide intermediate frames as in the smoothed choropleth animation at *https://cs.du.edu/~cdurso/data_animation/prev.diffs_1990_2010_eased.gif*.

In general purpose animation, the creation of frames to smooth transitions between essential frames is called betweening or tweening. Easing—the degree of difference between the successive tweens as the transition progresses—gives the designer

View percent population change from 1991–2010.

Smoothed

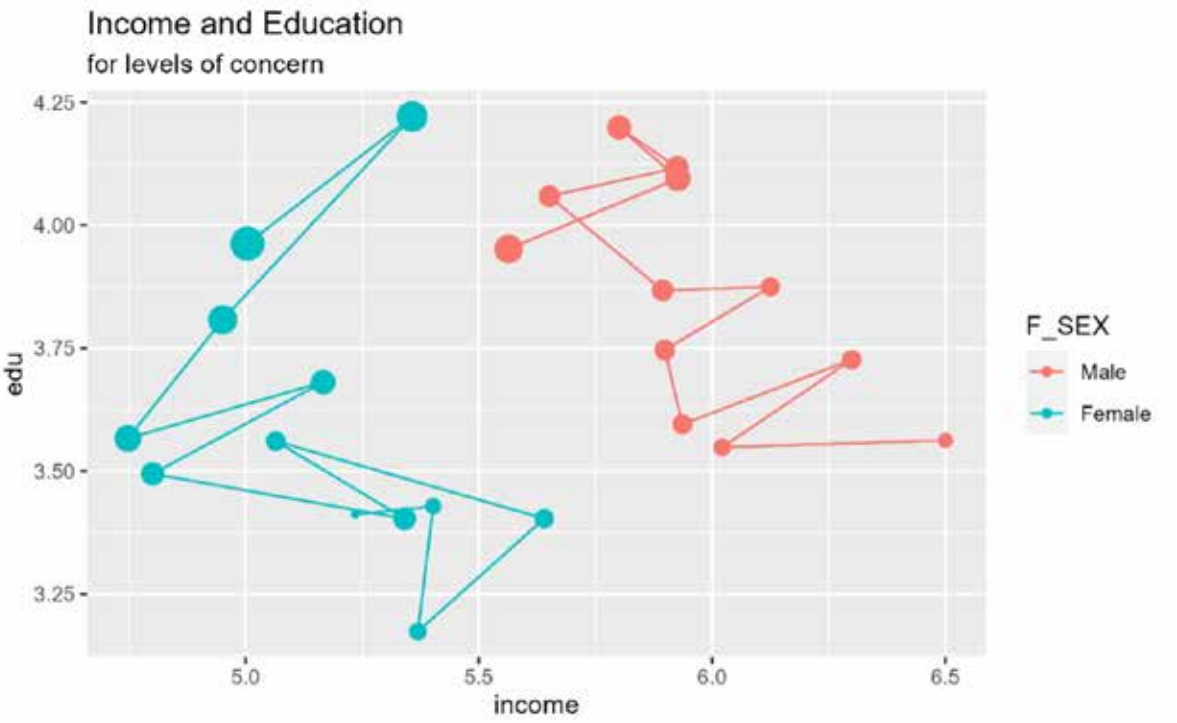


Figure 2: The final frame of the income, education, and priority animation

population pyramid over time

voting agreement

scatter plot animation

control over the apparent speed of the transition close to the beginning, in the middle of the transition, and toward the end.

A data animation can use pauses on frames directly drawn from the data to distinguish data values from interpolated values similarly to the way data points on a smoothed curve distinguish observation from interpolations.

Static data visualizations take a leap in flexibility with the representation of nonspatial variables using spatial coordinates. From there, using animation to represent data with a time coordinate and two or more other variables proceeds naturally. Each frame consists of a traditional planar representation of the non-time variables, while the time of display represents the time variable.

For example, a population pyramid can use space coordinates to represent population counts in age cohorts. For data with binary gender labels, back-to-back bar charts can represent counts in cohorts by gender category. Animating these visualizations with a frame for each year creates a population pyramid over time (*https://cs.du.edu/~cdurso/data_animation/full_1_year.gif*), giving a sense of the changes in population size and age distribution.

When an entity in the data is in different positions in different frames, the sense of an object in motion depends on mental connection of multiple images representing the same object in different positions. Continuity in the position may or may not be sufficient to allow the user to track the entity as the animation progresses. The designer may have to provide cues to identity.

In the population pyramid, a cohort of individuals—with additions and deletions—moves up the pyramid as the years progress. To facilitate tracking of a cohort, the hue of a bar acts as an identifier for the birth year, with repetition after 10 years.

For animations in which the construction of each frame includes a random component, the designer must take particular care to retain visual consistency. In the animation of voting agreement (*https://cs.du.edu/~cdurso/data_animation/WellmadeShadowyBergerpicard.mp4*), the graphs show the degree of agreement in voting between members of the US House of Representatives from 1949 to 2012. The vertices for each period correspond to members of the House and positioned using a linear-attraction, linear-repulsion model.

Proceeding by analogy from the visualization of nonspatial data with spatial coordinates, animation allows for the display of a nontemporal variable along a time coordinate. For example, the scatter plot animation at *https://cs.du.edu/~cdurso/data_animation/norm_dots.gif* displays successive cross-sections in the v3 direction of a simulated data set with variables v1, v2, and v3. The frames are scatterplots in the v1 and v2 values. Thinking about this carefully raises the question of time slices. If the v3

values come from a continuous variable, just displaying the scatterplots for each occurring value of v3 may result in sparse plots. In the case in which no v3 value is repeated exactly, each frame will contain a single point. The designer must consider how wide a v3 interval should contribute to the plot in each frame (shutter speed), whether position within this interval should be represented in the visualization (ISO, how fast a stimulus produces an effect of what size in a camera), and the extent to which past data should persist and future data should be prefigured.

Here, instead of thresholding the width of v3 displayed in each frame, older points are distinguished from newer points by color. Points close to the center time are more opaque than points from the future and past, but points persist throughout the animation. The viewer sees the v1 and v2 cloud move generally up and to the right as v3 increases.

Aesthetics used to indicate position in time should be distinct from aesthetics used to convey other variables. For example, if each case in the data represents an individual's score on four behavior scales and *x* position, *y* position, and color of the point represent an individual's score on three of the scales while time is used for the fourth, the color of the point shouldn't be used to indicate its position relative to the center time.

In movies and in data animation, the designer has flexibility in the relationship of the true time or time-like value to the playback time. Slow motion photography and time lapse photography illustrate this for motion pictures. The collector of the data and the data analyst will also make choices about the length of the interval between observations. In data collection, the question of the intervals at which to sample the data (e.g., annual observations of demographics or millimeter slices in medical imaging) corresponds to the frequency with which separate frames are captured with a camera. The designer may introduce a nonlinear relationship between playback time and the variable, call it *t*, represented by time. For example, intervals in *t* in which the remaining variables change rapidly with respect to *t* may be traversed slowly, while regions with little change may be traversed more rapidly, corresponding to slow motion and time-lapse movies.

Retention of data from smaller values of *t* can give a sense of the trajectory of the phenomenon being studied. Several aspects of poll data inform the animation of opinion and demographic data at *https://cs.du.edu/~cdurso/data_animation/pew_seg.gif*. The time variable represents the mean degree of priority given by the respondent to issues of high priority to self-identified "very liberal" respondents. The x-axis represents education level. The y-axis represents income category. The dots are positioned at the averages for respondents with each level of mean priority. The radii of the dots represent the proportion of the population estimated to have that priority value. The responses are separated into binary gender categories, indicated by color and segment connection. While the end point of the animation shows the values for all mean priority values, the animation allows the viewer to associate the values of income and education in each category at each level of mean priority without having to trace the sequence manually.

Departing from the movie paradigm and treating the time variable more symmetrically with other aesthetics suggests other features a data animation can use. Interaction allows the designer to leverage the user's sense of the control mechanism to increase the intuition that can be gained from the visualization. User control of playback speed and time gives opportunities for interaction comparable to controlling the scale and interval on a spatial axis.

Interaction allows the designer to leverage the user's sense of the control mechanism to increase the intuition that can be gained from the visualization.

For example, for two-dimensional projections of three-dimensional scatter plots, giving the user the ability to physically manipulate the choice of the projection axis in real time enhances the user's sense of the situation in three dimensions. As the user does this, the user essentially creates an animation. Allowing the user to run the animation forward and backward at different speeds provides the time axis analog of the ability to focus on a particular region in a static visualization.

These descriptions are not exhaustive, and human ingenuity abounds. I trust current and future data animators will use innovative methods to expand our ability to gain intuition from data. ■

demographic data

MORE ONLINE
Sources for the data used in animations generated by the author are shown in the index of animations at *https://cs.du.edu/~cdurso/data_animation/data_animation.html*.

Designing Spatial Experimental Plots for Comprehension

Alison Kleffner, Creighton University

Alison Kleffner is an assistant professor at Creighton University. Her research focuses on spatio-temporal modeling and visualization in environmental applications.

It is practical in many applications to visualize data in a way that maintains its spatial context. However, maintaining spatial context while visualizing the relationship between two variables is difficult, especially when the variables occupy the same spatial domain. Certain design choices increase the difficulty of interpreting these visualizations, potentially leading to longer graph comprehension times and an inability to correctly discover patterns.

With a projected increase in future crop demand, researchers have been conducting experiments (called on-farm precision experiments) to examine the effect of crop input application on yield to inform more sustainable farming practices. During these studies, researchers collect multiple pieces of data on the same field and store them in a shape file—a common file type for geospatial data. These pieces of information include the following:

- **Experimental Design:** Usually a Latin-square-based design, where the shape file contains the target crop input application (treatment) rate and locations on a farmer's field where they should be applied.

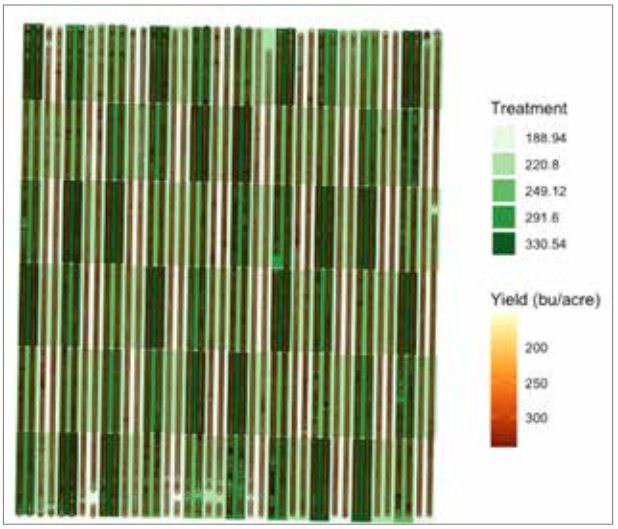


Figure 1: Original superimposed graph where the yield measurements (circles) lay on top of the experimental design (rectangles). This plot has issues with clutter due to the overlapping yield measurements.

- **As-Applied:** Treatment rates a tractor applied to the field and application location. These rates differ from the experimental design due to sensor errors and machine limits of planters.
- **Yield:** After the growing season, obtain crop yield measurements and their measurement location.

The data used in the visualizations in this article is a small example of the kind that would be collected and analyzed in one of these agricultural field experiments.

In this application, it is beneficial to visually explore the collected data to see how yield responds to different treatments across the field. Hence, when displaying this data, it is important to maintain spatial context, as yield can vary spatially due to variables such as soil content.

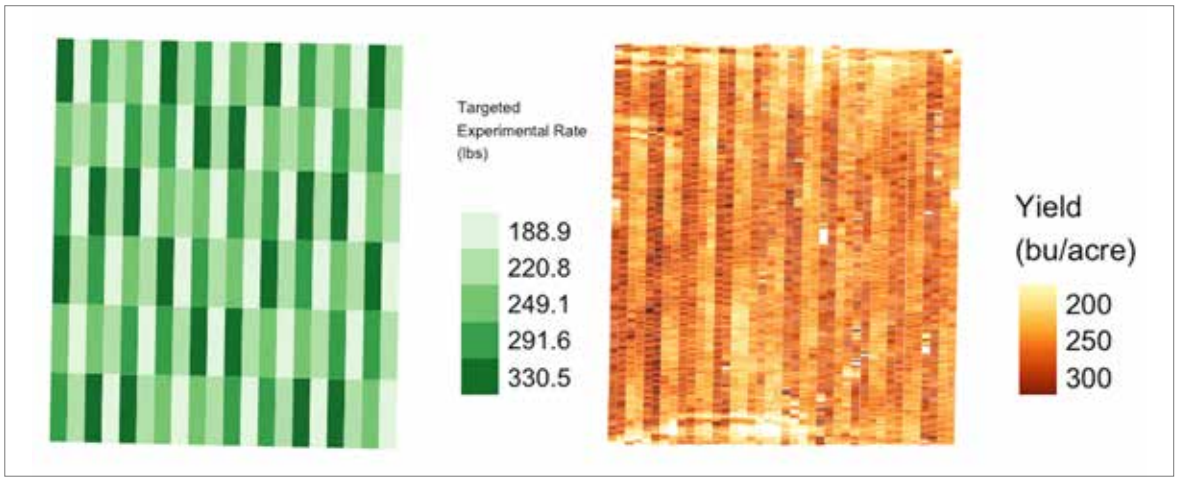


Figure 2: Original juxtaposed graph where the experimental design and yield measurements are placed side-by-side. In this visualization, the comparative burden is placed solely on the user, adding to the difficulty in interpretation.

Current Sub-Optimal Visualizations

A common method to visualize two variables in the same space is a superimposed comparative layout. Figure 1 is a re-creation of a superimposed graph typically used to display the relationship between treatment and yield in agricultural literature. Here, the yield map is superimposed on top of the experimental design. The colored rectangles represent the experimental design, with the different hues denoting the planned treatment rate. The overlaid yield information includes circles representing the measurement location and yield amount through its color.

One perceptual concern in Figure 1 relates to the yield circles. The circles overlap, obstructing the visual cue of color and adding difficulty to determining the number of yield measurements. Because of the color obstruction, users may incorrectly evaluate the relationship between the variables through space. Issues stemming from clutter are common drawbacks of superimposed layouts.

Figure 2 is another representation of the same data using juxtaposition. A juxtaposed layout places the experimental design and yield data side-by-side. In contrast to the superimposed layout, this visualization does not have issues with clutter. However, juxtaposition places most of the comparative burden on the user's memory. Due to the user's eyes shifting focus between images, they rely on a mental image for comparison. Unfortunately, the mental image might not form accurately in their working memory. Furthermore, users must identify corresponding regions in both plots and assess their correlation. These are demanding tasks that can result in incorrectly extracting the variables' relationship.

A common concern in Figures 1 and 2 is the color schemes. The chosen color scheme for this application uses green gradients for an experimental trial and a yellow-orange-red gradient for yield. Red-green color blindness is experienced by about 8% of men and 0.5% of women with Northern European ancestry, according to the *Nature Methods* articled titled "Points of View: Color Blindness" by Bang Wong. Those affected have difficulty discriminating between these colors and other colors containing a component of these colors, leading to potential errors when deriving relationships. So, when creating a visualization, choose colors more strategically to make them accessible to larger groups of users.

Suggestions for Improvement

We began with a superimposed layout, so users can use their perceptual system rather than relying on their memory. This layout is also generally recommended when data occupies the same space, as the user does not have to identify the corresponding regions for comparison.

To address the overlapping circles for yield, we transformed the circles into non-overlapping polygons using the distance between points, swath width, and harvester direction. The individual polygons were too small for users to extract information efficiently, so

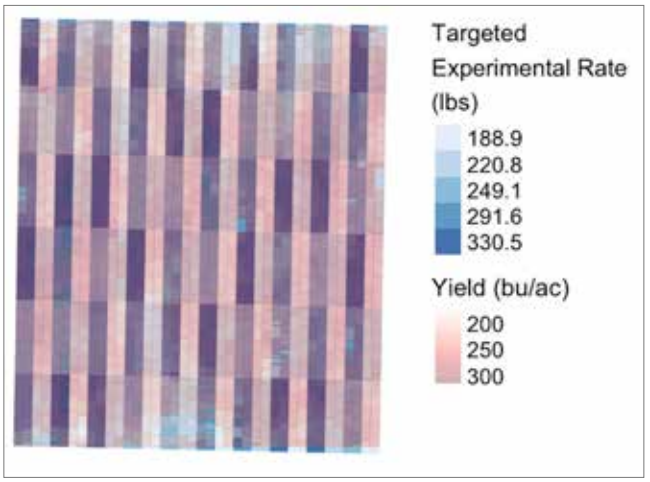


Figure 3: The first step in the redesign process maintained the superimposed layout but introduced non-overlapping polygons and transparency for color blending.

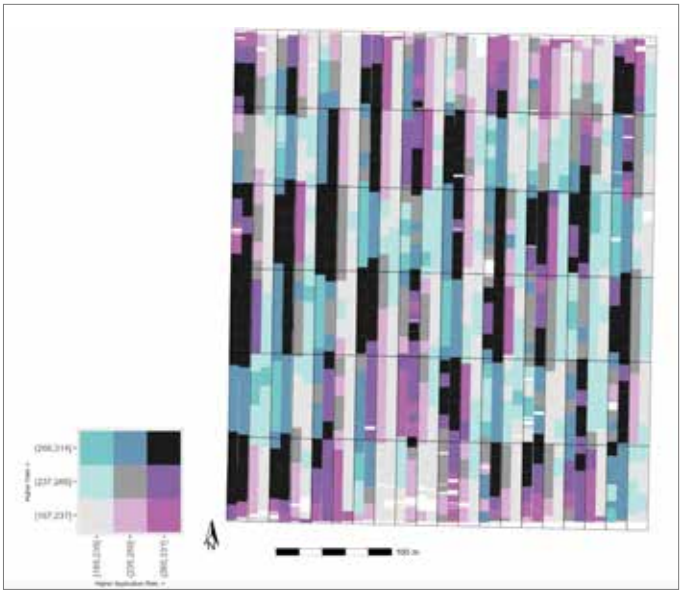


Figure 4: The bivariate color map is an alternative to color blending, where each variable is separated into categories to create a color scale.

Editor's Note: This work was completed under the direction of Susan VanderPlas.

we combined them into larger polygons, calculating the average yield within each combined polygon. The yield information no longer overlaps, so its color is no longer obscured. We blended the colors of the new yield map with the experimental design using transparency. Hence, users assess the relationship through the blending of colors for yield and treatment.

We also addressed the problematic color scheme by maintaining the red gradient for yield but using a blue gradient for treatment. An example of this plot is found in Figure 3.

While Figure 3 addresses perceptual issues, using color blending to derive a relationship through space is difficult. To further address these challenges, we used a common alternative to color-blending called a bivariate color map. This type of visualization can be informative if the relationship between variables is more important than individual values, according to the *Cartographic Perspectives* article titled "Operationalizing Trumbo's Principles of Bivariate Choropleth Map Design" by Georgianna Strode and coauthors.

We used quantiles to separate both variables (treatment rates and yield) into low, medium, and high categories to create a 3x3 color scale. A larger scale hinders interpretation due to the number of colors.

Since the focus is on the direct relationship between variables, the diagonal of the color scale uses a grayscale color scheme. The upper and lower triangles have complementary color schemes to show their separation.

The field is broken into polygons, where its color is determined by the average yield and average treatment found within that polygon. The outline of the trial plots from the experimental design is maintained to provide additional spatial context.

Although the polygons along the diagonal (positive linear relationship) stand out, this visualization is limited due to how we divided the data into only three categories due to working memory limits. Therefore, each category covers a wide range of values.

Additionally, when looking at Figure 4, the yield measurements of 237 bu/acre (low yield/low treatment) and 238 bu/acre (medium yield/low treatment), for example, are represented by different colors while only having a one-unit difference. The different colors may perceptually make the observations seem more distant than they are.

Finally, the user derived the relationship between the variables until now. To reduce their cognitive load, we used an explicit encoding layout, which directly displays the encoding of a relationship between variables. We wanted to directly encode the correlation between the variables in the visualization.

To maintain spatial awareness, we calculated the correlation between the as-applied treatments and yield in different field sections, outlined by black boxes. A diverging color palette was chosen to account for both the magnitude and sign of correlation values. Due to the common cultural associations of red with warmth and blue with cold, we used a red gradient for positive

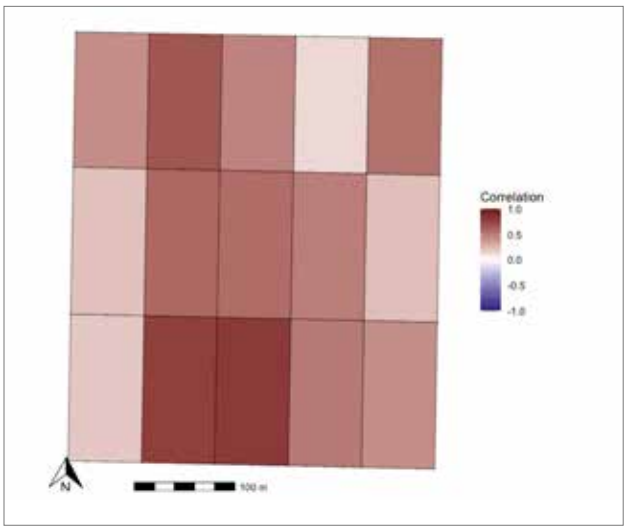


Figure 5: A plot that states the correlation between treatment and yield using an explicit encoding layout.

Further Reading

Information Visualization, *https://bit.ly/3yForNw*

IEEE Xplore, *https://bit.ly/3M67XBf*

Proceedings of the National Academy of Sciences, *https://bit.ly/3SLgZYg*

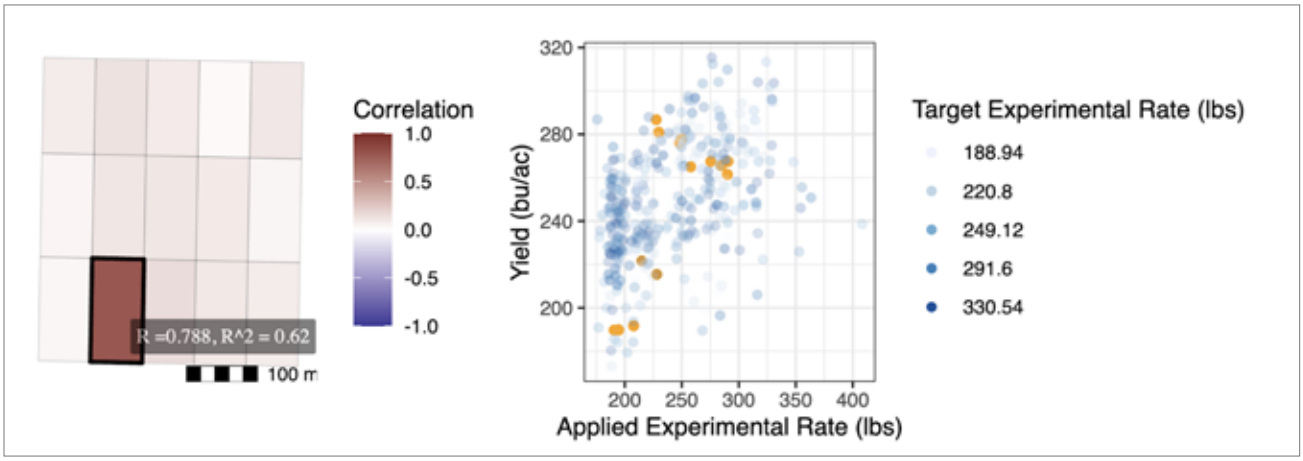


Figure 6: Screenshot of the interactive plot that overcomes the weakness of Figure 5 by including the values used in the correlation calculation.

correlations and a blue gradient for negative. Figure 5 is an example of this correlation plot.

Although Figure 5 directly provides the correlation to the user, it is difficult to connect the correlation back to the data. This is problematic, as a standard principle of data visualization is to show the data. To overcome this weakness, we used a hybrid layout. A scatterplot displaying the as-applied treatment versus the yield was juxtaposed with the previous correlation plot. Interactivity connects the plots, where hovering over a section in the correlation plot highlights the corresponding points used in the scatterplot to calculate its correlation. While interactive plots are not beneficial in static PDF reports, they can be advantageous in online tools designed to help farmers explore their data.

In the Future . . .

Future steps in this project will aim at obtaining user feedback to measure which proposed visualization is more suitable. Part of this will include testing user accuracy in estimating the relationship between variables in the current sub-optimal graphs versus the different iterations of the proposed graphs. Code for the eventual corresponding paper this article is based on can be found at *https://bit.ly/46Pd1Dt.* ■

Code for the eventual corresponding paper this article is based on can be found at *https://bit.ly/46Pd1Dt.*

Understanding How Viewers Interpret and Interact with Statistical Charts

Kiegan Rice, NORC at the University of Chicago

Kiegan Rice is a senior statistician at NORC at the University of Chicago specializing in static and interactive data visualization, reproducible data analysis pipelines, and scientific communication. She applies her skills to multiple research areas, including education, health care, and public health policy. Rice earned her PhD in statistics from Iowa State University, where she studied statistical and computational reproducibility in forensic evidence analysis.

Much of the existing guidance on data visualization design is based on practitioner opinion, and those opinions can be challenging to wade through. Some recommend minimalism, while others recommend catching your audience's eye with flashy colors or a catchy title.

Expert opinions are informed by years of practical experience designing data graphics and sharing them with audiences. However, they do not give us data-backed findings on how different elements of data visualization design affect a general audience's interpretation of the information conveyed, nor do they inform us about how different audiences interact with data graphics in different ways.

There have been many user studies on graphical perception that aim to do just that—provide data-based research on elements of data visualization design. The majority of these studies focus on the *form* of a data visualization and the mapping of values to a scale: Is a bar or pie more effective for faithfully representing the relative relationship between two values? How does changing the range of a Y axis distort a viewer's estimate of the severity of a trend? How small of a difference between two percentages can viewers perceive in a bar chart?

These are important questions to study, but they do not tell us how to *design* a data graphic: How should groups or categories be arranged to facilitate pattern identification? What level of detail in axes, labels, and legends is helpful without overwhelming a viewer? What patterns do viewers focus on when they look at a data graphic? Does viewer behavior and interpretation differ across demographics? The latter is a particularly poignant question to consider, as many prior graphical perception studies use convenience samples, often with highly educated participants.

We Asked US Adults

Nola du Toit, Ed Mulrow, and I collaborated with Heike Hofmann from the University of Nebraska-Lincoln to answer these questions using surveys. Over the course of several years, we fielded survey items to nationally representative samples of around 1,000 adult respondents to study how the US public interacts with and interprets data graphics. We varied the form of the graphics shown, context provided, and questions asked, but we showed the same data set throughout.

We first asked people to assess which element was larger out of two elements with similar magnitude. Aligning the elements in question, providing context like grid lines, and directly identifying the elements of the chart needed to make the comparison increase viewers' ability to assess the relative size of two elements.

When we directly marked the pieces in the chart shown and asked participants to identify which of the two marked pieces was larger, the rate of correctly identifying the larger piece was significantly higher than when we asked participants to, for example, identify whether "the percentage of 80–89-year-old females living alone in a household" or "the percentage of 90–99-year-old females living alone in a household" was larger. From these findings, we know visually emphasizing elements within a chart you want viewers to compare and making sure those elements share a scale and baseline improves viewers' ability to correctly make comparisons.

We also asked participants to directly tell us what they saw, saying "Please tell us in your own words: How would you summarize what this chart shows about the living arrangements of selected older age-sex groups"? Their answers, unsurprisingly, varied greatly.

Some participants picked up on the increase in the percentage of adults living in nursing homes in older groups, with varying interpretations of that pattern:

> *"It shows that the older you get, you are more likely to end up in a nursing home."*

> *"Old people end up in nursing homes eventually."*

Other participants picked up on differences between the male and female groups, responding:

"Elderly women are more likely than elderly men to end up living alone."

"Women are more likely than men to live in nursing homes or other group quarters as they age."

"Women are more likely to be independent in old age."

Some focused on the magnitude of those living in a household:

"[O]ur elderly are either living alone or with family."

"Very few elders live in nursing homes and group quarters. Most live alone or with others."

"Majority of people either live with other in their household then start to live alone."

All the above responses are based on observable patterns in the data graphic shown; however, they each focus on only one pattern or comparison. Respondents, for the most part, identified one key takeaway to share and shared the takeaway that stuck out most to them upon looking at the chart. For some, however, this takeaway or interpretation was unsupported by the chart or they simply shared their opinion of the chart itself (e.g., "It sucks." "Pie chart would be clearer." "Seems inaccurate." "I don't like it."). Some respondents also tied the graphic back to the implications of aging in their own life or related it to personal experiences with themselves or family members.

In addition to the variability in the content of responses, we observed differences in how participants engage with the open-ended interpretation questions. There were noticeable differences in engagement with the chart and content of responses across demographic groups, with higher-education groups spending more time and providing longer responses about what they learned from the chart. We also observed a higher prevalence of pattern-oriented language—words such as "more," "higher," "older" "likely," "than"—in higher-education groups, potentially due to higher degree of familiarity with data visualizations among those with advanced degrees.

What Does This Mean for Designing Data Visualizations?

Our main recommendation so far is as follows: *Make it easy for your audience.* Design your data graphics to visually emphasize the specific comparisons or pattern you want your audience to focus on.

No matter how much we, as data professionals, want to think of data visualizations as displays of objective information, every design choice has an impact on how viewers process information. Therefore, it's important to consider what the most important comparisons or patterns are and design your data visualization to facilitate viewers identifying these key points. This can be done in many ways, including grouping elements according to a high-level pattern, aligning visual elements when viewers should compare their relative sizes, and clearly labeling chart elements.

Moreover, think about who your viewers are; different audiences may experience data graphics in different ways. Every viewer will bring their own experiences, opinions, and perspective when they view and digest the data shown in a chart.

When communicating information to a group less familiar with complex data visualizations, default to simpler structures. Don't crowd graphics with more than one message or finding and try to reduce the possibilities of how a chart can be read or interpreted by providing context and example interpretations and keeping the scope of the chart narrow.

Visualizations are a powerful tool and, when carefully designed to meet your audience where they are, they can provide critical insights about your data. ■

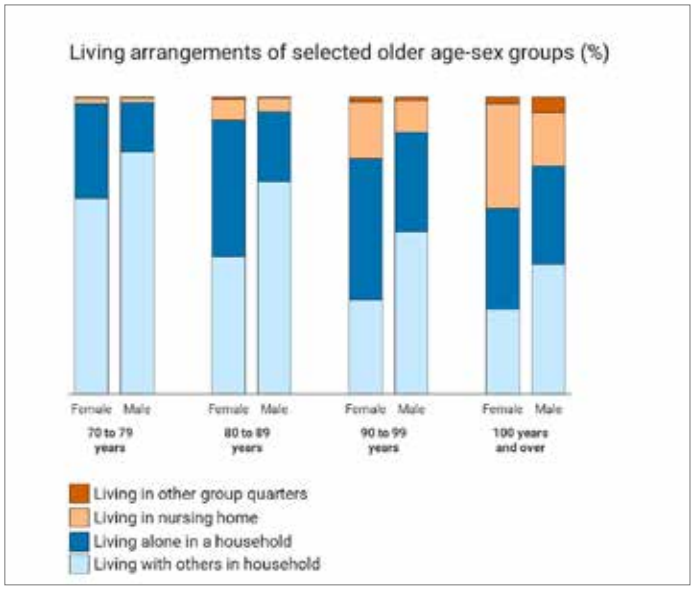


Figure 1: Basic data graphic shown to survey participants. Throughout tests, participants were shown marked and unmarked versions of the chart and versions with and without grid lines. Participants compared elements that were both aligned along the same scale and unaligned.

Deep Dive

For a deeper look at the survey discussed in this article, read "Testing Perceptual Accuracy in a US General Population Survey Using Stacked Bar Charts" in the *Journal of Data Science*. *https://doi.org/10.6339/24-JDS1121*

Save the DATE!

Join us as we work together to **empower tomorrow through statistics and data science.**

YOUR SUPPORT helps showcase the impact of statistics, drive advocacy and outreach, cultivate future leaders, and enhance K–12 education.

Stay tuned for details about how you can get involved, win prizes, and make a difference.

PRACTICAL SIGNIFICANCE

ASA Associate Executive Director Donna LaLonde

THE HOSTS!

ASA Executive Director Ron Wasserstein

Launched in January 2021, *Practical Significance* is the **flagship podcast** of the American Statistical Association and endeavors to inspire listeners with **compelling stories from statistics and data science.** Our guests include **leaders and rising stars** in the statistics and data science profession and ASA community—academics, researchers, industry executives, government heads, early-career professionals, and students.

THE TOPICS

- Statistics and AI

THE NUMBERS

Listeners in

- Real-World Data Science
- Ethical Guidelines for Statisticians
- A 360° View of ASA Journals and Publications
- Statisticians Making a Difference in the Community
- Mentoring Initiatives
- Diversity and Inclusivity in the Profession
- Leadership
- Collaboration with Other Disciplines

25,000+
Direct downloads and streams, averaging 600 downloads per episode

25
countries

15,917
Listens on IOS, the top platform

12,185
Listens on Apple Podcasts, the top app

GET CONNECTED

Subscribe to *Practical Significance*

Apple Podcasts
https://bit.ly/3y4Dy2V

Spotify
https://bit.ly/46nKWTO

Engage on LinkedIn
https://bit.ly/3LBfojw

Read *Practical Significance* Take Two in *Amstat News*
https://bit.ly/46khuhw

Scan me!

THE TOP THREE EPISODES FOR ENGAGEMENT

Episode 1
Making Meaningful Statistics—An Interview with ASA President Rob Santos
https://bit.ly/4cUjMX8

Episode 17
A Roadmap for Change—Recommendations from the ASA Anti-Racism Task Force
https://bit.ly/4dcODxS

Episode 14
The ASA Committee on Statistics & Disability Working to Improve Accessibility
https://bit.ly/4cUk5RM

2024 International Day of Women in Statistics and Data Science

OCTOBER 8, 2024 midnight to midnight, UTC

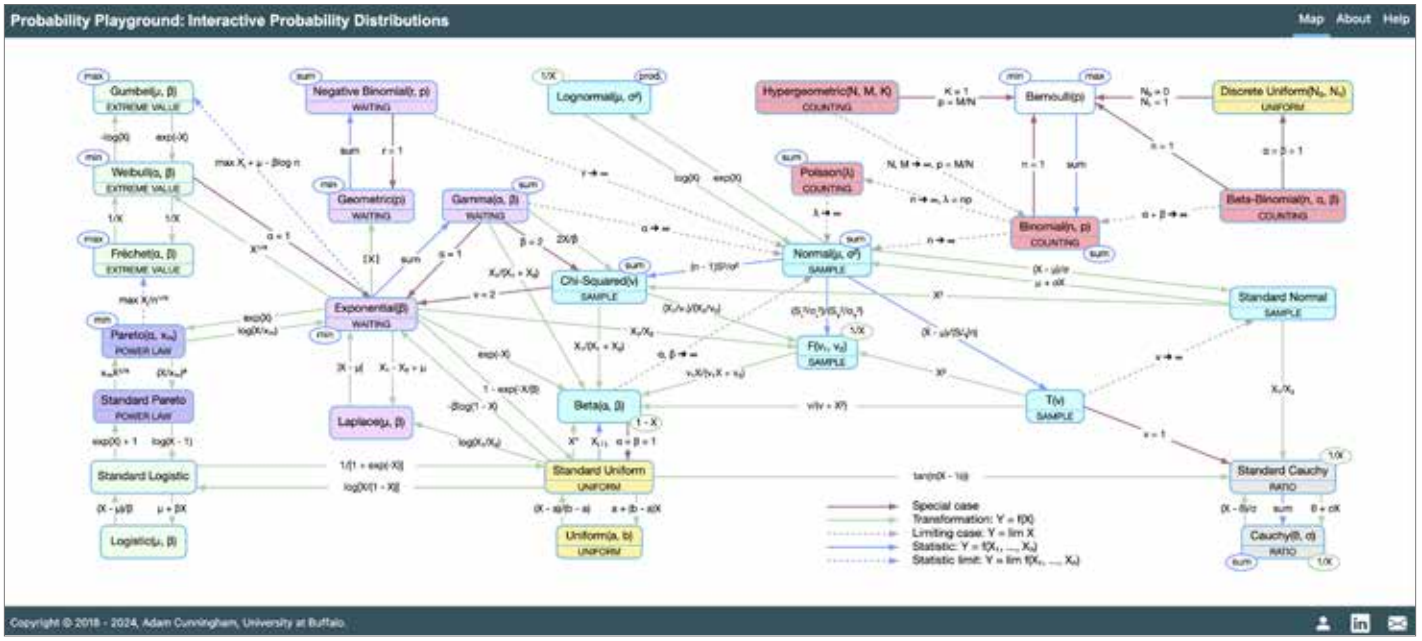
The year's theme is "Empowering the next generation of statisticians and data scientists.'

We will have over 40 sessions, covering a variety of topics: theoretical statistics, applied statistics, networking, professional development and more.

Register now to join us for this world-wide event!

Probability Playground

Adam Cunningham



The map page shows how distributions are related. Selecting a node or link loads its webpage. Distributions are categorized based on the similarity of their data-generating process. Link labels describe the relationships.

Adam Cunningham holds a BA in social and political sciences, an MS in intelligent systems, and an MA in mathematics and biostatistics. Currently an independent researcher, he has worked for the Ford Motor Company, Fujitsu, the University of Buffalo, and UB Concussion Management Clinic.

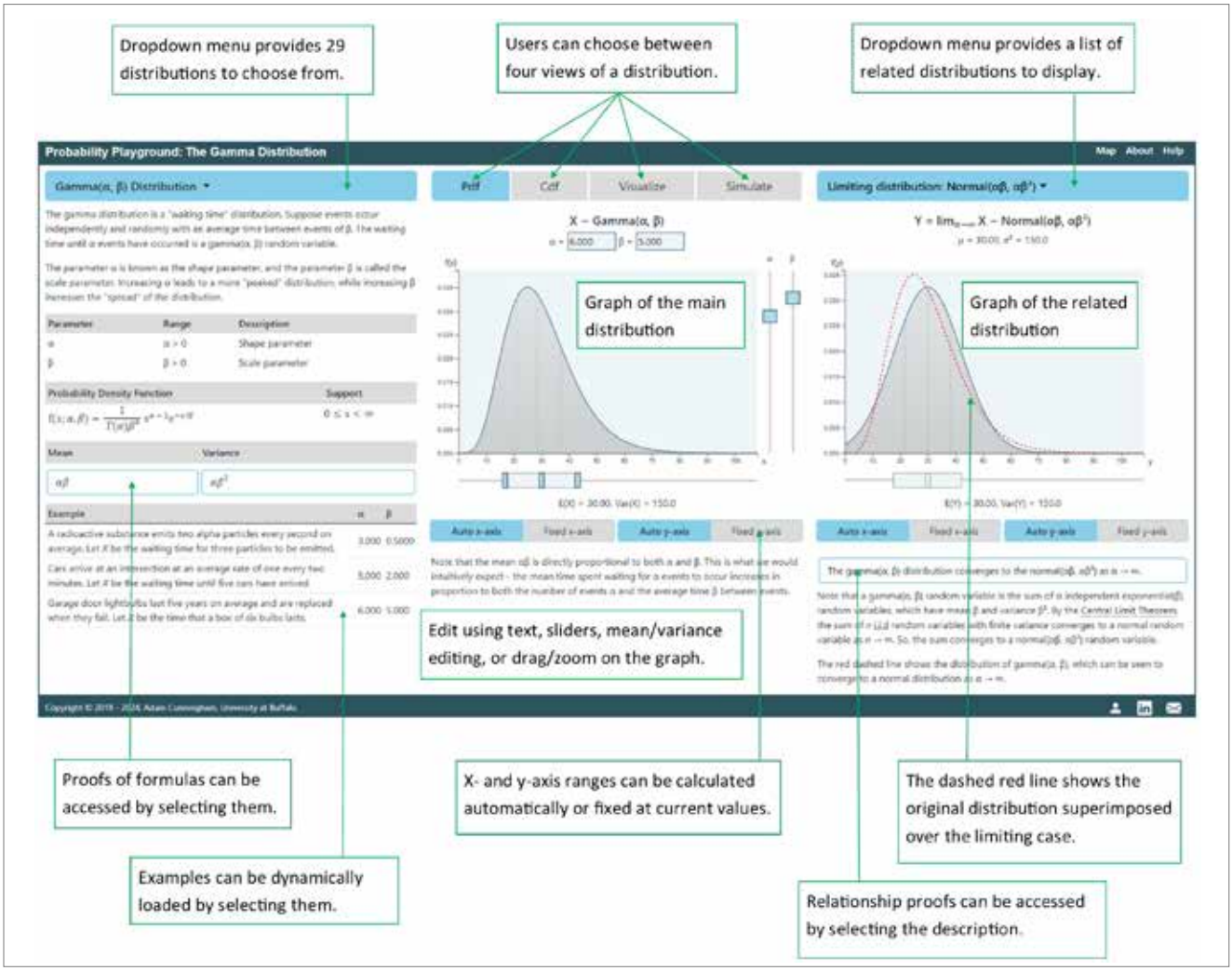
Way back in 2018, I was a graduate student working toward an MA in biostatistics at the University at Buffalo. As with most students studying statistics, the first class introduced several special probability distributions we hadn't seen before. There were a lot of algebra and formulas involved. We spent a lot of time on proofs of means and variances and analyzing how distributions were related. As someone who previously taught math at the University of Buffalo for several years, I could 'do the math' and work with these distributions algebraically by the end of the class. What I didn't yet have was an intuitive feel for many of them.

There are hundreds of websites about probability distributions. However, most are essentially 'books on the web,' with descriptions, examples, and sometimes several graphs on the same set of axes. The majority don't have any kind of interactivity, while the few that do often have limited editing capabilities and don't include relationships between distributions.

I was looking for something different. I wanted to be able to change the distribution parameters and see the density or mass graph dynamically update. If one distribution was a limiting case of another, I wanted to see how they converged. If a distribution arose because of some data-generating process, I wanted to see that process in action. So, as an educator with a background in software engineering and time on my hands, the 'obvious' thing was to build the kind of website I wanted to see. Additionally, the qualifying exams were coming up, and what better way to learn than to teach?

Designing an Interactive Website

Broad principles were used in designing the Probability Playground website (*www.probabilityplayground.com*). For a start, everything had to fit on one page, at least on larger screens. Users don't like to scroll to find information, and it's much easier to take in everything at a single glance. The website also had to be completely dynamic and interactive—whenever any variable changed, the change was immediately propagated to all depen-



Probability Playground provides a wide range of functionality, with several unique features to support interactive exploration.

dent variables, graphs, and animations. The interface had to be intuitive, using standard elements to provide a consistent look and feel that worked across multiple devices. Options for navigation and control had to be immediately visible and accessible using as few user actions as possible. Finally, a single unified framework that encompassed all univariate probability distributions would ensure that, once a user interacted with one distribution, they would know how to interact with them all.

Relationships Are Important

Rather than treating the relationships between distributions as secondary, they play a central role in Probability Playground. From the beginning, I envisaged having graphs for both a main and a related distribution side-by-side, with any changes made to the main distribution propagating to the related one. Building this in from the start meant all kinds of relationships—special cases, transformations, limits—were dealt with within a single framework. Users can see how distributions converge to their limiting cases; how the probability mass function of a special case such as the geometric compares to the more general negative binomial; and how the shapes of sampling distributions such as the chi-squared, F, or T change with sample size.

The map page serves as both a way to navigate the website and a graphical description of how distributions are related. Color is used to categorize distributions based on the similarity of their data-generating process. For example, the binomial, beta-binomial, hypergeometric, and Poisson distributions all involve counting discrete events and are grouped in the "counting" class. This was a later enhancement, with the distributions originally arranged to minimize the number of lines that cross. However, when color was added, the distributions already grouped together spatially under this categorization system, suggesting this is a natural way to classify them. Links are also classified and color-coded as special cases, transformations, limiting cases, statistics (those based on a statistic calculated from a sample), and statistic limits (the limiting case of a statistic).

An Interactive Tool for Exploration

The website was designed as an interactive tool that allows

Visit the playground online at *www.probabilityplayground.com.*

students to explore the 29 most commonly encountered distributions in probability theory and statistics courses. It can also be used as a teaching aid in classes where these distributions are introduced.

Probability Playground provides a wide range of functionality, with several unique features to support interactive exploration. Examples illustrating the range of shapes distributions take can be dynamically loaded by selecting them, resulting in an animated transition to the new parameter values. Parameters can be changed using either the text boxes above the graph or the sliders to the right. This allows students to either graph the pdf or pmf for specific parameter values or see how they change smoothly as parameters vary.

Distribution means and variances can also be independently edited, with an inverse mapping performed back to the parameter values. This can be done using either the slider below the graph (which shows the mean plus and minus one standard deviation) or by dragging and zooming directly on the graph. This is particularly useful when visualizing convergence to limiting cases, as the mean can be independently held constant while the variance is decreased.

The website also provides the option to choose between fixed or automatically calculated axis ranges. These are designed to display nearly all the probability mass on the graph in a visually intuitive fashion across a wide set of parameters.

Formulas for calculating the ranges for distributions with bounded, partially bounded, and unbounded x-axes were developed to ensure they are consistent across relationships.

Four views of each distribution are available: the pdf/pmf graph; cumulative distribution function; a visualization of the data-generating process underlying each distribution; and an animated simulation of this process. Visual representations of processes such as the Bernoulli or Poisson are consistent across the website, enabling the similarities between distributions based on these processes to be seen. Video-style controls allow simulations to be paused, run, reset, or stepped through one update at a time. A histogram below the simulation accumulates results, enabling students to see how the shape of a pdf or pmf arises from the data-generating process.

Last, to add mathematical rigor, more than 150 proofs are included for distribution means, variance, and relationships. These are accessed by selecting the formula or relationship.

Designed with Accessibility in Mind

Several months of work went into ensuring Probability Playground is accessible to users with disabilities. A wide range of improvements were made by following the Web Content Accessibility Guidelines of the World Wide Web Consortium. These ensure as many users as possible can perceive, operate, and understand the website with or without the use of assistive technologies.

To assist users in perceiving the website, larger and more legible sans-serif fonts were used, coded in such a way that font sizes can be controlled through web browsers. Contrast was enhanced for users with reduced visual acuity, and a color-blind friendly palette was used. Mathematical expressions and characters were coded to be accessible to screen readers, and all website elements were labeled to be navigable by users of these technologies.

Operating a website often presents difficulties for users with reduced motor control. Large buttons and icons make finding a target with a mouse easier, while the full range of functionality is also available using a keyboard. Extensive use of accessible rich internet application roles and attributes also makes the names and purpose of all controls explicit, so they can be accessed and controlled using speech recognition software.

A further advantage of designing for accessibility is it makes for a better experience for all users. A website that is well laid out with larger fonts and buttons is more accessible and understandable to everyone. Users with "temporary disabilities" such as injury, lost reading glasses, or bright sunlight can also find themselves needing to perceive or operate a website in a way they normally don't. I found this out the hard way last winter, when I was stuck in Buffalo for two days during a snowstorm without my reading glasses. I found I couldn't increase the font size to read my own website! Needless to say, this was one of the first issues I addressed.

Probability Playground has evolved from its earliest days as a personal project into a robust, extensive, and accessible resource for visually illustrating probability distributions and their relationships. It has been tested for compatibility with all major operating systems, web browsers, and devices, offering an off-the-shelf solution for educators looking to incorporate interactive and exploratory learning into their classrooms. ■

Establishing Your Role in Data Visualization

Robert Grant, BayesCamp and Kingston University

In 2018, my first book, *Data Visualization: Charts, Maps, and Interactive Graphics,* was published in collaboration with the American Statistical Association. The textbook market was already saturated with data visualization books at that point, and it actually took a while for my editor to convince me to write it at all! The aspect that won me over was that the audience included young scientists and students who might be considering a career in statistics. In the years that followed, I found myself continuing that theme of engaging with people who are new to communicating statistical outputs, no matter their age.

Most people in statistics, including me, were taught a lot about calculation and not very much about communication. Yet, as we all know, it is sadly all too common for good analyses to gather dust on the shelf—and fail to influence decision-makers—simply because the communication was not engaging enough. So, those critical communication skills are in short supply, and anyone who can demonstrate their abilities is liable to stand out from the crowd.

One of the most common grumbles I hear from clients is they struggle to recruit effective communicators into their data teams. A client once asked me where he could find an effective communicator and I responded, "I don't know, because we don't learn this in stats school, we acquire these skills in the workplace."

The Translator Role

When I wrote *Data Visualization,* job titles like data analyst and data scientist were somewhat fluid between organizations. Now, it seems clearer: Data analyst jobs focus on summarizing data, producing reports and dashboards, and answering quick-fire requests, while data scientist jobs involve managing pipelines of data from storage to query or model, or creating the models themselves for prediction or causal inference.

There is still little mention of the "translator" role, suggested in 2015 by Thomas Davenport in a blog post for Deloitte Consulting, alongside "light quants" (data analyst) and "heavy quants" (data scientist, statistician, machine learning engineer). In a software development environment, the translator is more likely to be called a product owner, whose specific responsibility is interfacing between the client, or boss, and the developers. They must define what success looks like for the project and check that it delivers value. This is central to effective statistics and links to all the steps taken for good data visualization. Let's consider how it fits into the statistics workflow.

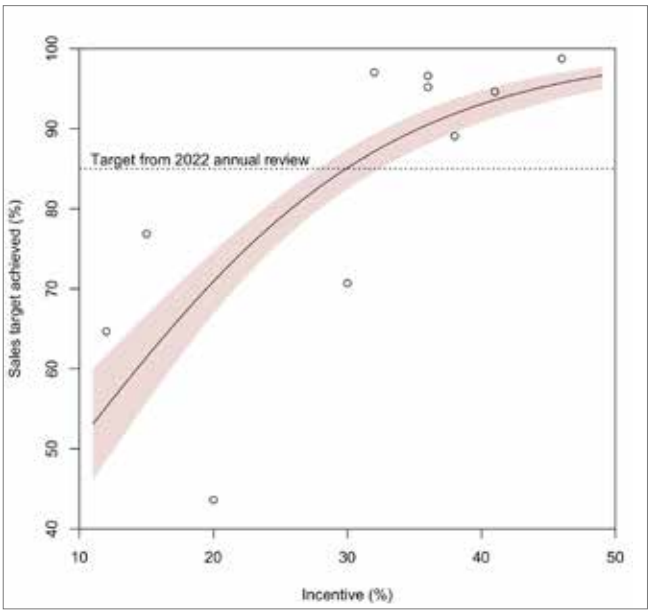
Partners, Not a Help Desk

The analysis can only deliver value if it answers the question the audience has. Someone must identify that question and map it to specific statistical goals. Often, the audience (e.g., your client, boss, or colleague) does not know how to ask for specific statistical goals and translation is needed. We are there to help them do this, as partners. They understand how they want to use insights from the data and we understand various tools for extracting insights.

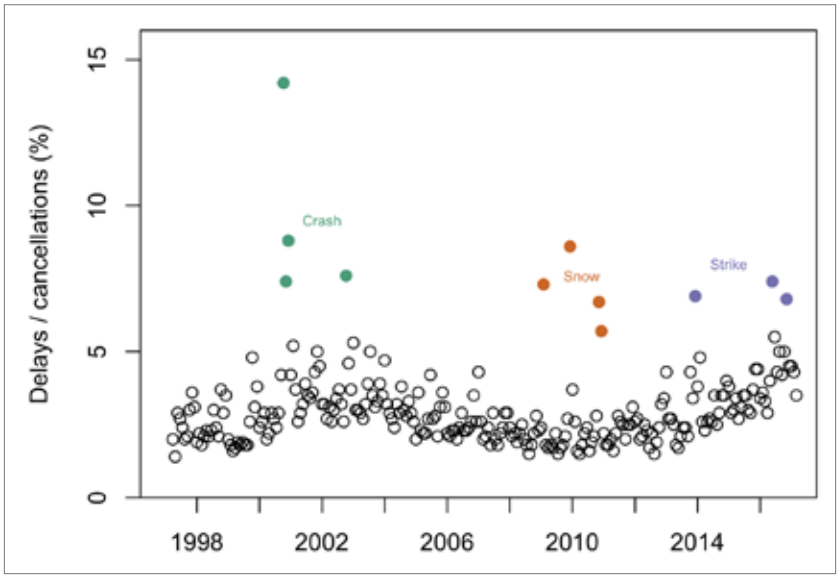
Sometimes, what they ask for might not be what they need most; it may be what they are accustomed to receiving or what they have heard others in the field are doing. They may expect asterisks for statistical significance when perhaps what they really need is a Bayesian posterior probability of an outcome being over or under some threshold. They may ask for a meta-analysis of previous A/B tests, when what they need is a multilevel regression prediction for a specific new scenario already being rolled out.

By taking time to first identify what they really want to be able to do and how they intend to use insights from data, we can help them get the most valuable information. This needs delicate collaboration, and as my years of consulting have gone by, I have become more convinced of this

Robert Grant is a freelance statistician based in Winchester, United Kingdom. His clients have included the World Bank, UK Cabinet Office, Harvard Medical School, and *The Economist.* He has written two books: *Data Visualization: Charts, Maps, and Interactive Graphics* and the forthcoming *Bayesian Meta-Analysis: A Practical Introduction.*



Adding critical annotation can help contextualize the image.



This graph shows a time series of railway delays and cancelations around London, with interesting caveats added.

provocative motto: *Statistics is an interpersonal skill.*

You may need to do considerable work over a long period to shift the relationship from help desk to partnership. The help desk mind set is characterized by expecting rapid responses that directly answer the question without further conversation about it. To be clear, there are times when urgent action is needed and rapid responses are justified, but they should not be the norm. It will be more useful to your audience in the long run to make this change.

In data visualization, this shift in the relationship allows us to adapt our graphics to meet the needs of the audience. An image that directly answers a question—that delivers value, not just eye candy on top of the analysis—is one that will be used. The same is true in private, public, and charitable sectors.

At the same time, you need to maintain some control over the image and how it is used. Adding critical annotation can help contextualize the image. In the graph of "incentive" and "sales target achieved" (invented data), we can see a horizontal dashed line dropped over a prediction from a logistic regression model, inviting the audience to compare predicted results to a "target from 2022 annual review." This kind of annotation can help put numbers in the context of team goals and organizational priorities.

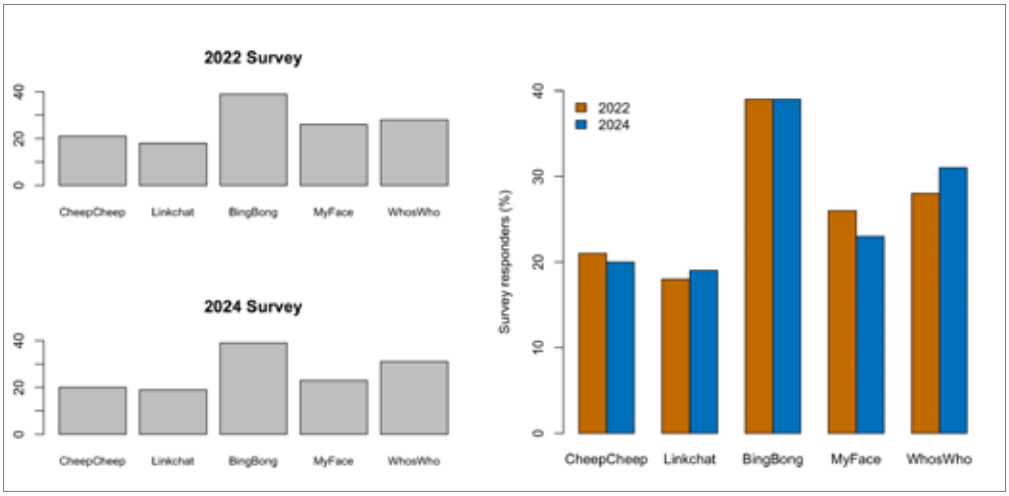
In a graph showing a time series of railway delays and cancelations around London, we can see a general, slowly undulating trend and some outlier periods when there were many more delayed or canceled journeys. If we want to acknowledge the outliers and show they are fundamentally different from the rest of the data, the best way to do it is by direct annotation. However, we must be careful not to add too much clutter, or "chartjunk," the enemy of effective visualization.

Another reason to annotate the graph is that if it is screenshot and used in someone's slides elsewhere without your involvement, the annotation comes along for the ride. Captions can easily be left behind in this way, not to mention lengthy footnotes and methods sections. Make the context and caveats as interesting as the bottom-line message. They are often critical to good decisions.

An Analytical Sandwich

Our work is sandwiched between the data source and audience. We do not generally collect the data, so we must find out what challenges might be hiding in it. These may constrain our analysis choices and our visualizations. Suppose you work for a supermarket chain and must analyze transaction data from the last five years. The data warehousing team can set you up to query the historical data and obtain neat data files.

However, by talking to them, you learn database definitions changed three years ago. The impact of that is the rate of some transaction types suddenly increased, while others dropped. This is not a true reflection of what's going on, but a data management artifact. You can adjust for this in some way, or you can annotate it in your outputs, and that choice will depend on the other end of the sandwich—your audience's needs.



On the left side of the bar charts, we cannot see whether a particular (fictitious) social media platform has gone up or down in popularity, but on the right, when we put them on a common baseline and next to each other, it is trivia.

If the decision-makers need to predict the volume of transactions by type, you would be wise to adjust for this artifact by some model or restrict the data to the most recent. Either way, this needs to be noted in the output, and therefore the visualization. If there are multiple models for adjustment, or options for cleaning, they should perhaps be shown, too, for full transparency and sensitivity analysis.

Another consideration is the statistical literacy of your audience—not just numeracy, but their ability to digest statistical jargon and reason with data and statistics in their decision-making. For less statistically literate decision-makers, we need to go further in summarizing and guiding the decision. Of course, we must be careful not to hide uncertainty.

Even for more engaged decision-makers, there are occasions when little time can be given to the output of your analysis. Especially in long, multifaceted meetings, where your work will be presented alongside many other items, it is a good idea to be prepared with a long and short version of your presentation, just in case you get trimmed—and that includes the visualizations.

Evergreen Tips

Now, let's reflect on some of the tips I have been providing to training course participants. Even when the client asks for a session as short as two hours, there is still a lot that can be conveyed to improve everyday data visualization.

1. Make sure the visual objects you want the audience to compare are close to each other. Related to this, there should be a common baseline, whether that is a horizontal or vertical line or some other aspect of the image. On the left side of the bar charts, we cannot see whether a particular (fictitious) social media platform has gone up or down in popularity, but on the right, when we put them on a common baseline and next to each other, it is trivial.
2. Make sure color schemes are compatible with house style, are color blindness friendly, and will survive black and white printing. You can generate color schemes with websites like colorhexa.com and colorbrewer2.org (the trains outlier chart above used colorbrewer colors). Lisa Charlotte Muth's blog post at *blog.datawrapper.de/beautifulcolors* will take you deeper into understanding color schemes.
3. Reduce clutter by linking elements. When our brains see two objects of the same color in an image, we assume they are linked. This is reinforced by proximity, and these are examples of "Gestalt principles," which are a valuable source of ideas for simplifying your images while still clearly conveying information. This is what happened with color and proximity in the train outliers scatterplot.
4. Use attributes such as shapes or color consistently throughout your entire presentation or report. Once the audience has learned how to read one graph, don't make them do it all over again.
5. Consider adding a "how to read this graphic" introduction. This could be a paragraph, a video, or a little time in a presentation in which you walk through one example of examining a data point or aspect of your visualization.
6. Remember a visualization is not always the right choice. A single percentage can be a short, impactful sentence and may carry more weight than a pie chart (for example).

My clients and learners are often keen to expand their repertoire of software. However, I think the software you use is not as important as the design and interpersonal skills you bring to it. For statisticians, I suggest you learn one fast package for sketching and trying out ideas (e.g., Tableau, the R package ggplot2, or the Python package seaborn) and one flexible programming language that lets you drop shapes, curves, and polygons where you like. This combination will let you craft your charts without constantly struggling to keep up with the latest fashion in software. ■

STAT*tr@k*

Start with the Hero, Not the Story

Mark Palmer

The data nerd in you might recoil: "I'm a scientist, not a creative writer." Wrong. If you work with data, you need to think like an artist.

To do a better job with data, take a moment to define your hero.

Kat Greenbrook, in her *Nightingale* article titled "Reasons to Visualise the Same Data, Differently," advises data analysts start with the who, not the story. Creative people often begin with a hero in mind. They wonder what obstacles their protagonist might face. What makes her act heroically? What is his mission?

The data nerd in you might recoil: "I'm a scientist, not a creative writer." Wrong. If you work with data, you need to think like an artist. Artists start by thinking about their hero.

Greenbrook proposes a framework that can help you find and define your hero. As part of this framework, ask yourself the following three questions:

Am I trying to *discover* insights for myself?

Am I trying to *inform* others?

Am I trying to *educate* others?

When the hero is you, your job is to find a story.

When you're discovering insights for yourself, you're like a detective, using data to find the story behind the facts. As you search, you compose new questions. But you're not working with physical evidence. Your job is to find your clues in graphs, charts, and tables. As you search, new questions emerge. As you collect, collate, refine, revisit, and organize facts, patterns appear. Those patterns are the backbeat of your story.

Once your discovery is done and you find your story arc, you have a story to tell.

When you're informing others, present facts.

When your hero is somebody else, such as a client who needs information, your job is to supply the facts, not tell a story. You're Agent M, feeding James Bond information along his way. Bond, James Bond, is on the hero's journey. In business, James Bond is the project manager who needs a Gantt chart, the sales manager who needs a pipeline breakdown, or the CFO who needs a cash flow statement.

Read Kat Greenbrook's "Reasons to Visualise the Same Data, Differently" at *https://nightingaledvs.com/reasons-to-visualise-the-same-data-differently.*

When you're educating others, you're a storyteller.

When your hero is hapless, helpless, or naïve, you must become a storyteller. Your job is to change hearts and minds. You must take the following steps:

- Find a story (return to hero #1).
- Choose visualizations that create "aha" moments.
- Find vivid words for annotations.
- Select compelling colors, callouts, and comparisons.
- Write titles that stick.

Like writing a book, storytelling with data is grinding, detail-oriented, and sometimes thankless work. I changed the title of this article 10 times—is it the best one, the one that will stick? I'm still not sure and tortured there might be a better one.

You're on a hero's journey, too.

Whichever path you're on, you're on a hero's journey, too. As you stumble through ideas, you discover what you're trying to say. As you explore, you enlighten yourself so you might enlighten others. As Issac Asimov said, "Writing is thinking with my fingers." The same holds for you when you work with data.

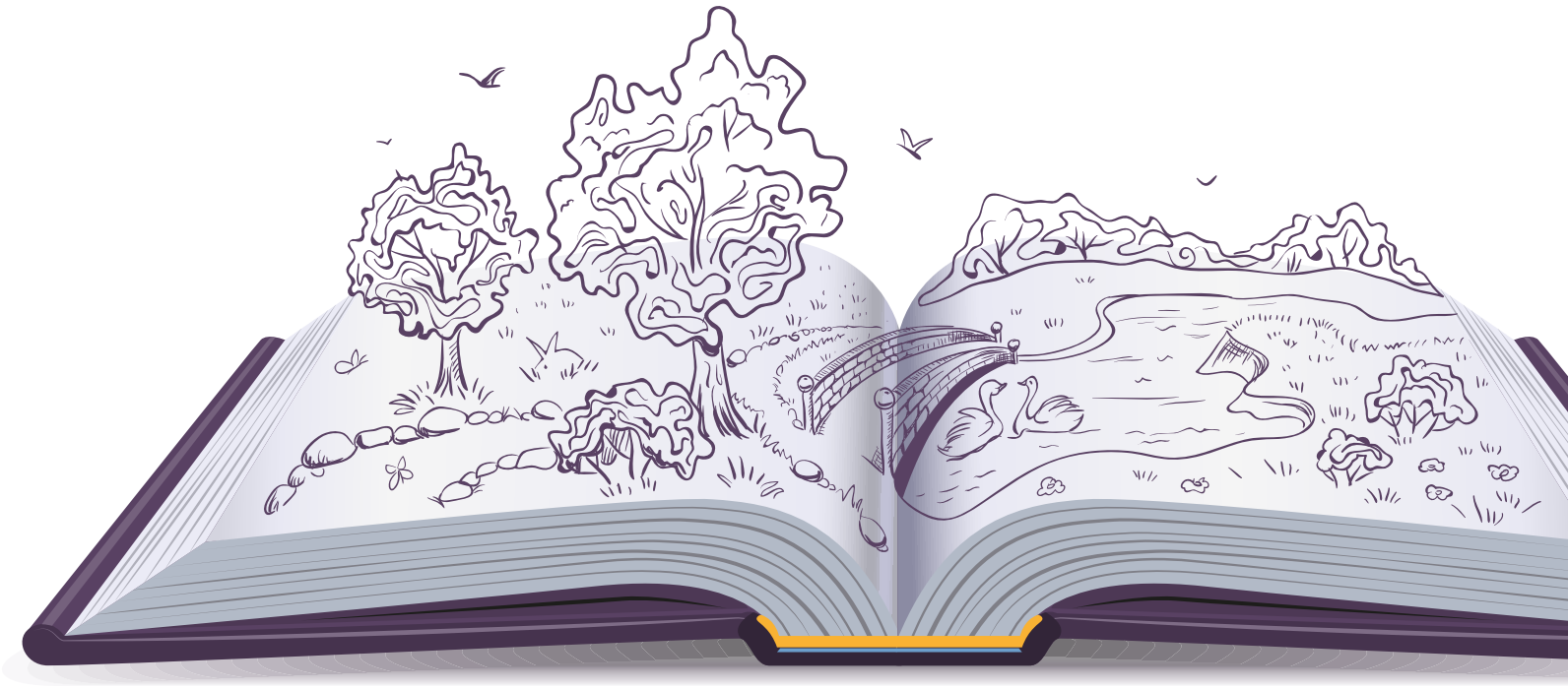
You're on a hero's journey when you're a data storyteller. You're out to change your hero's mind, and you're on the journey together. ■

Editor's Note: This article originally appeared in *Nightingale* (*https://bit.ly/4ds87Pq*) and is reprinted with permission.

Mark Palmer is the former CEO of StreamBase Systems and M&A adviser for Warburg Pincus. He's the coauthor of *Ten Things to Know About ModelOps*, a World Economic Forum technology pioneer, and one of *CDO Magazine's* leading voices in data. He serves on the advisory board of the Data Visualization Society.

STATS4GOOD

Storytelling with Data for Good Visualizations



David Corliss is the principal data scientist at Grafham Analytics and founder of Peace-Work.

In ancient times among Celtic peoples, a seanchaí (anglicized *shanachie*) was a combination of a historian, poet, teacher, musician, and diplomat. The synergy of these talents was used to create and tell stories, often to illustrate a particular concept or idea.

In a similar way, our work in Data for Good today draws on diverse skills in both science and the arts to tell a story directing attention to an important issue and informing ways to address it. One of the most important skills for telling stories with data is effective visualization. Combining statistical science with visual arts and communication, data visualizations are essential for communicating our scientific findings to guide effective action.

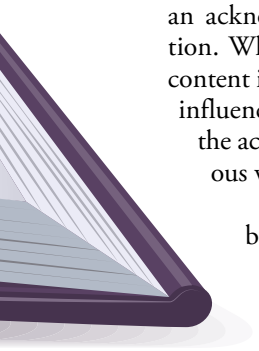
Keep in mind the following for telling your story with data visualizations:

- Have a clear purpose for each visualization. One way to implement this is to annotate the outline for the paper or presentation, specifying a visualization for each point.
- Keep it simple. Avoid crowded visualizations—each image should focus on just *one* story point.
- Make sure there is a legend and caption and all axes in graphs are clearly labeled and described.
- Underscore important information, helping the reader see the focal point.
- Get a review. Ask a person unfamiliar with the content to look at the visualizations, ask questions, and offer comments.

A good place to investigate different visualization ideas is the leading papers in your area. While a literature search is often seen as looking for peer-reviewed content, it should also capture the most effective ways for communicating the subject.

In my own D4G experience, the US Census Bureau does a great job at producing clear, simple, information-rich visualizations—often on the same D4G topics I am investigating. They even have a resource library and gallery of data visualizations you can search for ideas, techniques, and inspiration.

When a data visualization is influential in developing your understanding of a topic and how best to present it, it is appropriate to give an acknowledgement in the paper or presentation. While citations capture specific knowledge content important to the work, contributions that influence in a general way can be recognized in the acknowledgements. As always, being generous with credit is the best practice.



One significant quality to make—or break—a data visualization is the use of color. In a visualization, it becomes a vital part of telling the story. Use color like words to provide information, direct attention, and communicate subtle, nuanced meaning. Learn the color wheel and use it when selecting colors.

Related ideas in a visualization benefit from similar colors. For example, using warm colors for wealth metrics and cool colors for population numbers. Complimentary colors, which are opposite on a color wheel, are particularly good for distinguishing contrasting variables.

Consistent use of color to signify the same items throughout a paper or presentation teaches the audience the color language you are using and makes the data story easier to follow. For example, one could use orange for predicted values and blue for actuals in predictive analytics—high contrast because the colors are complementary and consistently used throughout the work to mean the same thing.

Maps are one of the most effective tools for telling the data's story. Maps are easily understood by all audiences, providing a powerful way to bridge the gap between data and understanding. Always include a legend and add notes on the map to focus attention on the most important information. Two maps side by side with study data in one and additional information in the other to provide context can be effective.

An excellent example of this is a pair of maps created by Alex Najibi. One shows the locations of high-definition surveillance cameras deployed in Detroit, Michigan, as part of a crime reduction program. A paired map has Census Bureau racial demographics, indicating more surveillance cameras in areas with a higher percentage of Black, Indigenous, and people of color population.

Learning practices for effective data visualizations can go a long way toward clarifying and strengthening the message you want to communicate, bringing out the D4G full story for maximum impact. ■

Getting Involved

In opportunities this month, check out Harvard's edX statistics and data science online courses, including one on data visualization. In addition to a low-cost certificate version, there is a free version offering temporary access to course material and nongraded activities.

Also this month, the Census Bureau has released 2020 household-level data with detailed breakdowns by race, ethnicity groups, tribes, and villages. This is an invaluable resource supporting a wide variety of projects. In addition to academic studies, it's also a great place to get the raw data needed for a hackathon targeting leading questions in Data for Good.

We interviewed a few current and former officers of the Statistical Graphics Section to find out why statistical graphics are an important component of the profession. Here's what they said.

Mike Jadoo
Economist,
Bureau of Labor Statistics

Why did you choose to be a statistician/data scientist?

I enjoy problemsolving and statistical analysis, which led me to a career in data science. With statistics, you get to do projects and research that have direct applications in the world. You can help solve a problem directly.

Why did you join the Statistical Graphics Section?

To network and gain a better understanding of the proper use of statistical graphics in statistics and data science.

What have you gained from being a member of the section?

A better understanding about how and why to use data visualizations in reports and dashboards.

List three of your favorite data visualization tools and tell us why you like them.

Horizontal bar charts—they show lots of categories in one chart that is easy to understand.

What advice would you give someone starting their journey in statistical graphics?

- Learn the basics. Start by familiarizing yourself with the basic concepts of statistics and data visualization.
- Choose the right tools. There are many tools available for statistical graphics, such as R and Python (with libraries like Matplotlib, Seaborn, and Plotly), and software like Tableau and Excel. Choose one that suits your needs and preferences and start practicing with it.
- Practice, practice, practice with real data. Theory is important, but practical experience is crucial. Start working with real data sets to apply what you've learned.
- Focus on interpretation. Don't just create visualizations. Focus on interpreting the data accurately and effectively communicating insights. Ask yourself what story the data is telling and how your visualization can best convey that story to others.
- Learn from others. Read books, articles, and blog posts and follow practitioners on social media.
- Stay curious and keep learning.

How do you see statistical graphics evolving in the next decade?

Statistical graphics will be integrated with machine learning techniques to provide deeper insights into complex data sets. This may include the use of algorithms to generate visualizations tailored to specific data sets or user preferences.

What three favorite books do you recommend to others who have an interest in statistical graphics?

- *ggplot2: Elegant Graphics for Data Analysis* by Hadley Wickham. If you're interested in using R for data visualization, this book is indispensable. Wickham, the creator of the ggplot2 package, provides comprehensive guidance on creating elegant and sophisticated graphics using ggplot2.
- *Information Visualization: Perception for Design* by Colin Ware. Ware explores the cognitive principles underlying effective information visualization. This book provides valuable insights into how visual perception influences the design and interpretation of statistical graphics.
- *Storytelling with Data: A Data Visualization Guide for Business Professionals* by Cole Nussbaumer Knaflic. This book focuses on the storytelling aspect of data visualization, offering practical tips and techniques for creating compelling narratives with data.

Lucy D'Agostino McGowan
Assistant Professor, Department of Statistical Sciences, Wake Forest University

Why did you choose to be a statistician/data scientist?

My father and grandfather are both statisticians, so I grew up knowing what a great field this is. However, it wasn't until I completed the Boston University Summer Institute for Research Education in Biostatistics that I really began my journey into biostatistics. I love being a statistician because it gives me the opportunity to help solve different problems across different domains.

Why did you join the Statistical Graphics Section?

I spend a lot of time thinking about how to best communicate complex statistical information to a variety of audiences. Being able to communicate via accurate, useful, and intuitive data visualizations is a key component of this.

What have you gained from being a member of the section?

Through this section, I have met so many statisticians I deeply admire who are similarly interested in finding the best way to communicate complex statistical information.

List three of your favorite data visualization tools and tell us why you like them.

When the intended audience is me, the analyst, the classic scatterplot is always a go-to. Any visualization that gets all the individual data points in front of my eyes is key for making sure I have a good understanding of several pieces of information like the range and spread of the data, any potential nonlinearity, etc. In terms of software, the R package ggplot2 is my go-to for building static visualizations. It offers a high level of customization, and I love the grammar of graphics style, which makes it straightforward to teach and use. For interactive visualizations, I really like Highcharts.

What advice would you give someone starting their journey in statistical graphics?

Find a community to support your journey! There are so many amazing statisticians who are willing to help incoming folks learn and grow in our field. Finding a community can really accelerate your progress (and it's fun!).

How do you see statistical graphics evolving in the next decade?

I hope we continue to learn from the audiences consuming the information we generate. The audience is a key component to consider when deciding how to communicate information, and as the audience changes, so must we.

What three favorite books do you recommend to others who have an interest in statistical graphics?

There are so many great books, and no book is perfect, but a few that have influenced my work include the following:

- *The Art of Insight: How Great Visualization Designers Think* by Alberto Cairo
- *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures* by Claus Wilke
- *The Visual Display of Quantitative Information* by Edward Tufte

---

Edward Mulrow
Senior Vice President and Director, Statistics and Data Science, NORC at the University of Chicago

Why did you choose to be a statistician/data scientist?

From high school through my early graduate school days, I had an interest in applied probability. As I learned more, I became interested in statistics and data analysis. Working in the government and as a federal statistical agency contractor provided me with an appreciation for quality data and proper analysis of that data.

Why did you join the Statistical Graphics Section?

I have always found graphics interesting, whether it was plotting a function in algebra class or visualizing data in statistics class. Because of that interest, I was drawn to JSM sessions related to graphics. I learned those sessions were sponsored by the Statistical Graphics Section, and I joined the section because others in the section had similar interests to mine.

What have you gained from being a member of the section?

The section provided me with an opportunity to be involved in the American Statistical Association and helped foster a great network of colleagues. Naomi Robbins, Rich Heiberger, Di Cook, and Heike Hoffman are a few of the people I have met through the section with whom I have had the good fortune to collaborate on projects.

List three of your favorite data visualization tools and tell us why you like them.

Systat was the first tool I regularly used for data visualization. I don't use it anymore, but I think it created good graphics. Systat was created by Leland Wilkinson, who was a big proponent of statistical graphics and often gave good JSM Statistical Graphics Section presentations. The Systat manual not only laid out how a user could create a good graphic but also explained the purpose of the graphic along with literature references.

Hadley Wickham's ggplot2 R package is my main go-to visualization tool. The "gg" stands for *The Grammar of Graphs,* which is a book by Wilkinson that provides an object-oriented way of thinking about statistical graphics. I find ggplot2 easy to use and intuitive. Once I have a basic graphic designed, it is easy to add or remove features so the graphic suits my needs.

My third favorite tool is my colleagues. Not only can I draw

upon them to create graphics using tools for which I have little expertise, but they are a wonderful group for feedback on my ideas. My favorite graphics are ones created with a team of colleagues who all interjected their thoughts into the process of creating an effective graphic.

What advice would you give someone starting their journey in statistical graphics?

Remember an effective graphic must be understandable to those in the intended audience. Creating effective data graphics can become technical, and some visualizations require complicated programs and techniques. Don't get too engrossed in the programing without considering how viewers will perceive the visual.

Critique your work and ask colleagues for their opinions on the effectiveness of your graphics.

My colleague Nola du Toit and I have a paper titled "Visual Communication of Data: It Is Not a Programming Problem, It Is Viewer Perception" in Computational Statistics in Data Science. Our advice is that creating effective graphics is not just about the technical tools for creating data visualization. It is also about the conceptual and perceptual tools you use to develop and evaluate your graphics.

How do you see statistical graphics evolving in the next decade?

There are a lot of hot topics on the horizon, including access to more data, improved tools for creating visualizations, data storytelling, and many more exciting concepts.

I think it would be good to take a step back and reconsider the makeup of graphics and the expert advice on creating effective graphics.

My NORC colleagues Kiegan Rice and Nola du Toit, along with Heike Hoffman from the University of Nebraska-Lincoln, are reconsidering the question of what viewers see when they view a chart. They plan to build on the foundations of Cleveland and McGill's work in graphical perception by employing a large, nationally representative, probability-based panel of survey respondents to test visual perception. Their approach affords the ability to dive into response patterns among demographics across perceptual tasks and identify significant differences among them.

Past studies that used nonrepresentative or crowd-sourced samples may have missed how viewers with different demographics, such as lower education or lower income, perceive statistical graphics. If so, we may need to come up with a revised playbook on the makeup of a good statistical graphic.

What three favorite books do you recommend to others who have an interest in statistical graphics?

- *Visualizing Data* by William S. Cleveland. After seeing so many references to Cleveland's visualization work, I decided to attend a visualization workshop present by him. I thoroughly enjoyed the workshop and could not wait to get his book. I read it cover to cover in a short amount of time. While there have been many advances in visualization since this book was published, the concepts laid out in the book are still relevant today.
- *Creating More Effective Graphics* by Naomi B. Robbins. This book is a great reference. Robbins provides step-by-step instructions for creating effective graphics. I have often heard her joke that her book is "Cleveland for dummies." While the book is written in a way that reaches a diverse audience, even the most knowledgeable visualization expert will find the book useful.
- *Visualizing Data Patterns with Micromaps* by Daniel B. Carr and Linda Williams Pickle. This book introduces a concept for including geography in a high-dimensional data visualization. Choropleth maps are good but don't handle multiple variables well. The use of micromaps is one way to overcome this problem. The book describes why micromaps are effective and provides many examples of different ways they can be used.

Joyce Robbins
Lecturer in Discipline,
Department of
Statistics, Columbia University

Why did you choose to be a statistician/data scientist?

I've always had a broad range of interests with a special pull toward math and computers. Statistics and data science have a role to play in just about everything in modern life, so it's a natural fit.

Why did you join the Statistical Graphics Section?

I joined to learn about current work in the field and meet people with similar interests. In addition, given the size of the JSM conferences, being a member of a section appealed to me since it provides a sort of "home base"—a smaller group with its own events, sessions, etc.

What have you gained from being a member of the section?

One of the most rewarding aspects of being a member of the section has been organizing sessions on new research directions and connecting people who work on similar topics. One of these sessions addressed the use of JavaScript in statistical graphics with speakers Karl Broman, Carson Sievert, Ramnath Vaidyanathan, and Joy Yang. At another, Jason Cory Brunson, Ursula Laa, and Hengrui Luo presented their latest developments in graphing multivariate data. Organizing a session gives you the power to custom build the panel you would most want to attend at a conference. It's rewarding and a great way for newcomers to the section to get involved.

List three of your favorite data visualization tools and tell us why you like them.

Number one is R/ggplot2. This is my go-to tool for statistical analysis and static visualizations. I could talk endlessly about the power of R, but I'd be preaching to the choir with this audience.

Number two is D3. I love the ability to make unique interactive visualizations, particularly for teaching.

Number three is pencil and paper. Perhaps I should have listed that first because it is so important to think about how you can best represent your data without the limitations of your coding skills or the constraints of your go-to graphics software. It's worth remembering in this context that John Tukey's *Exploratory Data Analysis,* the seminal work in statistical graphics, is based largely on paper and pencil methods.

As different as these tools are, they are all incredibly flexible. None come with a list of chart types but let the user design what works best for their use case. D3 and ggplot2 hit the right balance between high-level and low-level coding, so one can design custom graphics without getting too into the nitty-gritty details (usually).

What advice would you give someone starting their journey in statistical graphics?

Statistical graphics is a field with theory and principles that should always inform one's work. Sometimes, students measure their skill set in terms of the number of graphics tools they know, which is misguided. Keep an open mind and look for opportunities to learn both within the statistical community and outside it. Pay attention to the people who are doing work that interests you and those who are writing the packages you find useful. Follow them on social media and look for opportunities to hear them speak at conferences, webinars, and meetups.

How do you see statistical graphics evolving in the next decade?

Statistical graphics is already expanding from data exploration to model exploration. For example, it is critical in the developing field of interpretable machine learning. I think this trend will continue. In addition, it almost goes without saying that AI will play a bigger role, as it will in so many technical areas.

What three favorite books do you recommend to others who have an interest in statistical graphics?

For those with little experience with statistical graphics, including nonstatisticians, I recommend *Creating More Effective Graphs* by Naomi B. Robbins, which provides practical tips for improving graphs mainly based on [William] Cleveland's work on visual perception. (Full disclosure: I am Robbins' daughter). For those with some knowledge of R who want a comprehensive introduction to the field, I recommend *Graphical Data Analysis with R* by Antony Unwin. Finally, for a more in-depth look at visualization techniques for those with a basic foundation in statistics, I recommend *Visualizing Data* by William S. Cleveland.

Emily Robinson
Assistant Professor,
Department of Statistics,
California Polytechnic State

Why did you choose to be a statistician/data scientist?

Growing up, I enjoyed mathematics and attended Winona State University in Minnesota for secondary math education. When I started college, I knew little about statistical analysis and data science. I approached statistics from a black-and-white perspective. Thanks to the professors there, I gained a new perspective on analytical thinking and problem-solving as I realized the world was variable and subject to change.

Statisticians are challenged to make sense of constantly changing information and communicate this to the public. This was a challenge I wanted to accept. I added a statistics major, did a few summer internships in statistical industries, and ultimately earned my PhD in statistics at the University of Nebraska-Lincoln. I am now entering my third year as an assistant professor of statistics at California Polytechnic State University, where I enjoy teaching, mentoring students, and collaborating on research projects.

Why did you join the Statistical Graphics Section?

I began researching the perception of statistical graphics after my PhD adviser, Susan VanderPlas, printed out a paper on visual inference and graphical testing. I enjoyed the studies and read all the related papers over the next week. Subsequently, I joined the Statistical Graphics Section to connect with others doing similar work.

What have you gained from being a member of the section?

As part of the Statistical Graphics Section, I have built relationships with researchers and individuals interested in improving statistical graphics. It is a tight-knit and welcoming community. I have been invited to give conference presentations, lead workshops, and take on leadership roles through the section.

List three of your favorite data visualization tools and tell us why you like them.

- *d3.js.* I have recently enjoyed learning D3 as a tool for creating interactive graphics. This is used to create graphics in the data journalism sector, and I am intrigued by using it to understand user estimation processes.
- *Dear Data.* While this is not necessarily a tool, this book follows the journey of Giorgia Lupi and Stefanie Posavecas as they create a pen-pal relationship by

sending postcards of hand-drawn data visualizations about their lives that week. I love using this tool in classes to get students thinking outside the box when displaying and communicating data.

- *ggplot2.* This wouldn't be complete without a shout-out to ggplot2. I enjoy how this tool follows the grammar of graphics and layers each aspect of the graphic for easy adjustment between types of charts or separating groups with color or facets. Once you learn the language, you have a lot of freedom in creating high-quality graphics or quickly doing exploratory data analysis.

What advice would you give someone starting their journey in statistical graphics?

Join the TidyTuesday community! Each week, TidyTuesday posts a new data set that is often related to society at that time. Participating in TidyTuesday helps you practice your skills and allows you to see how others approach the same data set. Grab the data set, your favorite tool, and create!

How do you see statistical graphics evolving in the next decade?

In recent years, we have been able to generate graphics quickly, which increases the need for guidelines in creating effective graphics. Moving forward, I think we will continue improving graphics' readability, especially around accessibility (e.g., screen readers, 3D printers, etc.). Additionally, I can see us heading toward more interactive graphics, giving graphic users the autonomy to explore.

What three favorite books do you recommend to others who have an interest in statistical graphics?

- *Show Me the Numbers: Designing Tables and Graphs to Enlighten* by Stephen Few
- *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks* by Jonathan Schwabish
- *Fullstack D3 and Data Visualization: Build Beautiful Data Visualizations with D3* by Amelia Wattenberger

Susan Vanderplas
Assistant Professor, University of Nebraska, Lincoln

Why did you choose to be a statistician/data scientist?

As an undergraduate, I wanted to be anything other than a statistician. My dad is a statistician (he's now retired) and, obviously, the stuff he did all day was boring.

After trying my hand at cognitive psychology as an undergrad and bioinformatics as a grad student, I realized I liked working with data—I just didn't want to be confined to linear models and significance testing. Luckily, there's more to being a statistician than ANOVA and linear regression (something I definitely didn't realize as a teenager).

I ended up combining my interest in cognitive psychology with my interest in data and primarily work on problems related to the intersection of human perception, algorithms, and data visualization.

Why did you join the Statistical Graphics Section?

It's a fun place to be! I think I officially joined after I was recruited to represent the section in the Council of Sections but, as a student, I always went to the Graphics/Computing mixer at JSM because it was fun. It's still one of my favorite parts of JSM because I get to see all the people in my field in one place and hang out.

More generally, graphics are an important part of how statisticians communicate with the rest of the world and it's good to interact with people interested in improving how statisticians use graphics.

What have you gained from being a member of the section?

A chance to network and make friends across the space of graphics and visualization. I've also had opportunities to develop workshops, serve on award committees, and put together sessions at JSM and the Symposium on Data Science and Statistics relating to graphics and data visualization.

List three of your favorite data visualization tools and tell us why you like them.

- ggplot2 is my main tool for creating charts. It's lovely because of the clear implementation of the grammar of graphics, but it also has a ton of useful packages that extend its functionality.
- My next tool is my tablet and stylus or a set of felt pens and paper—sketching is probably the next most important tool I use regularly.
- Finally, shiny is a great interface for creating interactive displays. I use it for everything from data collection at the beginning of a project to data and model exploration to presenting results intended for others' consumption. As an interactive graphics tool, it's definitely limited in scope, but the features it has are extremely useful for many common tasks.

What advice would you give someone starting their journey in statistical graphics?

You'll almost always end up needing to learn multiple tools and/or programming languages. As you pick up new tricks, though, it's useful to think about how something is implemented in R and ggplot2 versus python and seaborn, for example, and use those comparisons to build a concrete understanding of both the tool and wider theoretical space. The grammar of graphics looks different in ggplot2 than seaborn objects or VegaLite, and recognizing those differences can help you pick the right tool for a job, but they can also help when you set out to design a better tool.

How do you see statistical graphics evolving in the next decade?

One area of research I expect to take off is in nongraphical representations of data. Data sonifications, for instance, have to be interactive because sound is instantaneous, so designing a system/grammar for creating data sonifications not only makes data more accessible to people with visual impairments, it also stretches concepts we take for granted about visual representations.

We also don't know as much about perception of sound, touch, etc., so there is a lot of work to be done on how to leverage the strengths of other senses to communicate data-related findings effectively. I love the idea of being able to add additional information to a visual display through sonification, as well as the potential to leverage touch using 3D printed data representations.

I also hope we get better tools for interactive and linked charts that are as well supported as ggplot2 and matplotlib. Having the ability to highlight and follow points through different visualizations generated on demand and linked, as you could do with tools like ggobi, is something currently missing from the space of tools available for modern operating systems.

The tech nerd in me would also love to see an implementation like ggobi that works with headset-based tools—so you can generate plots and move them around a room space—with each plot linked as specified. I

don't have those kinds of programming skills, and it remains to be seen whether VR headsets will become standard tools for offices and video gaming, but I'd love to see something like that in the future.

What three favorite books do you recommend to others who have an interest in statistical graphics?

- *Visualization Analysis and Design* by Tarama Munzner. This is a great introductory textbook that covers the perceptual concepts important for creating many types of good visualizations and approaches to testing and validating visualizations.
- *The Statistical Atlas of the United States.* There are atlases for the 1870, 1880, and 1890 US Census. They're available for free online at the Library of Congress (*www.loc.gov/item/05019329*). The atlases contain a visual exploration of the United States during westward expansion and come from a time before graphical conventions were standardized. Not everything they tried worked (I've written a paper about framed spine plots, for instance, which did not work very well), but the graphics are absolutely fascinating. It's amazing to think about the processes they had to use to generate the color images—adding wax to stone and scraping it off precisely for each layer of the three to four colors of ink required to create a full color chart. I'm quite grateful modern tools don't require that level of patience or artistic precision.
- *Getting (More out of) Graphics: Practice and Principles of Data Visualization* by Antony Unwin. This book has a lot of great examples of good graphics (and some fascinating data sets) and great discussions of the pros and cons of different design decisions and the insights that can be gained from different representations of the same data.
- *Thing Explainer* by Randall Munroe (a bonus book not specifically about statistical graphics). This isn't a statistical graphics book at all. It's written by the guy responsible for *xkcd*. In the book, he uses illustrations and the thousand most commonly used words in English to explain complex subjects such as how nuclear reactors work and the electromagnetic spectrum. What I love about this book (I keep a copy on my desk.) is it's a great example of how visual illustrations and simple text can be combined to communicate technical information without confusing people.

Emily Zabor
Biostatistician, Cleveland Clinic

Why did you choose to be a statistician/data scientist?

I became a biostatistician because I wanted to use my strengths in mathematics and critical thinking to address real-world research questions. More practically, I knew it was a field with many job opportunities across a wide variety of applications, so I thought it would be a good long-term career in terms of job options and ability to find a position that would interest me no matter where I wanted to live.

Why did you join the Statistical Graphics Section?

I strongly believe in the importance of graphics to convey information effectively, so I joined the Statistical Graphics Section to connect with others with similar interests and to learn about cutting-edge developments in statistical graphics. For me, graphics are not secondary to the main results of an analysis but are a critical piece of the analysis itself to understand the data and convey the results.

What have you gained from being a member of the section?

I have met many other statisticians and data scientists with expertise in diverse areas such as color theory and accessibility and have been exposed to many graphical topics I had not previously considered. As a result, I am better able to create graphics that are not only more visually appealing, more consistent, and more informative but also more accessible to a broad audience.

List three of your favorite data visualization tools and tell us why you like them.

I do all my data visualization in R and love using the ggplot2 package for its layered syntax and customizability, the plotly package for creating interactive visualization, and the viridis package for its wonderful color palettes.

What advice would you give someone starting their journey in statistical graphics?

I would advise someone to elevate graphics from something done as an afterthought to a critical part of every data analysis and considered up front. Think carefully about what question you are trying to answer and how a graphic can help address it. I often write the pieces of information I want to convey and then think about how to map the various parts to shapes, colors, dimensions, axes, etc. I usually make many changes to my plans before I actually begin making graphics.

How do you see statistical graphics evolving in the next decade?

I think interactive graphics will continue to be used more frequently to convey more levels of detail than what can be achieved with static graphics. I also think statistical software to standardize and streamline the most common graphics will continue to improve so even the simplest graphics will become more readable and visually appealing with little customization.

What three favorite books do you recommend to others who have an interest in statistical graphics?

- *Storytelling with Data: A Data Visualization Guide for Business Professionals* by Cole Nussbaumer Knaflic
- *Resonate: Present Visual Stories That Transform Audiences* by Nancy Duarte
- *Beyoncégraphica: A Graphic Biography of Beyoncé* by Chris Roberts (just because it's fun)

Professional Opportunity listings are shown alphabetically by state, followed by international listings. Vacancy listings may include the institutional name and address or be identified by number, as desired.

Professional Opportunities vacancies also will be published on the ASA's website (*www.amstat.org*). Vacancy listings will appear on the website for the entire calendar month. Ads may not be placed for publication in the magazine only; all ads will be published both electronically and in print.

These listings and additional information about the 65-word ads can be found at *ww2.amstat.org/ads.*

Employers are expected to acknowledge all responses resulting from publication of their ads. Personnel advertising is accepted with the understanding that the advertiser does not discriminate among applicants on the basis of race, sex, religion, age, color, national origin, handicap, or sexual orientation.

Also, look for job ads on the ASA website at *https://jobs.amstat.org/jobseekers.*

Massachusetts

■ The Harvard Faculty of Arts and Sciences, Department of Statistics seeks to appoint a tenure-track faculty in statistics. The application deadline is December 1, 2024. Please share our job posting with your network or apply here: *https://academicpositions.harvard.edu/postings/13763.* We are an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, national origin, disability status, protected veteran status, gender identity, sexual orientation, pregnancy and pregnancy-related conditions or any other characteristic protected by law.

Missouri

■ Missouri University of Science & Technology invites applications for a Kummer Endowed Full or Associate Professorship in the Mathematics & Statistics Department, starting August 2025. The department seeks an excellent scholar with a mathematical or statistical research focus in data science. Learn more about the application process at *https://hr.mst.edu/careers* (Position #00083283) and the department at *https://math.mst.edu.* Review of applicants will begin on October 15, 2024. Apply *https://bit.ly/3AhHwpn.* The University of Missouri System is an Equal Opportunity Employer. Equal Opportunity is and shall be provided for all employees and applicants for employment on the basis of their demonstrated ability and competence without unlawful discrimination on the basis of their race, color, national origin, ancestry, religion, sex, pregnancy, sexual orientation, gender identity, gender expression, age, disability, or protected veteran status, or any other status protected by applicable state or federal law. This policy applies to all employment decisions including, but not limited to, recruiting, hiring, training, promotions, pay practices, benefits, disciplinary actions and terminations. For more information, visit *https://www.umsystem.edu/ums/hr/eeo* or call Human Resources at 573-341-4241. To request ADA accommodations, please call the Office of Equity & Title IX at 573-341-7734.

New York

Assistant Professor in Statistics, Tenure-Track. Applications are invited for a tenure-track position in statistics at Vassar College to begin fall 2025: see full ad at *https://apptrkr.com/5410007*. The ideal candidate is committed to excellence in scholarship, undergraduate teaching, and working with the department's three statisticians to expand statistics and data science. Application review will begin October 7, 2024, and continue until the position is filled.

AMSTATNEWS ADVERTISING DIRECTORY

Listed below are our display advertisements only. If you are looking for job-placement ads, please see the professional opportunities section. For more job listings or more information about advertising, please visit *www.amstat.org*.

professional opportunities

software

Top Ten **Rejected Professional Development Course Ideas**

Amstat News continues its entertaining offering by ASA Executive Director Ron Wasserstein, who delivers a special Top 10—one that aired during a recent edition of *Practical Significance*. Wasserstein views the Joint Statistical Meetings as a means to professional growth and says, "JSM is loaded with continuing education and professional development opportunities, and you should engage with them if possible. But, alas, not *all* the ideas for professional development we receive are the best. So, in our relentless efforts to improve the lives of our podcast listeners, here are the top 10 rejected professional development course ideas. Don't sign up for courses like these!"

Wasserstein

10

The Lone Wolf Statistician: Collaboration Is for Losers

09

Writing Your Own Yelp Reviews

08

Using Transparencies for Your Presentations (Featuring a Mini-Course on FAX Usage)

07

Speaking with Your Back to the Audience and Other Skills for the Shy Statistician

06

MCMC on Your T I- 83: Sure You Can!

05

Making Your Convenience Samples Even More Convenient

04

Six Easy Ways to Make Your AI-Generated Work Look Like You Did It

03

How to Have Opinions on Things You Know Nothing About

02

Keeping Those Healthy Impulses in Check

#01

Living the Ron Wasserstein Way!

PODCAST

To listen to the *Practical Significance* podcast, visit *https://magazine.amstat.org/podcast-2*.

International Conference on Health Policy Statistics

Statistical Innovation to Improve Health Equity

January 6–8, 2025 • Mission Bay, San Diego, California

ICHPS provides a unique forum for practitioners, health service researchers, methodologists, health economists, and policy analysts to exchange and build on ideas to disseminate to the broader health policy community.

PARTICIPATE

Speaker Registration Deadline
October 15, 2024

ATTEND

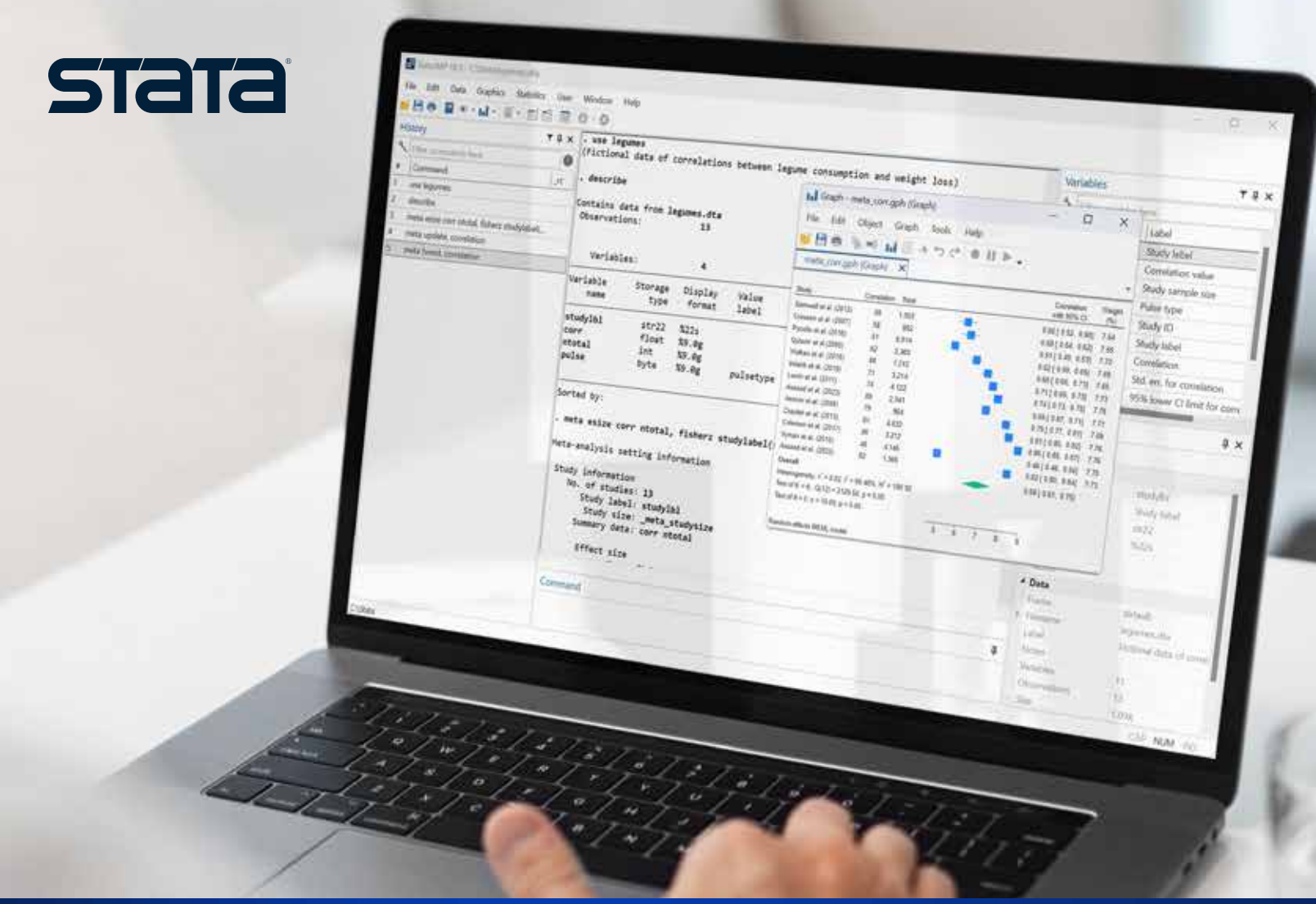
Early Registration
September 10 – December 4, 2024

Regular Registration
December 5, 2024, – January 8, 2025

Housing Open
September 10 – December 12, 2024

Learn more at *ww2.amstat.org/ichps* or scan the QR code.

STATA



StataNow™

New features straight from development to you.

Enjoy seamless access to Stata's latest features with StataNow. This continuous-release version provides users instant access to the newest statistical capabilities, reporting tools, and interface enhancements without waiting for major releases.

Whether you are a seasoned statistician or new to the world of data science, Stata is the software you need to power your research. With Stata's intuitive interface, broad statistical features, publication-quality graphs, and powerful programming tools, you can confidently explore and analyze your data.

stata.com/amstat-statanow