THE STATISTICAL DISCOVERY MAGAZINE. FROM SAS.

JMP® FOREWORD

Can data analytics help protect life below water?

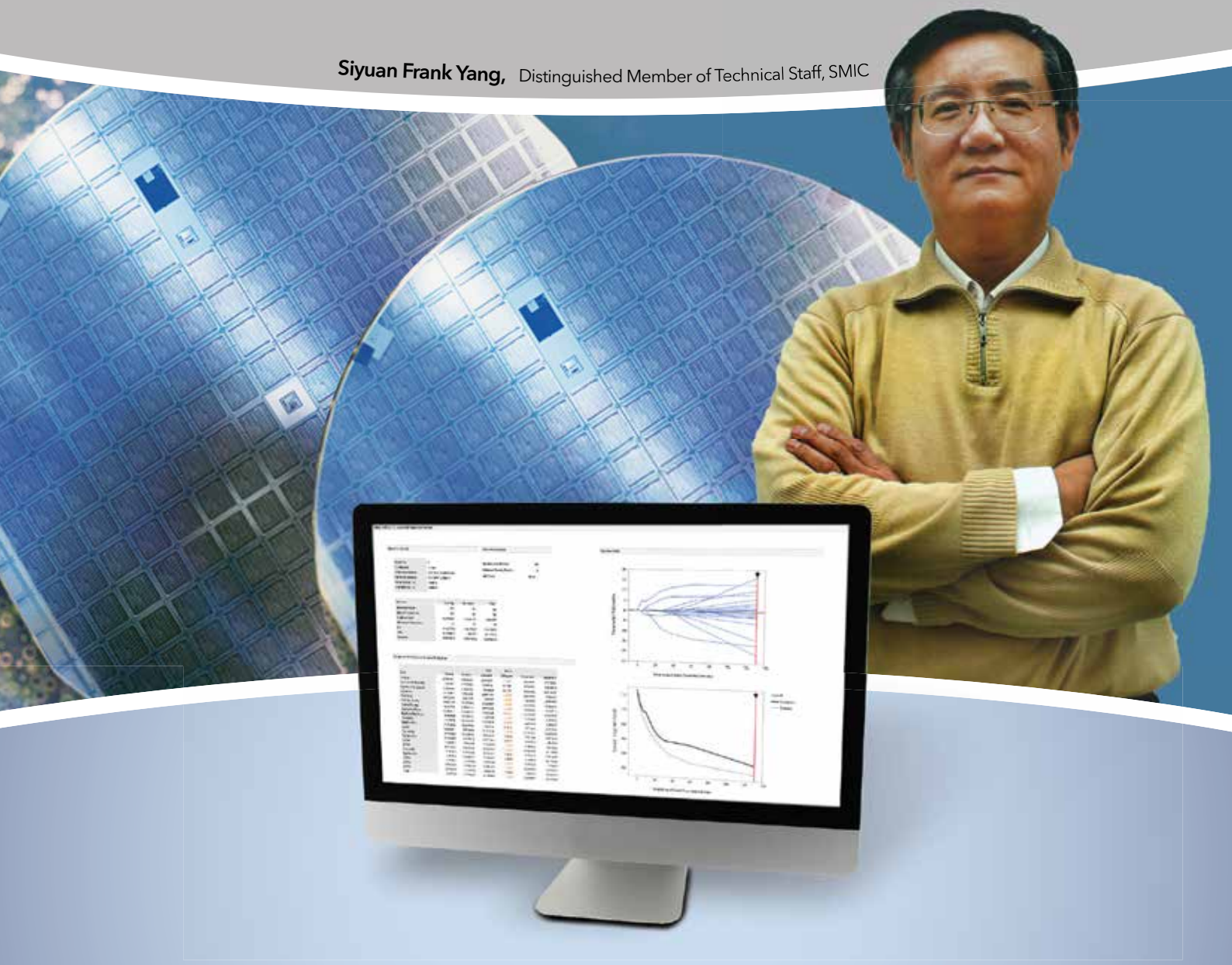
Quantifying your life's purpose

Winning the World Series

Machine learning in modern manufacturing

What do early-career scientists need to succeed?

Get ready to show off your data



Great software in the right hands can change the world.

At SMIC, semiconductor engineers speed up innovation and shorten time to market. In high-tech, timing is everything.

Read about SMIC's success, and find out how JMP can help you change your world:

www.great.jmp

jmp®

Statistical Discovery.™ From SAS.

WHO'S WHO

EDITOR-IN-CHIEF
Jessica Marquardt

FEATURES EDITOR
Meg Eberle

COLUMNS EDITOR
Anne Milley

TECHNICAL EDITORS
Landra Cunningham, Daniel Valente

ASSOCIATE EDITORS
Amy Stuart, Trey Whittenton

ART DIRECTOR
Brian Lloyd

EDITORIAL DIRECTOR
Diana Levey

EDITORIAL AND TECHNICAL CONTRIBUTORS
Nina Chan, Ian Cox, Xan Gregg, Curt Hinrichs, Ruth Hummel, Atsuko Iono, Sachiko Ishibashi, HoKeun Kil, Larry LaRusso, JinYoung Lee, Ryan Lekivetz, Rita Lobo, Chuck London, Rita Lv, Geoff Mann, Bernard McKeown, Arati Mejdal, ChulHee Min, John Mosier, Valérie Nedbal, Carla Paquet, Julian Parris, Elias Pickrahn, John Powell, Gregory Rideout, John Sall, Eric Schoeniger, Maria Astals Serés, IkJu Shin, Mia Stephens, Julie Stone, Isabelle Swaenepoel, Lou Valente, Jon Weisz, Ashley Wong, Gina Xie

PHOTOGRAPHY
Brahim Abdellaoui, Erik Bishoff, Meg Eberle, Diana Jarvis, Anderson Mayfield, Steve Muir, Julian Parris

WEB SUPPORT
Ryan Clare, Aaron Kennedy

CIRCULATION
Alexa Baltazar, Jeffrey Brittain, Amy Bullock, Tina Cui, Graeme Jones, Sean McLawhon, Stephanie Nelson, Susan Potts, Antoine Smith, Eric Sundstrom

PRODUCTION
Kristy Lynn Flynn, Melody Fountain, Lori Harris

HOW TO REACH US
SAS Institute Inc. World Headquarters

JMP Division
SAS Campus Drive
Cary, NC 27513
919 677 8000
jmp.com

For US and Canada sales inquiries, call 877 594 6567.

JMP also has regional offices in China, France, Germany, Hong Kong, Italy, Korea, Malaysia, Singapore, Taiwan and the United Kingdom.

ABOUT JMP AND SAS

First launched in 1989, JMP is a software solution from SAS. John Sall, SAS co-founder and Executive Vice President, is the chief architect of JMP. SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 83,000 sites improve performance and deliver value by making better decisions faster. Since 1976 SAS has been giving customers around the world THE POWER TO KNOW.®

This issue of JMP Foreword was published October 2019. The magazine also lives online: jmp.com/foreword

JMP®: Making discovery possible for 30 years

It was 1989. Computer scientist Sir Tim Berners-Lee proposed and successfully implemented the first communication between a Hypertext Transfer Protocol (HTTP) client and server via the internet. The World Wide Web was born.

Steve Jobs had unveiled the original Macintosh five years earlier, and by 1989 it was clear to SAS co-founder John Sall that he could exploit its groundbreaking point-and-click GUI to give more researchers a way to explore and analyze data graphically without the need to code. JMP was born.

Many scientists and engineers in the late '80s never fathomed they could do their own computing, let alone make breakthroughs like the ones you'll read about in this issue. Today marine biologists use predictive modeling to prioritize coral reef conservation efforts in the face of climate change. Engineers at Lynred design high-performance infrared technology for probes traveling deep into space on a voyage of discovery. NXP Semiconductors accelerates chip technology for vehicles connected to the Internet of Things.

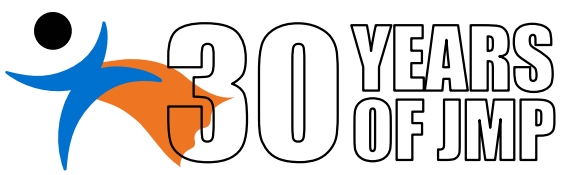
In his column, author Vic Strecher reminds us of another hurdle of that time: the tremendous cost of computationally exploring the nuances of what he calls "unmeasurables." He shares that his ability to study concepts like stress and purpose in life is "enhanced by the set of powerful, elegant data analysis and exploration tools offered in JMP."

Sometimes we see history repeat itself. While ubiquitous internet use has resulted in large amounts of incidental data, columnist David Salsburg notes that the pressure to find hidden "truths" in masses of interrelated (and often messy, incomplete or highly correlated) data is not a new problem.

The internet revolutionized how we learn and communicate. We believe JMP can revolutionize how an organization learns and communicates about its data. Our latest product, JMP Live, extends the interactive capabilities of JMP to the web, allowing organizations to publish, share and collaborate on data analysis projects.

As you read on, you'll find countless examples of discoveries and innovations made possible by the evolution of technology over the past three decades. I wonder what the next 30 years will bring.

Jessica Marquardt
Editor-in-Chief



COLUMNS

NEWS

FEATURES

The unmeasurables: Exploring the subtleties of what it means to be human

by Vic Strecher

Vic Strecher is a behavioral scientist, professor, and Director of Innovation and Social Entrepreneurship at the University of Michigan's School of Public Health. He is also founder and CEO of Kumanu, a company that integrates the science of purpose and well-being with advanced predictive analytics. His most recent book, *Life On Purpose: How Living for What Matters Most Changes Everything*, examines the science and philosophy of purpose in life and ways to develop and align with one's purpose.

"In a way, you are poetry material; you are full of cloudy subtleties I am willing to spend a lifetime figuring out."
Franz Kafka, *Letters to Milena*

"The most exciting breakthroughs of the 21st century will not occur because of technology, but because of an expanding concept of what it means to be human."
John Naisbitt, *Megatrends*

Back in the early 1980s, I was working as a research assistant for Jim House, a sociologist at the University of Michigan's Institute for Social Research. We were exploring the impact of job stress using a large longitudinal study of people living in Tecumseh, MI, a small town between Ann Arbor and Toledo.

From this study, we'd obtained self-reported levels of occupational stress at two points in time, followed up with over 10 years of mortality data. After nights of expensive analyses, entering thousands of lines of intricate code into the Michigan Terminal System, we found that people reporting job stress over both points in time had three times the mortality rate over the next 10 years as compared with people reporting inconsistent or no stress.

Three times the mortality rate over 10 years. What if the risk factor were not job stress but a suspected environmental carcinogen? Or the pressure against your blood vessels? The media would explode with news, resulting in massive amounts of NIH funding and political muscle to address the problem. But, to many "hard scientists," stress was an "unmeasurable" – an ethereal wisp of a concept that had no direct physical measure and might even change from one minute to the next depending on mood or circumstance. We'd often be asked: "Can you really measure stress?" Whatever we were measuring, we knew it was killing people, so we weren't going to stop.

As an undergraduate student in the hard sciences, I had a similar attitude. In my first psychology course, I once announced, "In physics, I let go of this pen, and it falls toward the earth 100% of the time. In this class, the pen floats up at least 30% of the time! That's not science!"

Besides being a jackanapes, I was a dope. I now believe that the wispy dark energy of our lives –– stress, depression, resilience, optimism, confidence, motivation, vitality, connectedness –– may be among the most substantial and important to study and require the most modern scientific tools.

In the 1980s, the costs of computationally exploring the nuances of unmeasurables was onerous, often hundreds of dollars a night – and that's when computer time was half-price. Everything in my career changed when, by complete happen-stance, I saw John Sall demonstrating a beta version of JMP. Like John, I was an early user of the Apple Macintosh computer. As a "visual thinker" (by the way, nearly everyone is a "visual thinker"), I was attracted to the graphic depiction of data in JMP. Even more important,

I now believe that the wispy dark energy of our lives – stress, depression, resilience, optimism, confidence, motivation, vitality, connectedness – may be among the most substantial and important to study and require the most modern scientific tools.

JMP gave me a way of thinking about data that other programs had never offered. From its inception, JMP announced, "Let's explore!"

For the past 30 years, my ability to study these concepts has been enhanced by the set of powerful, elegant data analysis and exploration tools offered in JMP. Subtle nuances are exposed using JMP, and data exploration has become the same source of joy and relaxation to me as reading a Sherlock Holmes mystery.

To many, I suppose my career has now gone off the deep end, moving far beyond stress to an even more existential concept – purpose in life – a goal or set of goals set to the things most important in your life. Here are some examples:

"To be The Tenor Saxophonist."
– Dexter Gordon (jazz saxophonist)

"To be a composer." – Samuel Barber (composer, when he was 9 years old; he also said at this time, **"Don't ask me to try to forget this unpleasant thing and go play football. Please."**)

"To keep trying to understand a little more deeply the universe we are all in, to try to take one more step on this unending quest. And to have fun along the way." – Ronald Graham (theoretical mathematician)

"be a dr. who takes care of kids with cansur so when they say 'Dr Jason, Sometimes I get so scared I'm going to die' or 'you don't know how weird it is to be the only bald kid in your whole school' I can say 'Oh yes I do. When I was a little boy I had cansur too. And look at all my hair now. Someday your hair will grow back too.' – Jason Gaes (in a book he wrote when he was 6 years old; he grew up to become a golf pro)

This ancient Aristotelian concept has, in large modern cohort studies, been associated with a decreased likelihood of mortality, depression, stroke, heart attack, Alzheimer's disease and hospital visits, an increased likelihood of preventive care activities, self-management of disease, better sleep and better sex.

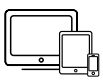
Experimental studies have now shown that one's purpose in life can be primed and strengthened, with resulting improvements in behavior change, health, well-being and resilience. Enhancement of purpose has even been found to increase telomerase, the enzyme that fuels our telomeres – those chromosomal aglets that maintain the integrity of our DNA.

If purpose in life were a drug, a diet or a new exercise regimen, the world would be injecting it, eating it, and working out to it in all the finest clinics. Since, however, purpose is an "unmeasurable," it receives faint regard from the scientific community or the media.

Will this be at our peril? Over the past 10 years, purpose in life has decreased among college students, along with a doubling of depression and suicidal ideation. And purpose is not just a nice-to-have at the top of Maslow's hierarchy of needs. As James Arinaitwe, founder of Teach For Uganda, told me, "Purpose goes hand in hand with hope. Hope for a better life. Purpose sustains poor people."

My own purpose at this stage in my life is to help others create more purpose in their lives. My latest favorite analytic tools are in Text Explorer in JMP, which I've been using to examine thousands of written purposes to determine whether some purpose typologies predict better outcomes than others. These tools, combined with the scientific community's openness to exploring and better understanding the dark energy of our lives, may help us become better human beings.

God knows we need it.



Online

How can you use data to align with your life purpose? Vic Strecher explains in this interview:
jmp.com/strecher

All charts are biased, but some are useful

by Nick Desbarats

Nick Desbarats has taught data visualization and information dashboard design to thousands of professionals in over a dozen countries at organizations including NASA, Bloomberg, the UN and Hershey's. He currently teaches Stephen Few's foundational workshops.

When I teach data visualization training workshops, I recommend that, if a chart is designed to communicate a particular message, it's often useful to explicitly state that message in the chart as a callout or even as the title of the chart. For example, a "neutral" chart title, such as "Water Levels, 2010 – 2019," might be replaced with the reason why the chart creator felt that the audience needed to see that data, such as "Water Levels Have Become Dangerously High in Recent Years."

Many workshop participants are clearly uncomfortable with this recommendation, and they sometimes raise objections: Aren't we biasing the audience's interpretation of the data by putting our own interpretation right in the chart? Shouldn't a chart just "let the data speak for itself," without the need to explicitly state what we think the audience should get from it? Shouldn't we be trying to "just show the data" in a neutral, unbiased way, and leave any interpretation of the data entirely to the audience?

These objections are predicated on the assumption that chart designers can choose to create completely neutral, bias-free charts. In my opinion, though, that isn't even theoretically possible.

Why not?

Well, the short answer is that every chart is a summary. We're condensing a potentially large amount of information into a smaller amount of information so that it can be consumed more quickly, and so that our target audience can think about the data more easily (see patterns, make comparisons, etc.). Creating any type of summary, though, means making judgment calls. We have to decide what to include and what to omit, what to focus on and what to de-emphasize, what to lump together and what to keep separate. When summarizing data into a chart, there are many such judgment calls that need to be made – for example:

- What data to show (gross expenses or net expenses?).
- How to filter the data (previous 12 months or 24?).
- What chart type to use (lines or bars?).
- What level of aggregation to show (daily? weekly? monthly?).
- What quantitative scale to use (include zero or not? narrow quantitative range or wide?).
- What external data to bring in (previous years' expenses? expense budgets?).
- What the chart title should be.

Most of these design decisions – and many others – are unavoidable. We can't NOT choose a chart type, or a level of aggregation, or a quantitative scale. Some of these decisions can be guided by research-supported best practices (e.g., lines versus bars), but others fall squarely in the "judgment call" category (e.g., what external data to bring in, such as data from past time periods or business targets). Judgment calls are made based on factors such as the chart designer's:

- Understanding of the reason(s) why the audience needed to see that data in the first place.
- Understanding of the audience's familiarity with the data and their overall level of sophistication.
- Data visualization training or books they've read.
- Creativity. (Designers who can think of a wider variety of good design alternatives from which to choose will make better judgment calls.)
- Mood that day, past life experiences and even genetic makeup, all of which can affect what they find easier or harder to cognitively and visually process compared with other people.

This is why, if two competent data visualization experts are given the same data and audience scenario, the charts that they design based on that same information are unlikely to be identical and may even be quite different. Even if both charts are well designed and useful, the fact that they're different means that the audience is going to perceive the underlying reality in at least slightly different ways. Certain trends may be

If two competent data visualization experts are given the same data and audience scenario, the charts that they design based on that same information are unlikely to be identical and may even be quite different.

more noticeable in one chart and less in the other. Expenses may look somewhat problematic in a chart that compares actuals to monthly budget targets, but more problematic in a chart that shows only the deviations between actuals and budget targets for each month.

Can we say that any of these charts are completely "neutral" or "unbiased," though? How would we even determine that? Because designing any chart requires making at least some subjective judgment calls, then, even the most honest, expert data visualization designer in the world can't create a chart that's completely "neutral" or "unbiased."

And it gets worse. Even before anyone decided to make a chart in the first place, there were judgment calls made by others regarding what data to collect, at which frequency to collect it, what related data to collect, etc. Any chart based on that data will inherit those judgment calls, which will, in turn, affect the audience's perception of the underlying reality that the chart is intended to represent.

What I'm not saying

I can almost see some readers' eyebrows making a break for their hairlines at this point, so let me clarify a few things. Firstly, I'm not saying that, because it's impossible to create completely unbiased charts, it's OK to create charts that deliberately represent data in a skewed way. When designing a chart, we should always strive to communicate our best, most honest understanding of the reality that underlies the chart, and to not knowingly distort it. We'll never be able to create charts that communicate a completely unbiased picture of reality, but we should still strive to do so.

Secondly, I'm not saying that, because all charts are biased, they're all equally biased. There's a big difference between a chart that usefully represents some aspect of reality and a chart that grossly misrepresents it. Creating a truthful, useful chart requires knowledge of data visualization principles and best practices, a good understanding of the reality that underlies the data and to have only the audience's best interests in mind. If any of those elements are missing, the chart will be unnecessarily biased, confusing or misleading.

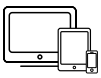
Should we explicitly state our message in our charts, then?

I certainly think so, yes. We never create charts to "just show the data" (although some people may believe that's what they're doing); there's always a higher-level reason. Perhaps a chart would be an effective way to communicate our answer to a question, or perhaps we've discovered a problem, and a chart would be a good way for our audience to quickly understand it.

Whether we realize it or not, the reason why we decided to create a chart in the first place will have guided many of the design decisions that we made while creating it, so we might as well just tell the audience what that reason was. They might not agree with our interpretation of the data or might see something different in our chart, but that's OK. Such discussions are useful since they uncover the fact that people are interpreting the same data to mean different things.

So what should our goal be when creating charts?

Instead of striving to create completely "neutral" or "unbiased" charts, I suggest that we strive to create the most useful chart for a given target audience and purpose. For a chart to be useful, it can't present a distorted picture of reality, so trying to minimize bias is always a "sub-goal" of usefulness. We might capture this goal by rephrasing statistician George Box's famous aphorism that "all models are wrong, but some are useful," as "all charts are biased, but some are useful."



Online

Learn best practices for designing data visualizations:
jmp.com/graph

Raking over the 21st-century muck heaps

by David Salsburg

David Salsburg, author of *The Lady Tasting Tea*, was the first statistician hired by Pfizer Central Research, where he worked on 15 successful products and hundreds of unsuccessful ones and rose to the top of the company's scientific ladder. His publication record includes more than 50 articles in refereed journals and three academic books. He is a Fellow of the American Statistical Association and recipient of a Lifetime Achievement Award from the Pharmaceutical Manufacturing and Research Association.

If anything epitomizes the first part of the 21st century, it is the accumulation of large amounts of incidental data associated with use of the internet. Service providers like Google and Amazon have developed algorithms that they hope will present users with ads they might find interesting. But the pressure is on to use the sophisticated statistical tools now available in software packages like JMP to discover new "truths" hidden in the masses of interrelated data now sitting on servers scattered around the world. How does one handle huge vectors of data where some of the items may be missing, incomplete or highly correlated to other items in this or related vectors?

A similar problem faced R.A. Fisher when he took a post as Chief (and only) Statistician at Rothamsted Experimental Station. For over 50 years, the agriculturalists at Rothamsted had been experimenting with different types of artificial fertilizers, different strains of grains or potatoes, and different fields. Meticulous records had been kept (in some cases) of rainfall, output of grain and straw, different measures of intrinsic "fertility," and anything else the experimenters thought might be useful to know. Formats for recording the data differed from year to year, and as Fisher later discovered, some of the most important information had to be dug out of alternative contemporary records.

Fisher pursued two different approaches to his problem. In the first, he developed designs of experiments that would yield useful information that avoided confounding in future studies. In the other, he dove into what he called "raking over the muck heap." He hunted for sets of data that provided information that could be applied to answer specific questions. These resulted in a series of six publications, titled *Studies in Crop Variation*. Study V does not appear in his collected works, nor is it noted in his bibliography. Like so much of the "muck" accumulated at Rothamsted, it may lie abandoned in some desk drawer. Except for Study III, each of the other published studies tries to dig out some useful subset of these huge, badly structured vectors. In Study III, he attempted to examine the influence of rainfall on crop yields across the years. Fisher developed the concept of multilinear regression to look at the influence of different variables measured at the time on yields of grain.

But such an ambitious look at all the data across more than 50 years was fraught with danger. Many of the measurements were correlated with others. Methods of measurement changed. The most important effect on crop yield was the use of children to pull weeds, and they ceased to be available with the introduction of mandatory childhood education in the middle of the 19th century.

Does all this sound familiar: measurements in different scales, major effects not recorded, unknown interrelationships across the various incidental measures? Throwing all this into one big pot and stirring up a regression does not work, because the interrelationships among possible regressors may take you into what Frank Anscombe called "will o' the wisps" – apparent relationships that are unique to this set of data

[The] pressure is on to use the sophisticated statistical tools now available in software packages like JMP to discover new "truths" hidden in the masses of interrelated data now sitting on servers scattered around the world.

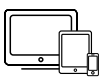
and have no value when it comes to extrapolation.

Fisher's solution was to take a small subset of the data (he suggested using the first n measurements for some small number, n) and look at all possible relationships. Use this initial sample to propose a small set of useful relationships. Then look to see if they hold in the rest of the data.

Thirty years later, John Tukey – now armed with digital computers – had a more formal procedure. He proposed that a small random subsample of data be determined by flagging random vectors as they are put up on the computer. In the 21st century, we have such vast storage capacity and speed of data transfer that we can start with a data file already on the computer and identify a small random subsample. Using this subsample, Tukey proposed what he called an "exploratory analysis." Forget about p-values or percent of variance accounted for. Just look for interesting relationships that make sense. Find a most interesting subset of these relationships. Then, with these in hand, Tukey would have us engage in "confirmatory analysis," where we test for these relationships on the rest of the data, using the tools of statistical significance.

There is another approach to these vast tabulations of what might be useful data. This is to simply look at the data. W. Edwards Deming would prepare pairwise two-dimensional plots, looking for structure within the various measurements, similar to the multivariate plot in the JMP Graph Builder. Personally, I like to see the data stream across a computer screen, using the remarkable ability of the human eye to see patterns. As a simple example, I might have a stream of three-digit numbers when a four- or five-digit number zips by. I once analyzed a clinical study of pulmonary functions measured on different patients for several hours, when I saw a stream of exactly the same values. They were all taken during the last four hours of a Friday afternoon when a single technician was tasked with measuring the patients by himself.

It's amazing what you can find by just looking at the data.



Online

A conversation with legendary statisticians J. Stuart Hunter and David Salsburg:
jmp.com/salsburg

The real work of data science: How to generate and benefit from information

by Ron S. Kenett

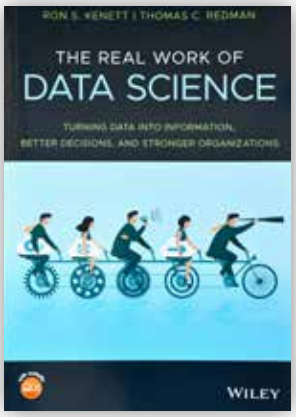
Ron S. Kenett, Chairman of the KPA Group and Senior Research Fellow at the Samuel Neaman Institute, combines expertise in academic, consulting and business domains. He is Past President of the Israel Statistical Association and the European Network for Business and Industrial Statistics. Kenett has authored and co-authored over 250 papers and 14 books on topics such as biostatistics, health care, industrial statistics, customer surveys, multivariate quality control, risk management and information quality. He was awarded the Royal Statistical Society 2013 Greenfield Medal and the European Network for Business and Industrial Statistics 2018 Box Medal for excellence in the development and application of statistical methods

It's 1962. John Tukey has just published a paper in the Annals of Mathematical Statistics, titled "The Future of Data Analysis," stating that "data analysis is a very difficult field. It must adapt itself to what people can and need to do with data." The paper goes on to describe much of what we now call data science. The real work of data science concerns what people can and need to do with data. It combines three critical aspects of modern data analytics work: 1) a need, 2) a methodology and 3) a delivery platform.

Take for example the Kaggle competition for petfinder.my, an animal welfare platform operating in Malaysia since 2008, with a database of more than 150,000 animals. The goal is to improve pet adoption rates by developing analytics tools to guide shelters and rescuers around the world, as they prepare optimized online pet profiles. Let's say you work as a volunteer in an animal shelter in upstate New York. You are new to the shelter organization, you have a background in data mining, statistics and machine learning, and you know about the analytic tools used in this competition. You also think that such tools can greatly enhance the activity in your specific shelter. However, several questions bother you: Is a project that works with data from Malaysia relevant to your shelter? Is the data that's available in your shelter adequate for the task? Is the shelter organization mature enough to adopt such an approach? To make an impact, you need a response to these questions. The answers to these questions are rarely found in traditional textbooks and courses on data mining, statistical learning or business analytics. Moreover, without a computing platform that implements analytic models in an accessible way, nothing will happen.

The Real Work of Data Science: Turning Data Into Information, Better Decisions, and Stronger Organizations, my 2019 book with Tom Redman, tackles those questions head on. In it, we examine how to meet the challenge of implementing data science in organizations. The message is relevant in any sector collecting data on its operations, its customer, its suppliers or its development activities.

Way back in 1962, Tukey envisioned the real work that data scientists are doing in 2019. In other words, the future is here, and it is called data science.

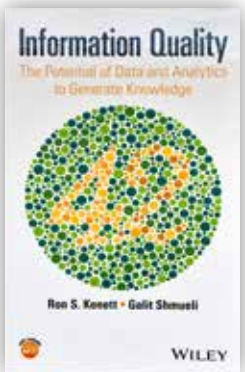


Today's data scientists are doing lots of good work, but there's always room for improvement – for good data scientists to become great ones. Tom offers some insight on the process in an article for the Harvard Business Review: "Good data scientists work to discover hidden insights in vast quantities of often disparate and often poor-quality data. It is a demanding job. Still, good data scientists discover new insights into customer needs, the causes of variability in processes, and how the business is performing that others cannot. They are rare and extremely valuable contributors.

"Great data scientists think about things differently. They are not simply interested in finding new insights in the data. They are interested in developing new insights about the larger world around them. Of course, they use the data to do so. But they also use anything else they can get their hands on."[1,2] Chapter 2 of *The Real Work of Data Science* expands on this.

What else must a great data scientist do?

Great data scientists make it easy for people to understand insights. They root out bias in decision making. In our book, you'll find 18 short chapters written for four personas: 1) a data scientist who works in a midsize department or company and produces management reports, 2) a chief data analyst who manages the data science department, 3) an industrial statistician who is simultaneously bemused and threatened by data science, and finally, 4) a division or company manager who is both excited about the possibilities of deploying data science and fearful that such efforts will fail miserably.



Then there's the issue of ensuring the quality of the information. In my 2016 book with Galit Shmueli, *Information Quality: The Potential of Data and Analytics to Generate Knowledge*, we address this key challenge facing statisticians and data scientists. We identify and define eight information quality (InfoQ) dimensions that can be used to plan and assess analytic work, including:

1. Data resolution.
2. Data structure.
3. Data integration.
4. Temporal relevance.
5. Chronology of data and goal.
6. Generalizability.
7. Operationalization.
8. Communication.

You need to properly address all these dimensions, given your goals, for your data-driven work to achieve appropriate impact. Like the cover of InfoQ, you need to see the pattern, and then you need to know its meaning (Hint: Google *The Hitchhiker's Guide to the Galaxy*). What's the significance of these dimensions? Let's say you are the production manager of a company developing and producing consumer goods. Efficiencies in production and quality of products are your main targets. Do you have data with the right resolution to allow you to achieve your targets? You need to know about the content in each product to ensure quality and track the production yields in the manufacturing plant to manage efficiencies. Simply reporting quarterly yields is not enough. Proper process control data is needed by batch, or shift, so that you can manage the process efficiently. Data resolution is the first InfoQ dimension. Communication is the eighth dimension. If the control charts you maintain are not communicated to the right person, at the right time, in the right way, the efficiency in running processes is hampered.

Turning data into useful information

From the emergence of the concept in 1962, we move to the practical applications of data science being used in 2019. Big data, machine learning and analytics affect everyone, every organization and every type of activity. A wide perspective for data science implementation is required if you want to derive concrete benefits. The data scientist must focus on generating information quality. The 2019 and 2016 books treat both subjects systematically, with examples. They also challenge you to develop your own solutions to your specific challenges. An understanding of the organizational needs, modern analytic methods, and an advanced and accessible analytic delivery platform are three important features of effective businesses, industries, services and national agencies, and pays huge dividends. Developing the competencies to do the real work of data science is a complex challenge, considered one of the most important facing today's organizations.

References

1. Redman, Thomas C. "The Best Data Scientists Get Out and Talk to People." Harvard Business Review, 26 Jan. 2017. https://hbr.org/2017/01/the-best-data-scientists-get-out-and-talk-to-people

2. Redman, Thomas C. "What Separates a Good Data Scientist from a Great One." Harvard Business Review, 28 Jan. 2013. https://hbr.org/2013/01/the-great-data-scientist-in-fo

Online

For more from Ron Kenett on these topics, see this interview:
jmp.com/kenett

Get the JMP add-in for evaluating information quality:
jmp.com/infoqscore

How does JMP align with the InfoQ framework?
jmp.com/infoq

How to get JMP® to make you look like a magician

by Julia O'Neill

Julia O'Neill, founder of Direxa Consulting, has over 30 years of experience bridging statistics and chemical engineering in vaccines, biologics, pharmaceutical and chemical development and manufacturing. Her focus areas include statistical, validation and regulatory strategy support for a broad range of novel accelerated products, including multiple breakthrough designation and orphan drugs. She has a passion for applying statistical thinking to more rapidly establish the evidence needed to bring forward medicines addressing unmet medical needs.

More than once, as I've shared my screen while working through a complex analysis, I've had someone say to me, "you're a magician." It's flattering; it tells me that I've been able to contribute something the team had struggled to do without JMP, or perhaps hadn't even dreamed was possible.

But that JMP magic doesn't happen automatically when you finish the install process. In this column, I'll share a few secrets that help me make the magic with JMP and point out some of the very best features I recommend you try.

The secrets behind the magic tricks

Some of the secrets to becoming a JMP magician reside completely outside of the software. Here are a few I rely on most:

Draw a diagram that matches the experiment, process or data collection that produced the data you want to analyze. This diagram will prove incredibly valuable. I start sketching with colored markers on paper, but then I like to clean it up enough to share digitally.

The first use of this diagram is for communicating. If any others had a hand in gathering the data, you need to check for agreement on the details of the data collection. This is especially helpful with virtual and international teams (although it's just as important for a lab team working all in the same building). In our world today, we may run an experiment on one continent, analyze it on another and transfer the results to a third. Visuals help a lot with accurate communication across time zones, language differences and particularly for sharing information with colleagues from different disciplines. The diagram should be as simple as you can make it (you don't need to show every detail) but should still show the flow of work or information. The devil is in the details here.

Write down the objectives for analysis using action verbs and concrete nouns. Use specific phrases, such as:

- To compare results from two different test methods.
- To identify the most important factors for increasing yield.
- To fit a model predicting potency as a function of process conditions.
- To find operating conditions that are robust to variations in input materials.

My clarity antenna goes off when I see the word "appropriate" – that's a flag to me that we're not quite sure what we want out of the data. Stop right there. Writing it down is much harder than you'd expect, but it's vitally important.

The most effective scientists I've worked with step in to slow things down. Instead of diving straight into a sophisticated experimental design or model fit, they back up to conceptualize the problem first. In one of the most powerful experiments I ever worked on, the lead chemist had the team place their bets in advance on which factors would influence the outcome and how. We were all wrong (including him, which he humbly acknowledged), but what we learned from the data rescued a line of busi-

ness that had been foundering from repeated defects in the product.

Back in 1957, A.W. Kimball wrote an essay called "Errors of the Third Kind in Statistical Consulting": giving the right answer to the wrong problem. His article describes blunders by novice consulting statisticians, and it's still fun to read today. The availability of powerful software in 2019 has democratized this problem so that any statistics enthusiast with a software license is equipped to commit this type of error. But writing down the objectives and drawing a representative diagram help you avoid errors of the third kind.

Wallow in the data. This is the transition point at which I wouldn't want to work without JMP. Graph your data sideways, upside down, sliced up, aggregated – any way you can dream up. This is a great way to find issues such as unintended repeats, typing errors, patterns of missing data and too many others to fully list.

My go-to method for wallowing in the data is JMP software's variability chart. This is where the diagram, written objectives and JMP graphics meet. When you can construct a variability chart that matches the diagram of data collection and addresses the objectives you've carefully described, you are ready to do some fancy statistics! As I begin specifying a model in JMP, I make sure the model matches the organization of the variability chart.

My favorite tricks in JMP® (or things I had to do by hand in graduate school)

Now for more JMP magic – as I jotted these down, I realized that most of them are things I learned in graduate school but had to do by hand at the time. Part of the magic is how quickly you can do these things with JMP. The top half of my top 10 list:

1. **Suspect an outlier?** Instead of agonizing about outlier tests, or arguing about distributions, simply run the analysis. Hide and exclude the suspect results and run it again. Compare the residuals and model results. Does excluding these suspect results change your conclusions in any meaningful way?

2. **Did your experimental design go astray?** If you've encountered unanticipated constraints, simply move to the JMP Custom Design platform to overcome them. If you've already run the design, apply the design diagnostics in JMP to see how much the missteps will hurt your conclusions. New practitioners of design of experiments are often put off by thinking that nothing short of a perfect classic design will do; this should hardly ever be a limitation!

3. **Dealing with random effects?** If you have factors that can't be dialed at different levels but will contribute variability (such as different lots of a raw material), use JMP software's REML feature to estimate variance components. Calculating mean squares by hand in grad school was valuable for learning, but boy am I glad JMP does these models for me now. Just make sure you specify the model accurately (see Part 1 above). If you've drawn a representative diagram of the study and constructed a variability chart to match, you're in good shape to fit the corresponding model in JMP.

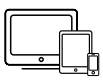
4. **Predicting results from a formulation?** Let JMP draw contour plots and use these to understand the formulation behavior. Understanding mixture models is a headache, but once you learn to read a ternary plot you can compare these for different models.

5. **Looking for patterns?** If you're old enough to remember overhead transparencies, you'll understand when I say I used to draw graphs and physically overlay them. Tedious but powerful. This is now at your fingertips using Graph Builder. And stepping fully into the 21st century, see what the software's clustering feature can identify in your data set.

For the full magic effect, I whip up combinations of these JMP techniques. But I always build on a solid foundation, which means:

- Draw a diagram.
- Write down the objectives.
- Wallow in the data.

Abracadabra!



Online

Julia O'Neill on bridging statistics and chemical engineering for biotech and pharma:
jmp.com/oneill

Could 'supersaturated' be the new superhero?

by Bradley Jones

Bradley Jones is JMP Distinguished Research Fellow at SAS. He is the inventor of the Custom Designer and Prediction Profiler tools and co-inventor of definitive screening designs. He is a two-time winner of the Brumbaugh and Lloyd S. Nelson awards. Jones is a Fellow of the American Statistical Association and the past editor of the Journal of Quality Technology.

Sometimes things align in such a way that the impossible becomes possible.

In the field of experimental design, the goal is to learn the most from a given number of experimental runs. Multifactor experimental designs have already revolutionized research by allowing us to learn a lot, even when there are many factors to consider. For screening designs, generally the number of experimental runs must be greater than the number of factors you want to screen. Supersaturated designs have more factors than runs.

JMP has been supporting a feature in the Custom Design model list where you can give the estimability attribute "If Possible" instead of "Necessary" – which allows you to specify fewer runs than factors. If there are only a few strong effects and the others are weak, then the experiment will find them. However, the statistical tests associated with these effects are weak and depend on assumptions that the weak effects are zero.

Chris Nachtsheim and I recently came upon an idea that would not only allow a supersaturated design to have an estimate of the error variance, but also deliver an error variance estimate that was independent of any of the effects – weak or zero or not. That idea evolved from our work on definitive screening designs, where we recommended using "fake factors" to get an independent estimate of error; we then realized that there was a construction to have a group of fake factors in a special mathematical structure (a Kronecker product of a Hadamard matrix with a subset of rows from another Hadamard matrix).

The first step is to arrange your factors into groups, next, apply this technique to create a design, which we called a "group-orthogonal supersaturated design." The first group is used for "fake factors" to estimate error independently. For the other groups, factors are correlated within the groups, but independent (orthogonal) across groups. You can easily isolate which groups are significant, and then use a selection technique to obtain the stronger factors within each group.

Look for the new design facility, aligned with its automated fitting facility, in JMP 15.

Armed with this new superpower, we can screen more factors than we have runs and still get real significance tests that are independent of the factors. "GO-SSD."

References
Jones, Bradley; Lekivetz, Ryan; Majumdar, Dibyen; Nachtsheim, Christopher J. and Stallrich, Jonathan W. "Construction, Properties, and Analysis of Group-Orthogonal Supersaturated Designs." Technometrics, August 2019

Online

For details, read the Technometrics paper on GO-SSDs:
jmp.com/gossds

Armed with this new superpower, we can screen more factors than we have runs and still get real significance tests that are independent of the factors. "GO-SSD."

Rethinking the statistics curriculum in the digital age

by Bernadette Lanciaux

Bernie Lanciaux is a faculty member at the Rochester Institute of Technology (RIT) School of Mathematical Sciences where she co-chaired the committee charged with rewriting the introductory statistics curriculum. Lanciaux has a PhD and an MEd and has been involved in professional development in statistics education for over a decade. She has been on the program for various professional meetings and has received numerous Faculty Professional Development Grants for her work revising the introductory statistics curriculum at RIT.

We have entered the digital age. Anyone who doesn't want to be left behind must recognize that. JMP is the tool that makes it possible for all of us to understand data the way Amazon, Facebook and Google do. And what is data? Everything is data. I'll save unpacking that for some other article, but the fact of the matter is, JMP puts the power to explore data in the hands of everyone. You don't need to code. Sure, if you want to code, there's a way. If you want to export to Python or some other program, no problem. But regardless of your abilities, JMP lets you play with your data first, which my students appreciate. As an educator, I appreciate how JMP helps my students to know, understand and do statistics fully in the real world. JMP can do it all.

My classes are never taken as electives; they are part of the core curriculum in many of the degree programs we offer at Rochester Institute of Technology (RIT). I have been teaching statistics for many years and have usually been successful, as have my students. But last year I had an unparalleled teaching experience. I was able to wipe the slate clean and rebuild my curriculum from the ground up to reach the learning goals outlined by the programs my classes support. Technology has transformed the world, and the traditional statistics curriculum does not reflect that. Students are both consumers and producers of statistics and data, yet we teach them which formula to use instead of how to analyze data. So I reviewed the list of concepts the students would be tested on for the common core and said to myself, "How should I cover critical statistical concepts? And if I could use the power of JMP to teach these things, how would I approach teaching them?"

I was lucky enough to start with a small summer class. On day one, students downloaded up-to-the-minute data about their hometowns or favorite cities. Within five minutes, they were able to bring the data into JMP. I showed them how to use Graph Builder and put in a background map. They were off! They explored data that had real meaning to them. They were actually curious about what secrets they could unlock with this amazing tool. They loved coming to class! One student couldn't get credit for the class but attended daily, staying in Rochester on his own dime. He recognized the value of what we were doing. The students came for an extra class on a summer afternoon to learn more. Two students mentioned they were looking forward to their final exam, which consisted of conducting analysis and constructing a poster to communicate what they had accomplished as a group in just three hours – an assignment that realistically mimics a professional world where they will work on teams and face tight deadlines. When the fall semester began, they were proud to show off posters of their work hanging outside the classroom. My colleagues asked, "Not graduate students?" Now that is the power of JMP.

I continued to make JMP central to teaching the required content in other courses, even with multiple sections and full classes. JMP empowered me to teach students, rather than material. Wow, what a teaching experience! I encouraged students to think of me as the consultant hired to help each of them build a portfolio to impress at job interviews. Many were scared – scared of failure. I told them that failure is impossible. Saying "This is what I see in the data, and it doesn't tell me anything" is not failing if they meet the curriculum goals. Some groups needed hand-holding; they didn't know where to start. But as they used the dynamic visualizations in JMP and the column switcher in play mode, they painlessly – and eventually playfully – explored their data. Then all they had left to do was describe it. They had something to say. They had a reason to learn.

Online

Check out graphs from RIT student posters showing hometown data in JMP Public:
jmp.com/rit

What's the point?

Just pointing at something tells you about what you are pointing at.

by John Sall

John Sall, co-founder and Executive Vice President of SAS, joined Jim Goodnight and two others in 1976 to establish what is now the largest independent vendor in the business intelligence market. Sall designed and developed many of the software's earliest analytic procedures; he also created JMP software in the late 1980s and remains its chief architect. Sall is a Fellow of both the American Statistical Association and the American Association for the Advancement of Science.

Pointing

In magical tales like the Harry Potter series, wizards can just point at something with their wand, speak an incantation, and magic happens. We do something very similar when we point the cursor at something on our screens, and without even an incantation or a click, magic happens. That is our world of tooltips or hover windows.

Identify points

Graphs of points in JMP are incredibly valuable in seeing the distribution of points over two dimensions. But when you see a point that is far apart from the rest of the data – an outlier – you want to identify that point. By hovering the cursor over the point, a hover window appears showing which point that is, plus you can click to select the point in the table to investigate further.

From the time JMP was first released, hovering your cursor over a point in a graph has brought up a small text window identifying a point. In the first version of JMP, it just showed the row number. But we've continued to add information and functionality with subsequent versions: the row number and any column values given the "Label" attribute; the values corresponding to the coordinates; the ability to "pin" the hover label to the graph persistently; label columns that allow images in the hover label window; an option to use the images as markers. But with JMP 15 there is much more.

What's new

JMP 15 introduces far more powerful hover windows – graphlets. They are built in to a number of platforms where points represent many values and can be used in a customization setting for almost any graph, providing drill-down features.

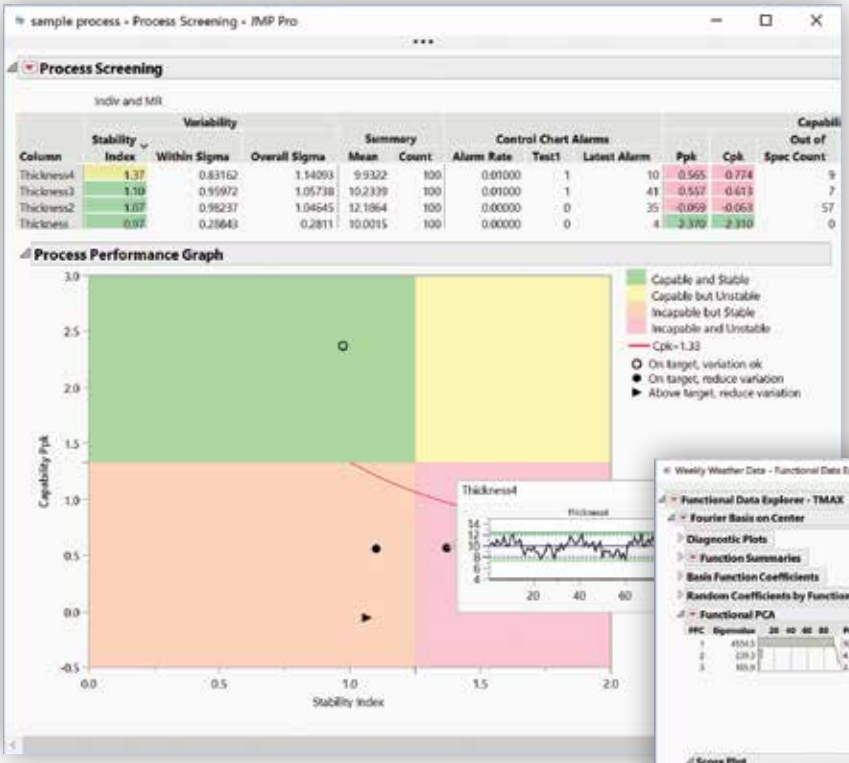


Figure 1: Process Screening

Process Screening

In Process Screening, you can see the health of many processes in one Process Performance Plot. In Figure 1, one process is healthy, the others are incapable, and Thickness 4 is unstable as well. To find out more about the instability, hover over that point and a hover window appears, showing a small graph of the process. Hovering is a lot easier than selecting the point and requesting a quick graph or control chart of that process.

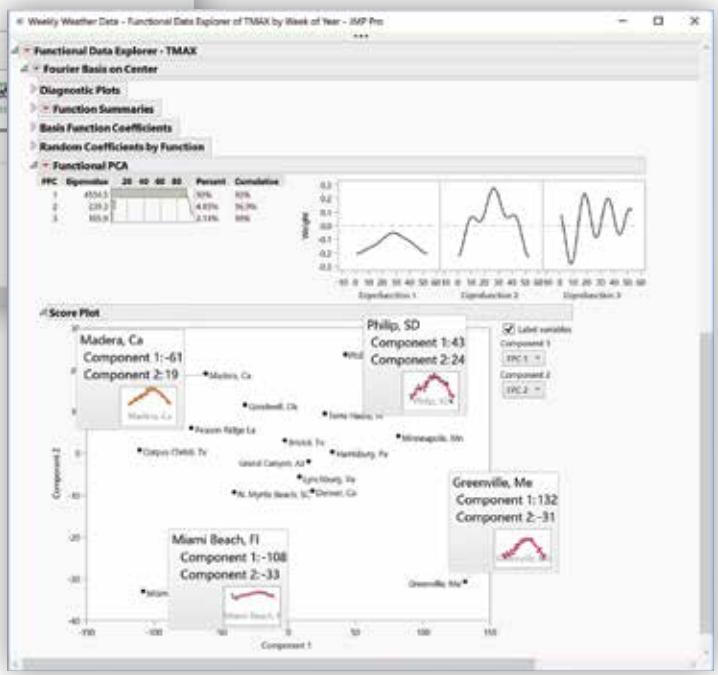


Figure 2: Functional Data Explorer

Functional Data Explorer

The Functional Data Explorer characterizes many trajectories using functional principal components. Figure 2 is a score plot of the temperature functions of 16 US weather stations. To determine what's going on with the score space, points in the four areas have their hover windows pinned. The first component on the X axis is basically how extreme the winter is, while the second component, with less than 5% of the variation, involves the shape of the seasons.

Multivariate Control Chart

In the new Model Driven Multivariate Control Chart (Figure 3), the fact that a point is far out doesn't explain which process is involved in it being extreme. By hovering over the point, the individual differences are shown – in this case, Process 6 is implicated.

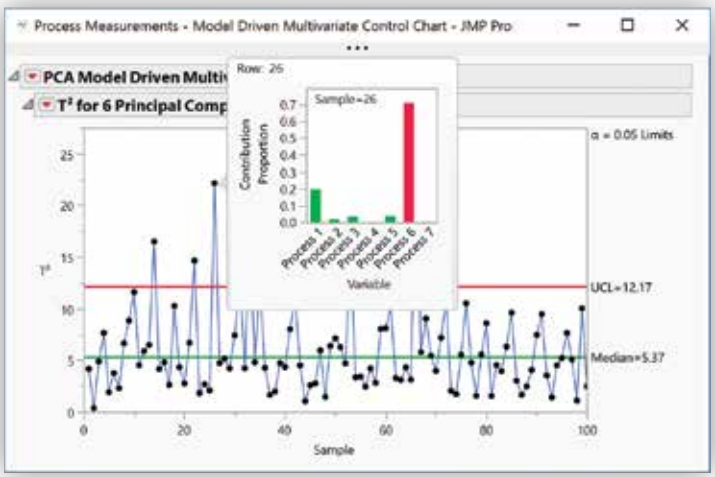
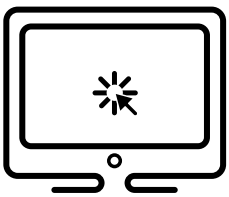


Figure 3: Model Driven Multivariate Control Chart



Good software can make things so easy, it seems like all you have to do is point a wand (or cursor) at something, and magic happens.

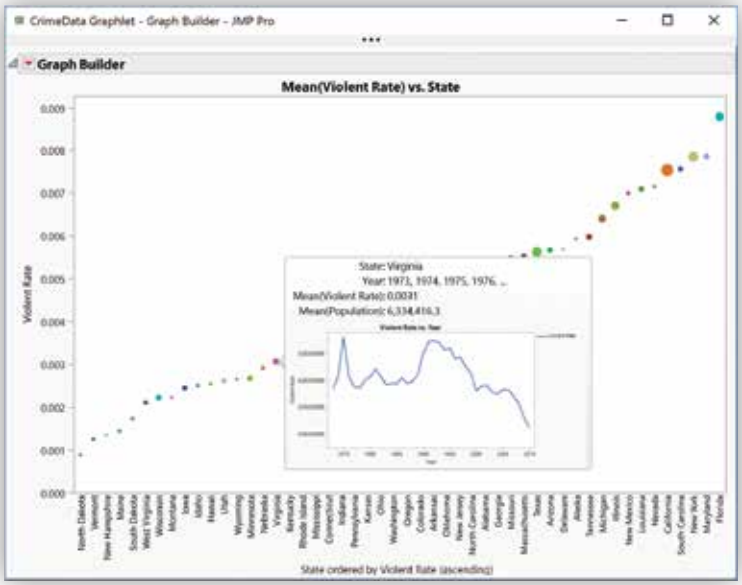


Figure 4: Custom graphlets

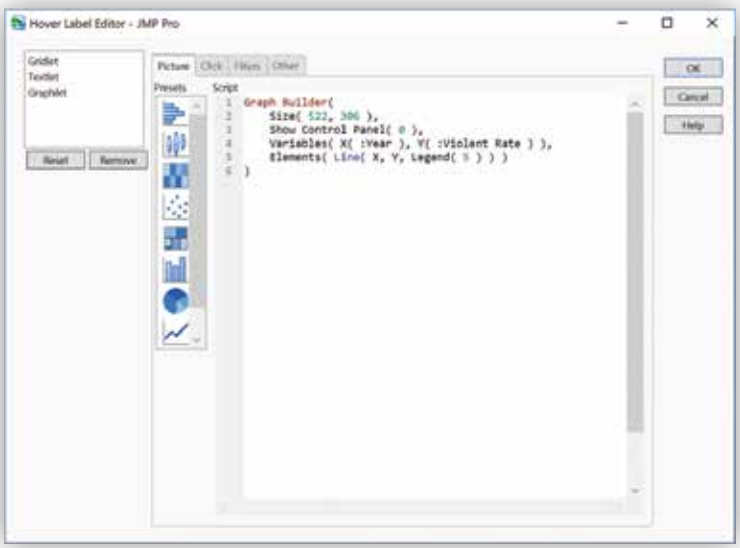


Figure 5: Graphlet Hover Label Editor

Custom graphlets

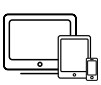
Figure 4 shows the mean violent crime rate for each state over years, but it doesn't show the trajectory. To see that, I made a separate Graph Builder showing the violent rate by year. Next I copied script to clipboard, then went to the original graph and right-clicked on Paste Graphlet. Now when I hover over a point, it makes the graph of that state by year, in this case detailing Virginia.

Further customization

Once you make a graphlet, you can customize its features in a Hover Label Editor. The Graphlet panel seen in Figure 5 shows the script that is run on the subset represented in the point. Gridlet provides control over which elements are shown in the hover window. Textlet allows you to customize the text, including the markups to stylize the text.

Conclusion

Good software can make things so easy, it seems like all you have to do is point a wand (or cursor) at something, and magic happens.



Online

John Sall highlights secret features of JMP:
jmp.com/secrets

What's new? JMP® 15 and JMP® Pro 15

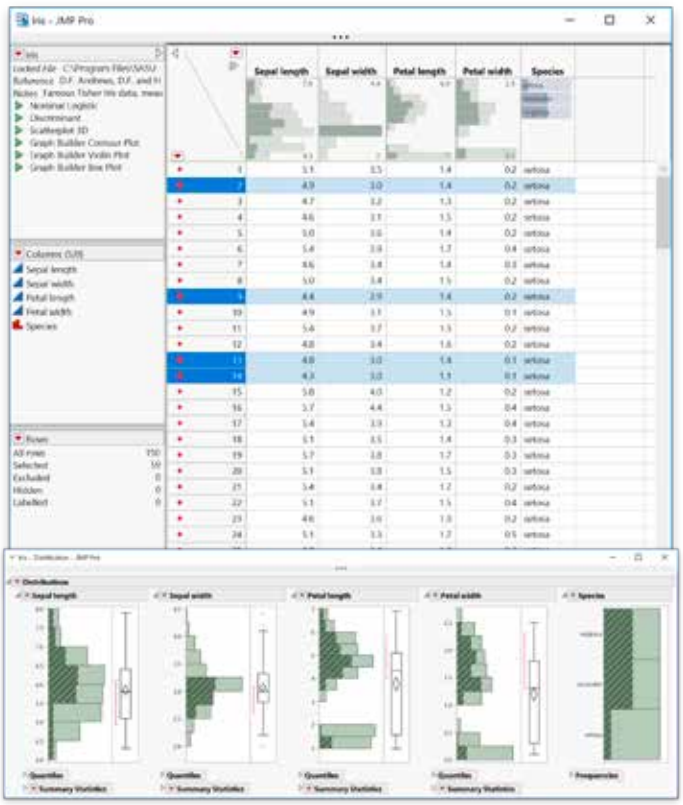
Explore and understand your data more fully

JMP data analysis software from SAS is the tool of choice for scientists, engineers and other data explorers that lets you explore data dynamically and more fully using powerful statistics combined with interactive graphics. With JMP, your analyses and discoveries unfold, driven by what the data reveals at each step.

And for analytics power users who have large, messy, incomplete or textual data, JMP Pro provides sophisticated modeling techniques to remove roadblocks, anticipate outcomes and act with confidence.

JMP 15 and JMP Pro 15 expand the options users have for richer exploration and data analysis and extend the possibilities for sharing results.

Here is a selection of the enhancements you'll find in the latest versions of JMP.

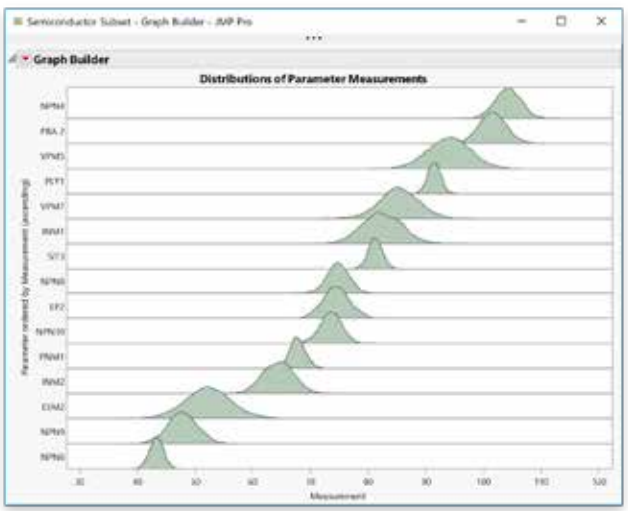


Header graphs in the data table show linked histograms for a quick view of the distribution of table variables.

Data Import, Data Table, Cleanup, Data Visualization, General

- New column header graphs display histograms above each column.
- Import wizard allows import and processing of XML, JSON and PDF file formats.
- Recode improvements:
  - o New column utility recodes column names.
  - o Corresponding cells in the data table are selected when their entries are selected in the Recode dialog box.
  - o New option to replace entire string without a regular expression.
  - o Ability to apply mapping from another table.
- New Value Order column property allows flexible, customizable sort orders.
- Virtual Join supports multiple links from a single column.
- Graph Builder enhancements:
  - o New customization options to enhance Heatmap and Box Plot graphs.
  - o Improved Contour element with additional options for lines between contours and alpha hull support.
  - o Bagplot (or 2D Box Plot) for continuous XY data added to the Contour element.
  - o Frequency option in column functions.
  - o New Interval > Band option allows band from low value to high value.

- o Time Series Forecast available in Line of Fit graph.
- o New Line of Fit regression line option for Robust Cauchy.
- o Ability to save prediction formula(s) for Line of Fit and save formula for Bar, Points and Line.
- o Upgraded Variability chart and Oneway Axis.
- o Summary statistic property added for Color and Size drop zones in elements that summarize color or size, specifically Points, Line, Bar, Treemap, Heatmap, Box Plot and Contour.
- o More control over the basis for percentages, allowing 100% stacked bars charts without data prep. Now distinguish percent by X, Y or X+Y.
- o New Ridgeline Plots use histogram with a categorical variable for Y.
- o Stacking and fill options added to Line element.

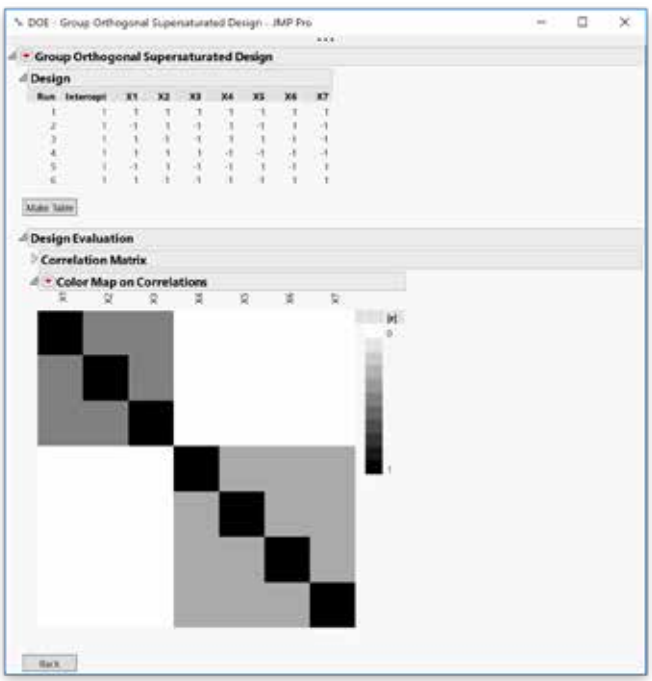


New ridgeline plots use histogram with a categorical variable for Y.

- Local data filters now include a histogram in the column filter control.
- New hover label options extend information available in graphs:
  - o Graphlets display a graph created by a JMP platform inside a JMP hover label. These contextual graphs add the ability to further drill down into the data without returning to the data table or launching a new platform.
  - o Textlets in hover labels expand options for the appearance of label text.
  - o Gridlets in hover labels offer options for formatting numeric values.
- Enhanced Capability Analysis report in Distribution platform (now Process Capability).
- Help menu now navigates to documentation on jmp.com for smaller software installs and more frequent Help updates.
- Support for Dark Mode in macOS.

Design of Experiments

- Group Orthogonal Supersaturated Design is a new platform for creating and analyzing supersaturated designs (more factors than runs), having groups of correlated effects. Should confounding persist after the initial analysis, the factors involved are identified and an appropriate augmentation scheme (requiring only a few runs) is advised.
- Instantaneous creation of a globally optimal design when runs are multiple of four and all factors are two-level continuous or categorical.
- Custom Designer provides modeling scripts for Generalized Regression and Fit Mixed. PRO



Group Orthogonal Supersaturated Design is a new platform to create and analyze supersaturated designs having groups of correlated effects.

Statistics, Predictive Modeling and Data Mining

- Explore Patterns platform investigates data integrity by looking for abnormal features in data including long runs, duplicate long sequences and unusually formatted values.
- Generalized Regression enhancements provide one-click relaunch options and the flexibility of changing distributions without relaunching.
- Enhancements to the Explore Outliers and Explore Missing Values platforms.
- Multiple frame contour profilers can be viewed simultaneously.
- New Model Driven Multivariate Control Chart (MDMCC)

platform allows users to monitor large amounts of highly correlated processes. The platform can be used in conjunction with the PCA and PLS platforms to monitor multivariate process variation over time, give advanced warnings of process shifts and suggest probable causes of process changes.

- Functional Data Explorer (FDE) has workflow improvements to bypass intermediate table creation, reshaping and joins. This allows functional response DOE, Custom Designer, FDE and Generalized Regression to all work together for a more convenient and robust workflow. **PRO**
- Improvements to Make Validation Column offer more control and an improved interface. **PRO**
- Structural Equation Modeling, a flexible new predictive modeling platform, allows you to model relationships among both observed and unobserved (latent) variables. **PRO**
- Support Vector Machines platform rounds out the JMP Pro modeling toolbox. **PRO**

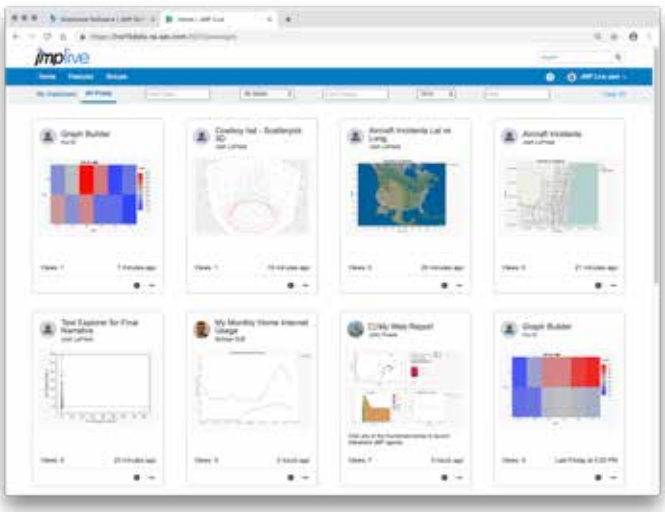
Improvements to Functional Data Explorer let users bypass intermediate table creation, reshaping and joins for a more convenient and robust workflow.

Automation and Scripting

- New Copy Columns menu item copies to the clipboard a script to create the column using JSL.
- Table script option, Copy Table Script (No Data), copies the table template without data.
- Scripts generated by operations in the Tables menu can now be saved to the source table.
- Generate JSL for Where clauses via point-and-click.
- Integration with SAS Cloud Analytic Services (CAS) server provides a simpler scripting mechanism, better authentication/authorization control, easier data set retrieval and the ability to run CAS actions from JMP to take advantage of high-performance, large cloud environments built on top of AWS, Azure or OpenStack.

Data Sharing

HTML5 reports include additional profiler support and connection to JMP Live.



Publish and share discoveries from JMP 15 to JMP Live, a private site for collaboration and interactive decision making.

JMP® Live

Publish to JMP Live – the newest member of the JMP product family. JMP Live takes the robust statistics and visualizations in JMP and extends them to a secure online portal – with data filtering and column switchers for interactive collaboration with colleagues.

Online

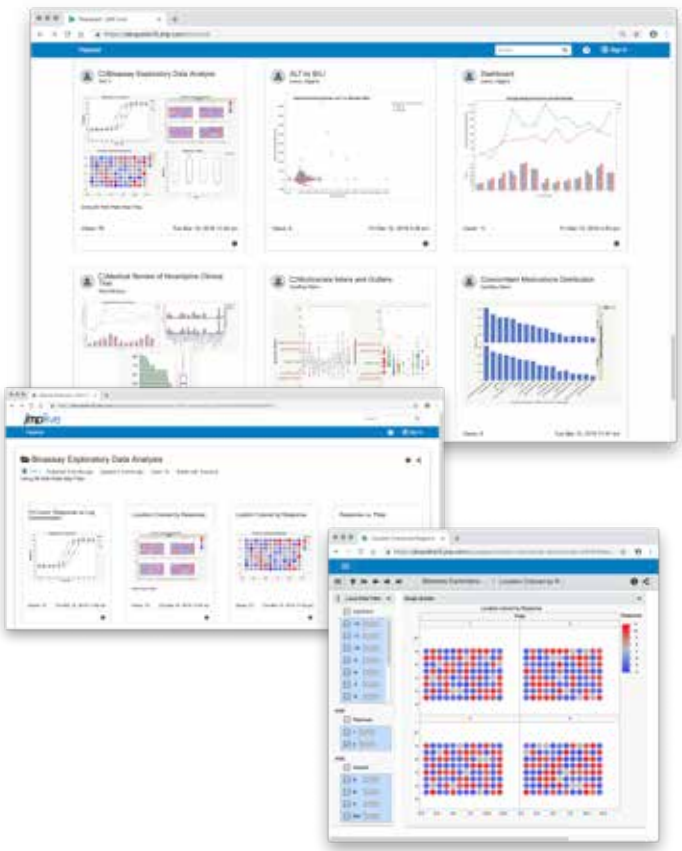
Play with JMP! You can try it for 30 days for free:
jmp.com/try

If you need answers from JMP experts around the world or want to share and download add-ins, scripts and sample data, join our online community:
community.jmp.com

JMP® Live: Discovery, Delivered.

Share and collaborate on data analysis projects across your organization

Scientists, engineers and other data analysts know the value of contributing their insights as those findings emerge. Silos of information lead to inefficiencies in organizations, and the absence of shared knowledge hinders solid decision making. JMP Live is the online platform where organizations share common organizational knowledge and access analysis conclusions. With JMP Live, a JMP user publishes discoveries to the JMP Live repository of JMP reports. JMP Live users can interact with those published reports privately and securely, even without a desktop version of JMP.



Discoveries are meant to be shared, so JMP Live extends the interactive capabilities of JMP to the web, reducing the amount of time from discovery to delivery.

JMP® Live makes it possible to:

- Distribute interactive JMP reports in a single click for informing and decision making instead of emailing tables or using shared folders.
- Make JMP reports available to explore on the web (with functioning local data filters and column switchers).
- Publish reports that are created regularly, so that report viewers always see the most up-to-date data.
- Save data and reports to a central repository.
- Surface decision alternatives to managers, who may not be JMP users.
- Automate routine but time-consuming data preparation without the typical pains of access administration.

JMP Live fits naturally into the JMP data analysis workflow. After exploring and visualizing your data, you may want to share your discoveries with others. Rather than attaching files to an email, converting reports to a static presentation or requesting report hosting by IT, you can publish reports directly to JMP Live to share with collaborators.

Why JMP® Live?

Organizations understand the value of sharing data between stakeholders. Ongoing efforts to transition the enterprise mindset from "my" data to "our" data can grow the impact of data and the conclusions drawn from it across the organization, inform collaboration, increase efficiency and provide support for decision making in other business areas. JMP Live can maximize these efforts by making it possible for those who share data to:

- **Access a single source of truth.** Most organizations have data in many locations. JMP Live provides a central location of managed reports that gives stakeholders the latest details from one source.

- **Focus on the analysis.** Instead of spending time converting JMP graphs to a format to allow sharing with non-users of JMP, JMP Live gives JMP users the freedom to concentrate on the tasks that move them toward impactful findings or integrate automated publishing of findings to their existing JMP Scripting Language (JSL) scripts.

- **Make informed decisions.** Sharing JMP tables, graphs and interactive HTML reports on other cloud-based file sharing tools is possible but lacks the interactivity of working data filters and the organizational abilities of a purpose-built platform like JMP Live. This compelling interactivity and access control gives teams the confidence to make solid, timely decisions.

JMP® Live in action

With a few clicks, JMP users can publish their reports online to share with collaborators. Immediately, JMP Live users are able to interact with these reports using exploratory data analysis tools such as local data filter or column switcher for another view of the data in seconds. These capabilities make it possible to share interactive reports, manage content and automate report publishing.

Share interactive reports

Reports you publish to JMP Live are dynamic and interactive. Because JMP Live can use the JMP analytics engine, these interactive reports allow you to drill-down into subsets of the original data using exploratory data tools. You can fit updated models right in JMP Live by selecting subsets of the data or different columns, allowing the data to answer any new questions that arise while you view the reports. And, with JMP Live, your report is just a link away. Which means it's easy to collaborate with others in your organization who can then explore your findings for themselves. Additionally, ad hoc visualizations and discoveries are easily shared by using point-and-click publishing tools. It's easy for JMP users to publish one-off reports and notify consumers that it's available in JMP Live.

Publishing the results of experiment design help with "institutional memory." In JMP Live, you can easily see the results of a design of experiment (DOE), the model that was fit and interact with the Profiler. Easily pull the results down from JMP Live in the form of a JMP table if follow-up analysis is needed. For projects that require multiple steps by multiple individuals (for example, if one person is responsible for data prep and another for modeling), creating a group for sharing with a specific subset of collaborators and publishing a data table to JMP Live is a great way to hand off the analytics project between team members while tracking each step in the analytic workflow.

Automate report publishing

For daily analysis and reporting, JMP Live can be incorporated directly into your automated analytic workflow, so a few clicks can take you from analysis to publish. When new data is available, you can easily update the applicable reports on JMP Live so content is always up to date. By automating the publishing of these reports and making them available to the team, better collaboration and communication is afforded.

With JMP Live and publishing features from JMP, you can share discoveries without needing to redo or rework dashboards or results just to put them on the web. For those reports that just need to be consumed and acted upon, having a regularly published JMP Live report is much easier than passing around scripts, controlling database access or distributing JMP add-ins.

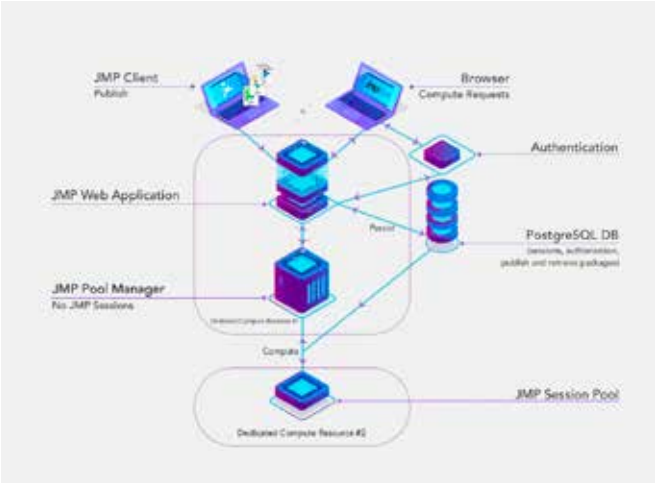
Manage content

JMP Live is a content management framework for your JMP output and supporting files. Easily add or change titles, descriptions or thumbnails, so you and your collaborators can quickly find what you are looking for. With JMP Live as a content management system and easy repository to publish and share results and data, your team can move toward managing JMP reports in one location without an individual scientist or engineer losing the agility or autonomy afforded by JMP. JMP Live lets you collaborate when you need to and control who sees your reports, without relying on IT resources to run reports or manage servers. JMP Live offers the capability to publish analyses to a central location or embed reports in web pages for quick access – at the moment you're ready to share. Plus, when data changes, you can replace reports quickly or automate the publishing of routine reports to provide the latest details from one source of truth.

Collaboration across the organization

With JMP Live, share discoveries using simple report publishing that fits seamlessly into the JMP analysis workflow:

- **Explore** data with JMP on the desktop and develop visualizations that reveal complex relationships.
- **Publish** these saved visualizations to JMP Live in a few quick steps.
- **Share** access with others across the organization, allowing them to interact with the visualizations. Consumers can use local data filters and column switchers to explore the data themselves, recalculating results based on various conditions.
- **Enrich** the impact of shared analyses by automating report publishing – for the most current results in JMP Live, all the time.



JMP Live extends sharing capabilities for richer collaboration.

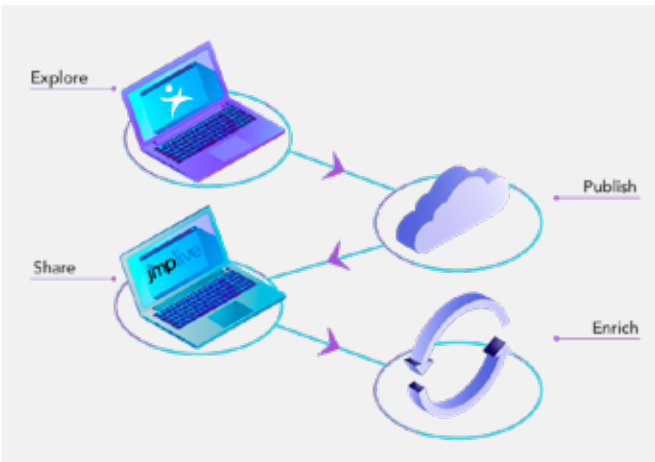
Deployment and administration

JMP Live offers flexible options for deployment in your organization.

- Integrate JMP Live with your existing security protocols for authentication at the user and group levels. JMP Live supports managed resource access for most directory services, including Active Directory and OpenLDAP, or for cloud implementations, Okta, Azure Directory Services, AWS Directory Services and SAS profile. JMP Live also supports use with a single sign-on provider for enhanced security and ease of use for end users.

- Incorporate JMP Live into your current framework, no matter how you store and access your data. Install JMP Live on-premises with your current data infrastructure or use a public cloud platform (Amazon Web Services, Microsoft Azure or Google Cloud Platform).

- Choose from scalable architecture options based on the number of users, size of reports or frequency of access in your organization.



JMP Live has scalable architecture options, including the most common, standard, medium.

Whether you are an IT expert or embarking on the journey of setting up JMP Live for your organization without a lot of IT training, there are guides available to help you get JMP Live up and running.

JMP® Live offerings

JMP Live can be hosted by JMP or your organization.

JMP Live Hosted

- Installed and maintained by JMP.
- Results published and shared from JMP on the desktop.
- Unlimited reports.
- Built-in security and authentication.
- Automated publishing using JMP Scripting Language (JSL).
- User licenses to view JMP reports.

JMP Live On-Premises

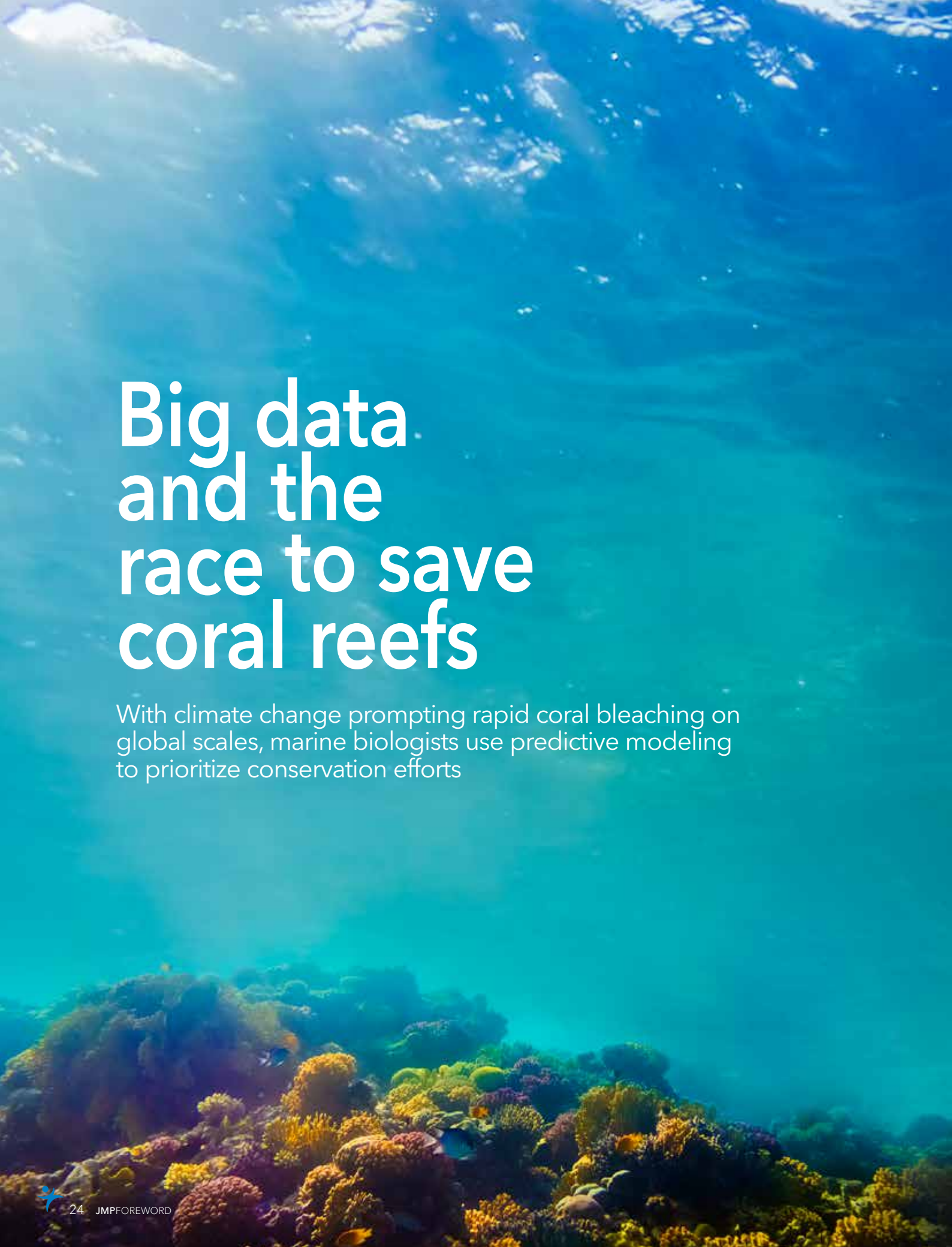
- Installed and maintained by your IT staff.
- Results published and shared from JMP on the desktop.
- Unlimited reports.
- Built-in security and authentication.
- Automated publishing using JMP Scripting Language (JSL).
- User licenses to view JMP reports.
- No additional hosting fees.

Online

Get all the details about collaborative analytics software from JMP:
jmp.com/live

Take a look at the public version of JMP Live, a website using the same technology as JMP Live, where anyone can share JMP content with the public:
public.jmp.com

Access complete deployment and administration details:
jmp.com/liveadmin



Big data and the race to save coral reefs

With climate change prompting rapid coral bleaching on global scales, marine biologists use predictive modeling to prioritize conservation efforts

Challenge

With the rapid advance of climate change, coral reef health is in decline globally. Scientists are working against the clock to optimize reef preservation efforts and prevent further degradation and bleaching.

Solution

By increasing the level of statistical sophistication in their research, marine biologists can draw more substantive conclusions from reef data sets. JMP makes statistical modeling less daunting for non-statisticians.

Results

Through enhanced collaboration and data transparency, scientists are closer to developing new approaches aimed at best driving future coral reef conservation efforts.

Though coral reefs constitute less than 2% of the Earth's oceans, they play a critical part in ocean health and are among the most diverse ecosystems on the planet. Not only are they home to scores of marine species, but their immense impact extends to local economies, supporting fisheries, encouraging tourism, and protecting shorelines from erosion – an impact valued at billions of dollars each year. With an outsized role in both environmental and economic success, protecting these fragile reefs should be a global priority, yet coral reef health around the world continues to decline. Climate change, in particular, has accelerated the degradation of reefs, and scientists now estimate that more than half of the world's coral reefs have been lost over the past four decades due to climate change and other human-associated factors. Although scientists broadly understand the causes of coral reef bleaching and (later) death, predicting specifically which reefs are most at risk, which are most resilient, and which should be prioritized for conservation continues to be among the field's foremost challenges.

Organizations like the Khaled bin Sultan Living Oceans Foundation seek answers to these questions, and marine scientists the world over are investing in international coral reef research, conservation and restoration efforts. With their support, Anderson Mayfield, PhD, has devoted his career to the study of reef coral physiology, arguing that scientists need a more rigorous, statistically driven approach if they are to understand – and mitigate – the breakdown of corals in response to environmental shifts before it is too late.

Statistical approaches speed research

In the field of marine biology – where so much is hidden from casual view – scientists' first challenge is to observe and understand how many corals are out there, where they're located, how healthy they are, and the risk factors they face.

Documenting changes in reef health over time is paramount in efforts to understand how this fragile ecosystem is being affected by climate change-associated factors like increases in seawater temperature and acidity.

But with the immediacy of the threat now facing coral reef ecosystems around the world, Mayfield – an assistant scientist at the National Oceanic and Atmospheric Administration's Atlantic Oceanographic and Meteorological Laboratory in Miami – believes there is a need to be more proactive about marine science, and especially marine conservation efforts. That's why he and his collaborators are making the case for a statistical approach. Predictive modeling, he says, can help to determine which reefs are more susceptible to environmental stress – and which are most resilient – thereby enabling conservation funding agencies to prioritize targeted mitigation efforts.

Healthy coral polyps rely on an endosymbiotic relationship with microscopic algae that live in their tissue. These algae give the coral their color and serve as their primary food source. When water temperatures rise, however, the symbiotic relationship comes under stress and the algae begin to leave, causing the coral to turn white or "bleach" and starve.

"Some corals will bleach regardless," Mayfield explains. "But there are other reefs where if we alleviate some of the local pressures, they're going to be more likely to survive." Although it's difficult to decide which reefs are most worth protecting, some, like a series of reefs in Southern Taiwan, have shown high resiliency in the face of increasing environmental pressures. "This seems to me like a good candidate for a reef we should prioritize for conservation and research [to determine] why it's so resilient," he says.

"Predictive modeling can help to determine which reefs are more susceptible to environmental stress – and which are most resilient – thereby enabling conservation funding agencies to prioritize targeted mitigation efforts."

Anderson Mayfield

Identifying critical survival factors may allow researchers like Mayfield to better predict which reefs are more prone to stress and thwart future bleaching elsewhere.

Boosting the statistical power of scientific conclusions

Mayfield's statistical models represent a significant departure from his previous reliance on experimentation with coral conspecifics grown in laboratory tanks where he says results were not sufficiently statistically powerful to account for natural species variation. By contrast, "big data" statistical modeling is enabling him to make use of worldwide data surveys (such as those undertaken on the recently completed Living Oceans Foundation "Global Reef Expedition," in which he was a participant) that are both deeper and more geographically expansive than data amassed by any individual scientist.

For example, in recent papers published in Platax (2018) and Journal of Sea Research (2019), Mayfield and his co-authors looked at data sets from understudied regions of Fiji's Lau Archipelago and the deep South Pacific (Austral Islands of French Polynesia and the Cook Islands), respectively, using a combination of univariate and multivariate methods. Critically, the researchers explored 12 environmental factors (e.g., temperature, reef structure, fish biomass) expected to influence coral physiology. Best-fit models produced by stepwise regression and partial least squares showed that only a subset of such routinely assessed environmental parameters were needed to explain a significant portion of the variation in physiological response. Though several of the models in the studies were found to be of relatively low predictive capacity, Mayfield urges that the method is nonetheless a proof of concept and showcases the idea that it is worth attempting to use previously collected data to make predictions about marine animal health.

From a graphing calculator to JMP®

When Mayfield began his training in marine biology first at Duke University and then at the University of Hawaii, Manoa, he was far from the outspoken proponent of statistical methods he is today. Having not yet been exposed to formal multivariate statistics in any real way – and because the variables in the tank experiments of his early career were largely known and tightly controlled – Mayfield jokes that his analysis was limited to what could be done on a graphing calculator (e.g., one-way ANOVA).

Mayfield's reliance on rudimentary tools like calculators and Microsoft Excel, however, had to change as his research expanded into more field-based projects for which a new level of statistical sophistication was necessary. Fortunately, it was around that time that Mayfield also got started with JMP statistical discovery software.

Now he won't use anything else.

"At first, I just used JMP to look at distributions and to carry out simple comparative analyses: t-tests, ANOVAs, linear regression and such. I probably took advantage of 5% or less of what JMP could do," he recalls. Eventually, however, Mayfield resolved to pursue the other 95% – a move he credits with not only changing the way he analyzes data but also how those data are collected.

"Even when I'm out on the reef, I'm seeing JMP tables and figures, thinking about how I can get the most information out of a sample," he says. Knowing the statistics and the capabilities available to him through JMP, Mayfield says he can make strategic decisions about where to focus his attention and which samples to collect while he's underwater. "These are the kinds of things that I can directly test with JMP ahead of time," he says.

Furthermore, Mayfield contends that with JMP, "even with the same amount of data, the amount [of information] it tells us has expanded exponentially." Bringing the statistics in-house also limits his reliance on external statistical resources, which inevitably add more time to the research process. And time is a precious resource in light of the growing crisis of coral bleaching. "We don't really have the luxury to sit around for five years and think about how to analyze the data," he explains.

From exploration to prediction

Having taken the time to deepen his modeling skills within JMP, Mayfield is empowered to be creative with his data sets and devise a statistical approach that extracts previously hidden information. For ideas, he now looks to how other companies and institutions handle their large data sets. Analysts in other fields routinely use behavioral data to develop predictive algorithms about, for example, shopping preferences or maintenance schedules – why couldn't the same principles be applied to coral reef behaviors and their survival?

"We may not be able to predict with 100% certainty whether a coral is going to die," Mayfield theorizes, but "if we have enough data already about how corals behaved in the past ... we might say based on our past data that a one-degree [seawater] temperature increase results in a 30% change in coral growth."

Potentially, he and his collaborators can use the worldwide data survey developed as part of the Living Oceans Foundation's Global Reef Expedition to make the transition from explanation to prediction. "These are the kinds of things I want to explore now that I know more about what JMP can offer," Mayfield says. He, however, was quick to echo a key message gleaned from the pioneering work of Galit Shmueli (2011) in that, just because a data set excels at explaining past observations, it does not ipso facto signify that this same data set will allow one to make predictions on the future behavior of the target animals. This is not to say that explanatory and predictive capacity are never in sync; just that one must validate this relationship empirically.

Though still a nascent idea, Mayfield is excited to explore the opportunities big data affords in the field of reef coral physiology. With new models developed in JMP, he can test their predictive power with an extensive field data set. "We won't know until we test [a model]," he says, but luckily, JMP has in-built platforms that support this type of model accuracy testing; when you build your models, JMP will tell you whether your model worked for each sample. "JMP's predictive modeling platforms already have the pieces of the puzzle in place; they're just waiting for this kind of data to come so that we can play around with it."

Mayfield says he's taken advantage of not only the modeling features of JMP, but also the neural networking and outlier analysis platforms. Regarding the latter, rather than throw out data from corals displaying aberrant behavior (i.e., the outliers), he argues that, in many cases, these oddly functioning individu-

als – which tend to be the most stress-resistant in his data sets – might actually be those of most interest to physiologists. In other words, such features of JMP allow Mayfield to identify the coral biopsies he should prioritize under the real-world limitation that, in many cases, there are insufficient funds to analyze every biopsy to its fullest capacity.

Accelerating research through data transparency

As Mayfield is painfully aware, copious data languish unused by researchers who lack either the time or statistical knowledge to analyze them. Moreover, given the pressure to publish and competition between scientists, most are reluctant to share their data (even data they have already used in published works). With such a time-sensitive problem like coral reef bleaching, though, data sharing could be a key factor in accelerating the state of the science and determining the most effective conservation strategies. "Especially in our field, [scientists] should be more open about how they analyze the data and make real efforts to get it out there. It's not doing any good just sitting on a computer," he says.

And Mayfield leads by example – JMP gives him the tools for promoting transparency in his own research, such as interactive HTML features that allow for real-time data analyses. "JMP has given me so many ideas – new ways of making my data more interactive and not just in terms of visualization but showing people the thought process behind how I analyzed it," he says. In addition to posting data sets to his personal website coralreefdiagnostics.com, Mayfield now uses JMP Public as a means of disseminating dynamic visualizations. The software's data filtering capacity, he says, permits his collaborators and other interested individuals to "find" corals demonstrating certain characteristics in the data set – a bleached coral from Fiji that was over-expressing fluorescent proteins, for example. This feature in JMP Public will allow both present and future collaborators to quickly identify samples they may wish to analyze in more detail, be they archived biopsies or the GPS coordinates of the originally sampled coral colonies themselves.

"Most people will fall asleep if you give them a one-hour lecture on statistics, but maybe if they could just see how you went from this data table to this figure... viewers would then be more confident in your data." Mayfield hopes that by disclosing the details of his own research, and by boosting the reproducibility of his methods, he can encourage colleagues toward similar transparency, bringing the field together in pursuit of tangible solutions to coral reef decline.

"We might already have all the data we need to address certain coral reef issues," says Mayfield. "We just aren't analyzing them the right way." JMP could change that. And now, Mayfield's research is only limited by the questions he asks.

References

Mayfield AB, Dempsey AC, Inamdar J, Chen CS (2018), A statistical platform for assessing coral health in an era of changing global climate-I: A case study from Fiji's Lau Archipelago. Platax 15, 1-35.

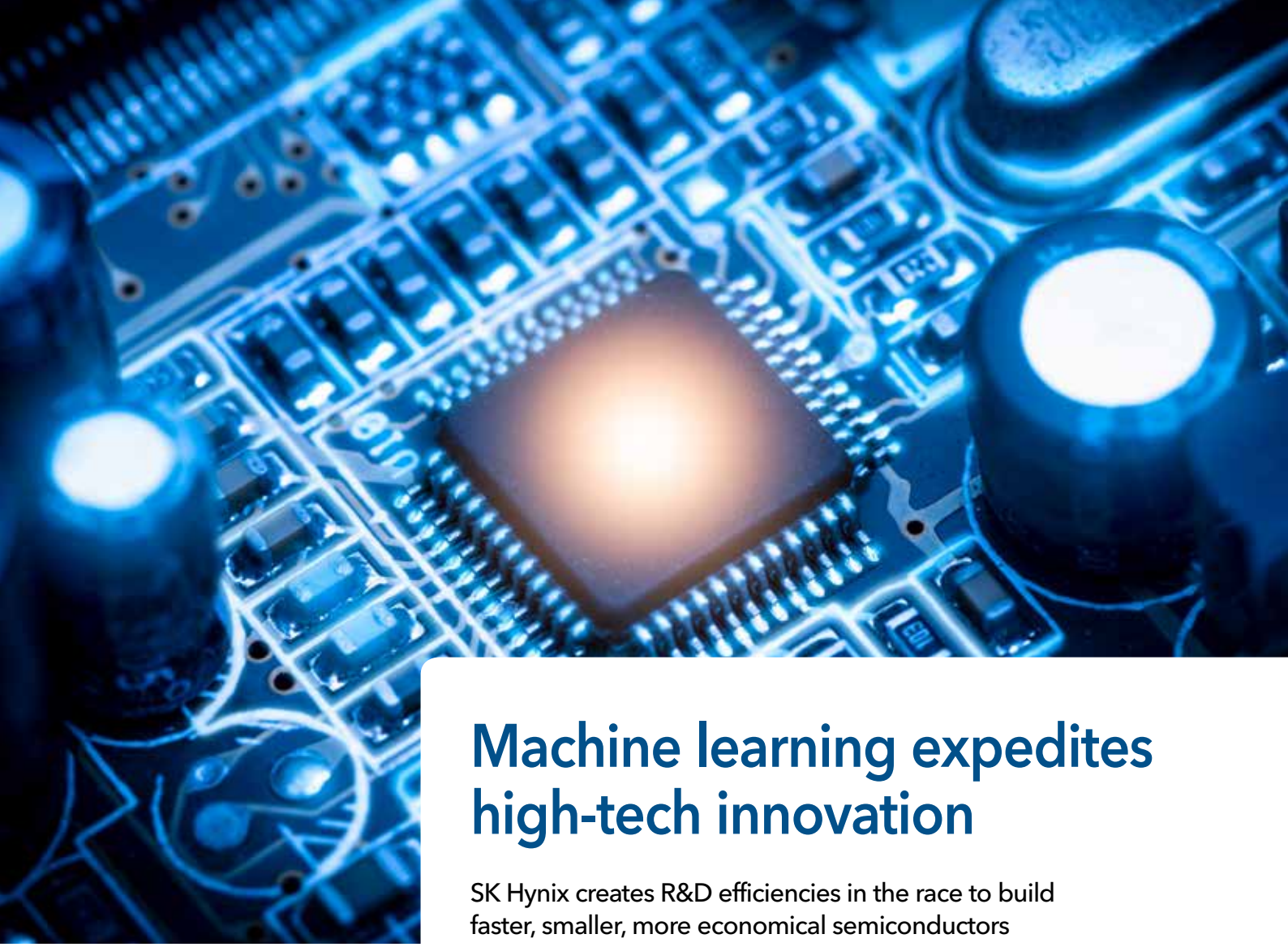
Mayfield AB, Chen CS, Dempsey AC (2019), Modeling environmentally mediated variation in reef coral physiology. Journal of Sea Research 145, 44-54.

Shmueli G (2011), To explain or to predict? Statistical Science 25(3), 289-310.

Online

Anderson Mayfield has shared some of his coral reef data and discoveries. Explore it on JMP Public: jmp.com/coral

The Khaled bin Sultan Living Oceans Foundation provides science-based solutions to protect and restore ocean health: livingoceansfoundation.org



Machine learning expedites high-tech innovation

SK Hynix creates R&D efficiencies in the race to build faster, smaller, more economical semiconductors

Challenge

Develop the know-how and data infrastructure to help design of experiments processes that will increase automation, create time savings and produce efficiencies in semiconductor R&D.

Solution

Use JMP to create machine learning algorithms that can expedite the use of data in manufacturing processes. Standardize the use of JMP to scale up efficiencies that have the potential to affect the whole organization.

Results

In addition to widespread time savings, SK Hynix has realized intra-company synergies with the rollout of new methods and a software toolkit that starts with JMP. The further development of proprietary technologies has expanded the potential impact of machine learning-based operations.

In the semiconductor field, much of the rapid innovation that characterizes the industry is driven by fierce competition. Players are embroiled in a constant race to develop faster, smaller, more economical parts that has led to a doubling in the number of transistors in a dense integrated circuit approximately every two years. That's great news for buyers but can mean tight margins and an intense emphasis on research and development for market-makers.

Data analytics has long played a part in this expedient growth, as researchers look to gain synergies in every step of the development and manufacturing process. So far, data has been most useful in the testing phase, but as technology advances, it is being deployed earlier in production. And SK Hynix, the world's third-largest producer of semiconductors, has been on the forefront of this development.

"With such an abundance of data, we can perform a huge variety of types of analysis," explains Yonghan Ju, head of the Data Innovations Team under the Data Science Section at SK Hynix in South Korea, where his team mainly focuses on analytics for R&D and the automation of SK Hynix's manufacturing environment. Along with his colleagues, Ju is tasked with resolving some of the organization's most complex statistical problems, sometimes using technology to produce an automated process solution. "A broad range of areas can be targeted regarding statistics and data analysis, from

"On the R&D side, we normally use data from experiments, and in many cases we can outline [experiments] using a general statistics methodology. Better yet, we can use machine learning and start changing the technology itself."

Yonghan Ju

conventional SPC analysis to deep learning and optimization," he says.

Data visualization plays vital role

SK Hynix is undoubtedly one of the most important players in the global semiconductor industry with factories in South Korea and China – and a history of technological leadership dating back to 1983. Today SK Hynix products are widely used in many popular IT devices. SK Hynix's impressive list of customers is testament to its R&D methods, as well as its capacity for shift innovation. Ju's team members are responsible for developing the data side of this process to ensure consistent progress.

"We mainly receive data from the data engineering team or on-site organizations for analysis. Such data is related to optimal process conditions or the yields of equipment and products," explains Ju. He and his team work to develop new statistical models for testing that will streamline experimentation depending on the variables being measured. Data visualization can play a vital role in developing a suitable approach to test design. "It is difficult to evaluate the fitted statistical model with only the coefficient of determination or p-value," Ju explains. "The outlier effect is a representative example. In addition, visualization tools are highly necessary for more efficient communication with on-site engineers. I believe that JMP is an intuitive and fast tool for this purpose. However, it is still difficult to visualize data in three or more dimensions."

Machine learning in data analytics

These processes can be challenging given the amount of data involved, and its quality; incomplete data sets pose a considerable obstacle to successful analysis. "In the case of very big data, a data set may sometimes be impossible to analyze on a PC," he explains. "In addition, problems may arise as data is sporadically loaded. Therefore, we are attempting to manage data in an integrated manner this year by means of a concept like data lake. We are also expanding our GPU servers to perform models such as deep learning."

Full automation is increasingly a defining trend in contemporary manufacturing. Ju and his team are working to develop solutions that will allow for machine learning to expedite SK Hynix's R&D processes as well as production. "The semiconductor industry produces a lot of data," he says. "On the R&D side, we normally use data from experiments, and in many cases we can outline [these experiments] using a general statistics methodology. Better yet, we can use machine learning and start changing the technology itself." He describes developing algorithms to help clean data sets and remove noise without compromising the quality of an experiment. "I feel that machine learning algorithms are more fault-proof than a statistical approach in some particular areas. This is a very good opportunity in data science."

JMP has been particularly important when testing for practical range, limitations and endurance, according to Ju. "In this situation, there are a lot of experimental variables," he adds. "We may have 10 or 20 variables for which we need to consider interaction effects. [Traditionally, engineers] did not use statistics [to guide experiments] because a lot of it dealt only with relatively simple data sets. But when we began using JMP for DOE, we were able to dramatically reduce the required sample size of the experiments so that we could apply appropriate statistical analyses. This has really improved time spent on analysis."

JMP® is shaping the technology landscape

For Ju and his team, it is important to be able to communicate experimentation methods and findings with others within SK Hynix – particularly non-data-scientists – and the visualization tools in JMP can be invaluable in facilitating conversations. "Perceptions about data analysis have changed significantly," he explains. "Data analysis is no longer the exclusive realm of data scientists, but also provides essential backup materials for decision making in actual development and manufacturing conditions. This enhanced awareness of the importance of data analysis in big data environments, especially among leaders, has led to the uptake of JMP site licenses."

Ju heads only one of a number of divisions within SK Hynix to license JMP. And he says the company as a whole has prioritized JMP over other data analytics software packages because of its intuitive connectivity facilities and

rich capabilities in predictive modeling and machine learning methods. Though JMP is increasingly becoming a software of choice at SK Hynix, Ju recalls a time when different divisions used different applications and there was no unified process: "The data structure was changing," he explains. By contrast, with JMP, "most SK members can connect [to the data] more easily than before. So through data visualization, they can perform tasks and see results."

Today, JMP is such a vital part of SK Hynix's R&D process that new employees are given training in its use as a matter of course, Ju says. "We teach them JMP and the statistical method. The environment is very good." And for an industry leader like SK Hynix, this kind of environment is exactly what is shaping the technology landscape not only in Korea, but in the world beyond.

Online

Using smart machine learning in modern manufacturing: jmp.com/smart

A leader in the global semiconductor industry, SK Hynix innovates chips that make your digital devices smaller, faster and cheaper: skhynix.com

Analytics for the win

How the losingest team in baseball found its way to the World Series

Challenge

Fulfill a long-shot prophecy. When in 2014 Sports Illustrated projected a 2017 Astros victory in the World Series, few believed the worst team in the American League was up to the task.

Solution

An analytically minded front office working in tandem with astute scouting talent to discover practical insights; a forward-thinking coaching staff prepared to embrace data in on-field decision making.

Results

The Houston Astros went from losing more than 100 games for three years in a row to being crowned World Series champions, proving that analytics, no longer a novelty, is now "table stakes" in baseball.

Pitcher Charlie Morton was on the mound the moment the Houston Astros clenched their first-ever World Series title with a 5-1 victory over the LA Dodgers in Game 7. It was something of a poetic ending to the season in which both team and player rose from the ashes of back-to-back losing streaks.

Following a 2016 season with only four starts, a mediocre ERA and a series of injuries that had him on the bench, the Philadelphia Phillies cut Morton loose. At 32, it looked like the death knell of his MLB career. But when the Astros' front office noticed a velocity jump in Morton's pitching, they signed him to a two-year, $14 million contract. One year later, he closed out Game 7 of the World Series.

The potential the Astros saw in Morton – and how the team climbed their way from the worst record in baseball to winning the championship – was based largely on insights derived from novel statistical analysis techniques.

Creating decision tools for decision makers

Though many in 2014 saw the Sports Illustrated projection that the last-place Houston Astros would win the 2017 World Series as greater than a long shot, the Astros' front office was confident in what was at the time an unconventional strategy. That strategy hinged on the work of former NASA engineer turned sabermetrician Sig Mejdal, who first walked into the offices at Minute Maid Park in 2012.

If you're going to try to build a championship team from scratch, "you want to figure out how to take advantage of what's discoverable and acquirable," says

Mejdal, then the Astros' Director of Decision Sciences. Now Special Assistant to the GM, Mejdal describes his remit as improving the processes by which the Astros' decision makers – scouts, coaches, even players – apply the findings of the team's analytics group. Often in exploring the wealth of data the team has now amassed, analysts can see potential in a player that traditional scouting may have missed or share coachable insights with decision makers in the bullpen.

"Analytics permeates the entire [Astros] organization," Mejdal says, "the decision to draft a particular player, the instructions he gets in player development, the decision to promote him, how he attacks a specific pitcher, where he stands on defense – analytics is a part of all of that. It's impossible to separate analytics from what we do."

Analytics, no longer a novelty, is now 'table stakes' in baseball

Baseball has always been a game in which stats mattered. But Mejdal says that as the means of collecting information have gotten better and better, data has become truly ubiquitous. For example, new technologies like radar devices can now measure the release point, speed, movement, spin and location of every pitch – and record the same for batted balls. When combined with traditional scoring methods, these technologies provide teams with a vast number of data points. "We can get information on a center fielder's reaction time, his ability to accelerate, the efficiency of the path he takes to track down the ball, his top speed, and of course, whether he made the out or not," Mejdal explains.

In fact, every major league team in America now has at least one quantitative analyst – someone like Mejdal working behind the scenes to interrogate the numbers. Whereas in the beginning of his career in sabermetrics, nuggets of statistical insight in any amount were enough to give an analytically savvy team the edge over a data-naïve competitor, quantitative analytics in baseball today is just "table stakes," Mejdal explains. "When there were only four analysts in baseball, it was easy to stay ahead of the curve – the inefficiencies were giant. But now, like in any mature industry, the inefficiencies become smaller and harder to come by." Take all this readily available data and throw in capable sabermetrics programs at every front office around the country; what you get is as competitive a playing field as baseball has always been. But that doesn't mean there's not a competitive edge to be found.

A competitive edge through teamwork

When analytics is an arms race, as Mejdal describes it, the competitive advantage comes increasingly less from data access – or even the kinds of statistical insights you can glean from the data. Rather, teams that are most successful in their analytics programs are those that, like the Astros, have built a culture where individuals across all parts of the organization work together to take advantage of the data they're given.

"I don't think what we are discovering is too different from what others have already or soon will discover," Mejdal says. "Where I think we (the Astros) do very well is in taking advantage of the findings. From the GM to the manager to the coaches to the scouts to the players – they all are embracing analytics and the efforts to do something a bit better. It's incessant."

This harmony within the organization is what makes the Astros great. For one, the growth of decision sciences in the Astros' front office hasn't detracted from the team's reliance on expert valuations and the opinions of scouts and coaches; in other words, the art of baseball has not been eclipsed by the science. And coaches and players have been eager to experiment with the insights they receive from analysts.

Being the first to discover a pattern may still be a challenge but, Mejdal says, "also important in the formula is having a culture in which you're prepared to take advantage of the findings once you discover them. And that's what our GM, Jeff Luhnow, has been so good at creating."

Intuitiveness, capability and speed

More astute analytics programs like the Astros' front office understand that data insights go both ways; this is a key part of the culture Luhnow has fostered. Analysts listen to what the experts intuit and then help test the value of certain attributes. "Analysts have come into baseball, and what they're bringing in is a process to test hypotheses in a structured way," Mejdal explains; the scientific method is the strategy. And sabermetricians are positioned to apply that method – with the help of new technology and statistical tools – to investigate the merits of conventional baseball wisdom.

Having come into sports analytics in the field's early days, Mejdal had to build that toolkit for himself. "At the time, we had R and XL Stat," Mejdal recalls. So it came as a surprise when Mejdal got a call one day inviting him to speak at a conference hosted by statistical software company JMP. "I had not used JMP, I did not own JMP, and then when I got to the conference I could see the passion these users had with this software program and I couldn't believe it. That got me curious to try JMP, because frankly I didn't care for another off-the-shelf software program that I needed to learn. But then when I took a look, I was just wowed with the product….That combination of intuitiveness, capability and speed is hard to find."

In baseball, Mejdal says, data exploration is not only useful – it's mandatory. Analysts may not know exactly what they're looking for until they see a pattern in the data to investigate further. "Before you jump in with a specific statistical

"Analytics permeates the entire [Astros] organization, the decision to draft a particular player, the instructions he gets in player development, the decision to promote him, how he attacks a specific pitcher, where he stands on defense – analytics is a part of all of that. It's impossible to separate analytics from what we do."

Sig Mejdal

technique, it's a best practice to look at your data, explore it, visualize it and get a general understanding as best as you can of what you're working with. It helps you decide which statistical technique to use," he advises. "And [when it comes to data exploration,] I haven't seen anything that could compare with JMP Graph Builder. Just that ease of use, the flexibility, the capabilities. If you want to encourage data exploration, JMP is so inviting."

Out of the black box of analysis

Data visualization helps the end user too; in baseball, the front office's data insights are only as useful as the players and coaches make them. "We need to connect to the [on-field end users]. We're asking them to some degree to change their behavior. In my experience, that doesn't just come with saying 'Hey, I did some analysis. Here's a number. Use it.' Instead, the more you illustrate what's behind [the stats], connect an anecdote to it, the better off you'll be."

As they say, a picture is worth a thousand words; similarly, graphics help on-field experts understand and operationalize data insights. "The more you can remove [coaches and players] from the black box of analysis and toward an understanding – a visualization of what's going on – the better. And JMP is great at creating intuitive illustrations of your data quickly," Mejdal explains.

That the Astros would use a tool whose key value lies in its exploratory potential is no coincidence. Mejdal says that same curiosity is at the heart of Luhnow's vision for the team: to drive, question, improve and then take advantage of any insights they can glean. "JMP is an excellent tool with which to determine these relationships and predictive ability," Mejdal says. "Rarely does a day go by in which I don't use JMP."

Fast forward to Game 5 of the 2017 World Series. "It was the loudest crowd I have ever heard in Minute Maid park," Mejdal recalls. The pandemonium was set off by Astros third baseman Alex Bregman who, late into the game's fifth hour, put the torturous 13-12 game to an end with a walk-off single. Bregman was seen as a highly unconventional pick when the Astros drafted him in 2015. Mejdal says that at the time, the team's analysts noticed a number of attributes pointing to his potential success. That insight combined with scouts' significant interest in the player was enough to draft Bregman out of LSU. So, when two years later he became only the second player to drive in a run in each of his first five World Series games, the whole affair had something of a prescient shine to it. "It's wonderfully satisfying to have played a role in acquiring a player and then seeing him years later succeed," Mejdal says.

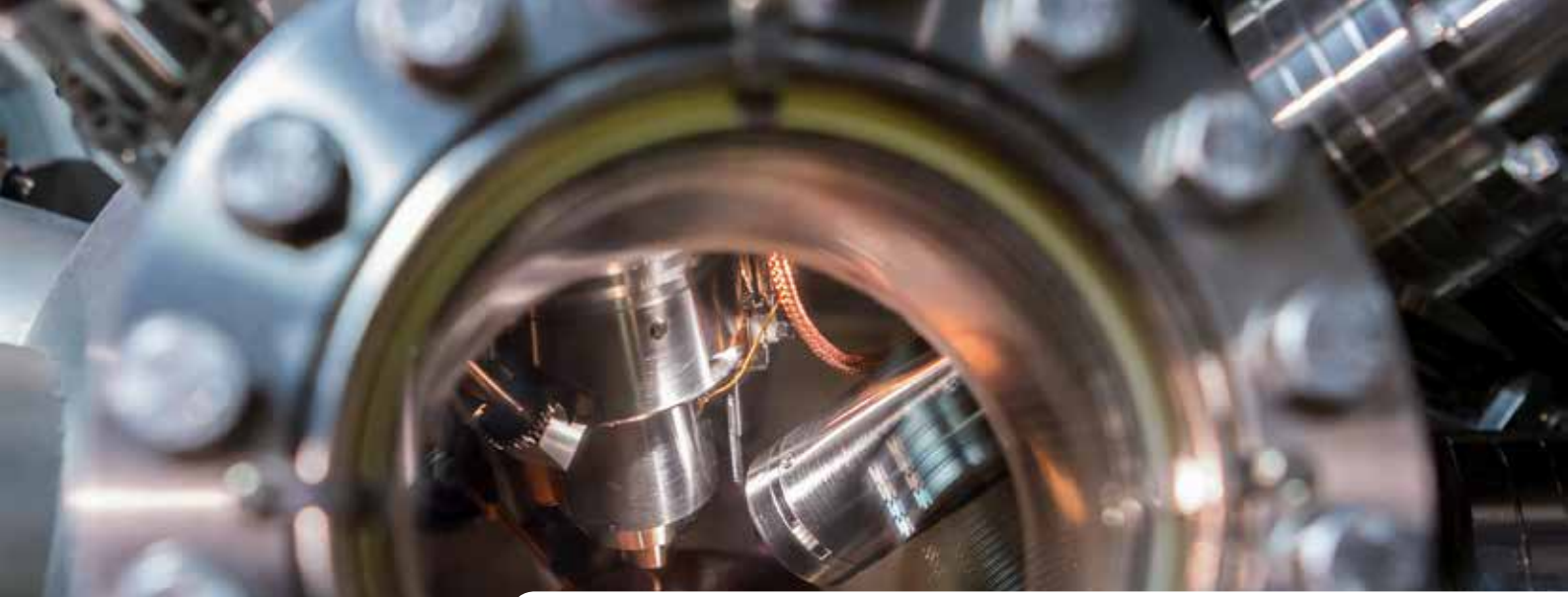
Even more satisfying, the Astros finished off the World Series with an infield shift – a move designed to defend against batters with a penchant for pulling hard right hits through the gaps between players. Seeing the team's infield move into this now ubiquitous defensive position, Mejdal took a moment to reflect on just how far they'd come. In earlier years, the Astros had been the first to take the shift to a level that had seemed almost laughable at the time. "I thought it was appropriate that the last out in Game 7 of the World Series was into the shift," says Mejdal. Yet further proof that analytics programs – at least in baseball – are here to stay.

Sig Mejdal is now Vice President and Assistant General Manager of the Baltimore Orioles.

Online

Sig Mejdal shares analytics lessons from the baseball field:
jmp.com/mejdal

Get the latest news from the Houston Astros:
astros.com



What do early-career scientists need to succeed?

The Master's Industrial Internship Program at the University of Oregon prepares graduates with a complete and relevant skillset

Challenge

Slews of talented young people graduate each year with bachelor's degrees in physics, chemistry and engineering, but for new graduates, the path to becoming a professional scientist is often unclear. And at a time when there are more jobs in applied science than ever before, even the industry's leading companies struggle to fill positions with candidates who are ready to hit the ground running.

Solution

The Master's Industrial Internship Program at the University of Oregon is designed to help students launch their careers through a combination of intensive experiential learning on campus and on-the-job internships with industry partners. Students focus on honing skills valued highly by the industry, from practicing design of experiments to mastering JMP software, an industry favorite.

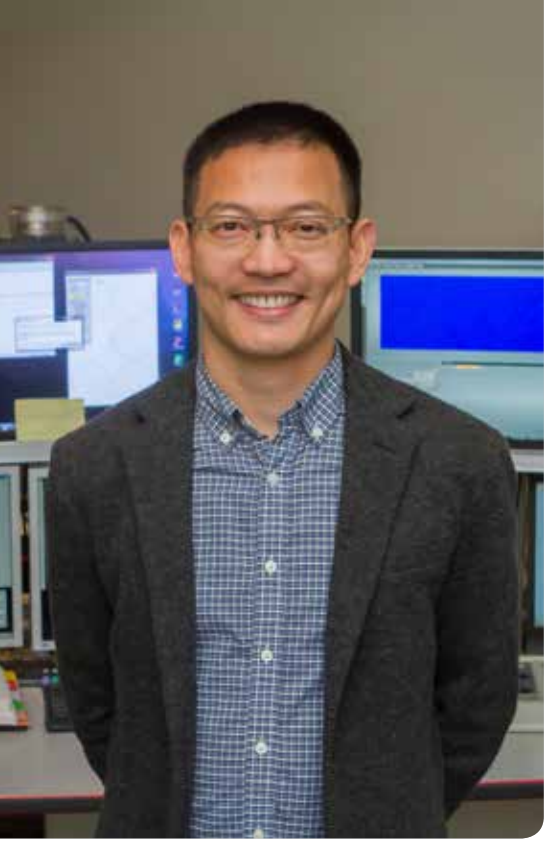
Results

Competitive internship salaries ($54,800 annualized in 2017) and 90% post-graduation employment rates for graduates are indicative of the widespread positive impact the program is having not only on individual students but also on the companies that hire them.

When the Master's Industrial Internship Program was launched at the University of Oregon (UO) in 1998, there were few programs designed specifically to facilitate students' transition from academia to employment with applied science companies.

"There was an unmet need in industry for master's level scientists," explains Stacey York, PhD, Director of the Master's Industrial Internship Program. The university has long maintained robust corporate partnerships, and it was by listening directly to voices from industry that administrators realized there was a growing market for "students who had sector-specific understanding but also an interest in gaining hands-on experience with the type of work being done [in the industrial setting]," York says. Over the past two decades, the program has developed further in tandem with the evolution of the job market. "The fields of chemistry and physics are really broad," notes York. That's why, in order to meet the demands of the industry, the faculty began to discuss which subfields of chemistry or physics would help make UO graduates most employable. "Three different tracks are currently offered with others on the horizon; the photovoltaic and semiconductor device processing track [which was first founded in 1998], polymer science and most recently optics," York says. "We build [our curriculum] around an industry sector that has significant career opportunities."

"In the Master's Industrial Internship Program, we really emphasize experience," says manager of the photovoltaic and semiconductor program Fuding Lin, PhD. "Students get that experience by going through the process together with very close guidance from us. Before joining the program, students don't necessarily know how to do the research and development work… or have the skill and confidence to work independently. Our courses allow them to acquire a more complete skillset directly relevant to what they will need in their careers."



"Our courses allow them to acquire a more complete skillset directly relevant to what they will need in their careers."

Fuding Lin

DOE gives students an edge

Throughout the evolution of the Master's Industrial Internship Program, York, Lin and their colleagues looked to the university's industry partners for feedback on the curriculum: Was it relevant? Were new UO graduates entering the workforce prepared to make an impact? "It was through their feedback, and through my assessment of early career positions [around two years ago], that we decided that a design of experiments (DOE) course using an industry standard software such as JMP was going to give our students an edge," York says. "They're only here for a finite amount of time – and we have a finite amount of resources to work with. So [adding a DOE course] was also a strategic decision: what would have the largest return on investment in terms of students' marketability?"

A key part of the program's experiential philosophy is the belief that, while they are at UO, students should preferably work with the same tools they will later encounter in the industry. So when York and Lin began investigating software needs, they again consulted the program's corporate partners. "Direct feedback from partners showed that JMP was the software they were using," says York. "When I try to understand what an early career would look like for one of our graduates – what the semiconductor scientist or polymer scientist or optical engineer needs – you can clearly see a pattern. It typically shows up in the preferred qualifications: JMP based DOE experience."

Lin now teaches the program's DOE elective, a course that has burgeoned in popularity with a majority of students enrolling each year. But before they get to the statistics, Lin says, students receive a crash course in JMP. "I project my screen whenever we're using JMP in class. That way, I can show them how to do the [exercises assigned to them] and they can follow in real time with the JMP software installed on their laptop. To make learning statistics more fun, we play with dice and use the data to demonstrate important concepts such as central limit theorem, confidence interval and hypothesis testing. Once they have the basic statistics and principles of DOE down, we then move on to class projects that most people can easily relate to." One such project involves a series of experiments related to coffee; how brewing parameters like water temperature, brewing time and the amount of sugar affect the taste. The coffee project also teaches students how to deal with subjective response data where individual preference plays an important role. By the end of the course, though, Lin pulls more advanced examples from industry and gives students a project to optimize a chemical vapor deposition process.

Advancing industry standards

Trevor Shear is an alumnus of the Master's Industrial Internship Program, and former intern with the Los Alamos National Laboratory (LANL) in New Mexico. After hearing from industry recruiters that DOE would be an important skillset to master while in graduate school, he enrolled in Lin's course and learned to use JMP. "I never took stats as an undergrad because I wasn't required to. But [once I got to UO] my mindset was that I wanted to make myself as valuable as I possibly could to an employer… I wanted to make myself too valuable to lose," Shear recalls.

After previewing the projects he would be working on at LANL, Shear was convinced his experience using JMP for DOE would be beneficial. "There were so many variables with so many levels [that could be adjusted]," Shear explains. "You could spend a whole career trying to figure this stuff out and maybe still not get the answer that you wanted."

Upon beginning his internship at LANL, however, Shear learned that his new colleagues were using a different software package than the one he'd become familiar with at UO. But after reviewing LANL scientists' experimental designs, he quickly pinpointed an opportunity to reduce both cost and experimentation time. "I prepared a DOE in JMP and walked them through what the work could be like," he says. To illustrate, he pulled up the coffee taste experiment data file from Lin's course. "The coffee project had good statistical data I could show them. Just going through that in JMP really sold them. It is not just about the analysis, but also the graphical representation of your data. Other software might be able to do this, but it takes them three or four hours where the same thing would take literally one click of a button in JMP. That made the group excited."

Shear's use of DOE was so effective that JMP was rolled out across his working group at LANL, and eventually they even upgraded to JMP Pro. "Working with x-ray tomography," Shear explains, "we could actually use JMP's statistical function and obtain a distribution of the void space in high explosives that would previously have taken [my colleagues] three or four days to process. But it took me about a half an hour with JMP. LANL management was really happy with that result."

A program with widespread impact

There is no question that the Master's Industrial Internship Program has been a success; 98% of those students enrolled undertake successful industry or government lab internships, York says. "In the past three years we've worked with 75-plus organizations, from start-ups to Fortune 50 companies to national labs. And the average internship salary this past year was $54,800." Since 2012, she says, students in the program have been funded in the magnitude of $11 million set forth by both industry and government labs. "This metric alone shows the commitment outside of academia to our educational model." And then there is the success of program graduates. York says 90% of participants are employed within three months of receiving their master's.

In 2017, the Master's Industrial Internship Program became part of the new Phil and Penny Knight Campus for Accelerating Scientific Impact, a $1 billion initiative announced in 2016, made possible by a $500 million gift from Nike co-founder Phil Knight and his wife, Penny. "It's a pretty monumental time to be at UO," says York. "The Knight Campus is designed to accelerate scientific impact. And our [Master's Industrial Internship] program has been at the heart of that effort all along; using an experiential educational model to get students really prepared to make an impact."

Online

Here's a free, online course that teaches statistical thinking for industrial problem solving: jmp.com/statisticalthinking

Webinars on how to teach and use statistics concepts, including DOE: jmp.com/academic-webinars

More on how UO trains students in the real-world knowledge and skills necessary to be successful in the industrial research laboratory: internship.uoregon.edu

"When I try to understand what an early career would look like for one of our graduates – what the semiconductor scientist or polymer scientist or optical engineer needs – you can clearly see a pattern. It typically shows up in the preferred qualifications: JMP based DOE experience."

Stacey York

Statistical Thinking for Industrial Problem Solving

A free online statistics course

In virtually every field, deriving insights from data is central to problem solving, innovation and growth. Learn core statistical concepts and when to apply them. Learn how to interpret and communicate results.

jmp.com/statisticalthinking

jmp®
Statistical Discovery.™ From SAS.

Seven self-paced modules include instructional videos, JMP® demonstrations, questions and practices.

Statistical Thinking and Problem Solving
Learn how to map a process, define and scope your project, and determine the data you need to solve your problem.

Exploratory Data Analysis
Learn how to describe data with graphics and use interactive visualizations to find and communicate the story in your data.

Quality Methods
Learn about tools to quantify, control and reduce variation in your product, service or process.

Decision Making With Data
Learn to draw inferences from data, construct statistical intervals, perform hypothesis tests, and understand the relationship between sample size and power.

Correlation and Regression
Learn how to study the linear association between pairs of variables, and how to fit and interpret linear and logistic regression models.

Design of Experiments
Learn the language of design of experiments (DOE) and see how to design, conduct and analyze an experiment in JMP.

Predictive Modeling and Text Mining
Learn how to identify possible relationships, build predictive models and derive value from free-form text.

Stephen Chambers, CEO, Imperial College SynbiCITE

A new ecosystem for commercializing biology research

Statistical analysis brings discipline to state-of-the-art research at Imperial College's synthetic biology accelerator SynbiCITE

Challenge

To improve the reproducibility of data obtained from engineered biological systems for translation of research and ideas into commercial opportunities.

Solution

SynbiCITE uses JMP to deliver a systematic approach to experimentation, based on a robust statistical framework combined with cutting-edge automation technology in the London DNA Foundry, to help research scientists achieve the data quality necessary to validate and scale successful commercial applications.

Results

An increased number of new products and services are now reaching the marketplace; these successes have led to the creation of new companies, jobs and return on investment of public funding.

In academia today, the lines between scientific disciplines have become increasingly blurry. One of the newest of these intersectional disciplines is synthetic biology, a field that unites biology, chemistry and engineering. Synthetic biology is radically changing how we engineer biological systems; moving beyond the confines of academic research, it is now responsible for creating new products and services with industrial, technology and biomedical applications. Synthetic biology is today what molecular biology was 40 years ago – and generating a new wave of applied biological innovation that is driving a modern scientific economy.

Established in the UK in 2013 with a £28 million (US$39 million) commitment from government, private and academic partners, SynbiCITE, a synthetic biology accelerator, seeks to connect emerging biological research with commercial opportunities. The organization's core facility is the London DNA Foundry, located at Imperial College. Here, scientists have access to cutting-edge robotics equipment that enables the design, construction and validation of gene constructs from end to end. Beyond access to equipment, the Foundry provides educational programs, mentoring and engineering and commercialization support through a diverse staff with a wide range of expertise.

This expertise starts at the top. The CEO of Imperial College SynbiCITE, Stephen Chambers, is a serial entrepreneur who holds a PhD in molecular biology, an MS in bioengineering and a BS in bio-

"The visualization really helps. That's the key part. Scientists can then start visualizing their experiments and understanding them."

Stephen Chambers

chemistry. He also leads the teaching faculty for the Lean LaunchPad for Synthetic Biology at Imperial College, an entrepreneurial training program, and maintains ongoing research in the advancement of engineering biology through automation and design of experiments (DOE). More than just an invaluable resource, Chambers provides the entrepreneurial vision that keeps SynbiCITE on the forefront of engineering biology commercialization.

From biologist to designer to entrepreneur

Synthetic biology also marks a shift in how biology is performed. "The traditional biologist is no longer a wet lab scientist. They are now more like a designer," Chambers explains. "Essentially, these biologists – or designers – are programming organisms through DNA code to make things." However, validating and scaling these experiments can be complicated and require expensive equipment and financial resources far beyond the means of most startups and small-to medium-sized enterprises.

"This is where the London DNA Foundry has a role, by providing a maker space for synthetic biologists. The London DNA Foundry allows them to rapidly build and test a prototype to see whether it works without making a huge investment." Engineers at the Foundry work alongside "designers" to execute on these ideas, facilitating a faster, systematic, reproducible process or prototype. With these working prototypes in hand, biologists can more clearly demonstrate the viability of results to investors, attracting capital and developing scalable infrastructure.

The role of the London DNA Foundry has taken on more significance with the increase in startup activity, which is driving much of the innovation in synthetic biology. Traditionally, technology was developed at a university and spun out into new companies. "The people involved would typically be heads of departments, professors, etc.," Chambers says, and these spinouts would be supported and nurtured by the universities. "What we are now seeing, especially in synthetic biology, is the growth of startups still founded by university personnel, but also students and graduates, who no longer license technology from the universities. These startups do not have the same support or resources available to university spinouts, and it is these companies that are flocking to the Foundry.

"Even the undergraduates are now starting companies," he says. "It's just amazing. They are outpacing the traditional university spinouts."

Preaching the importance of reproducibility

To facilitate consistent quality and success – and in some cases, regulatory approvals – well-designed and reproducible experiments are key. "As we scale, we do a lot of Quality by Design ... it's all about having confidence in the data and the rigor of that data so that you can present to other people and convince them that what we talk about is real," Chambers says.

Quality by Design (QbD) is a systematic approach to development that highlights the importance of process understanding and control while designing the project to address targeted quality measures and achieve specific results. While this sounds reasonable, biology is complex with various factors to consider in the course of experimentation, making it sometimes difficult to identify the source of issues and to reproduce outcomes. Often, biologists lack the statistical expertise or foresight to implement a QbD approach from the outset of a project, forcing them to rely on a combination of best guesses and conventional wisdom in their experimental design.

As CEO of SynbiCITE, Chambers preaches the importance of statistics to his biologists and engineers, as he has personally learned their value from his own previous experience in the lab. Early in his career, Chambers recalls, "I was doing a lot of automation around biology and generating a lot of data. The guys in the lab brought [the data] to me and said, 'What does it mean?' But I had no idea."

Hoping for guidance, Chambers brought his questions to a colleague with more expertise in clinical research, only to find that this colleague had questions of his own: "'Is it balanced? Is it powered? Did you use design of experiments?'" Chambers recalls. "When I said no to all these things, he said to me, 'You really should be using JMP.' That was the first time I realized there was this statistical package that would be amenable to biologists."
As Chambers well knows, not all scientists are knowledgeable about, or interested in, robust statistical analysis. "When you introduce [statistics] into a biological setting, there's an initial resistance to it," he explains. "They don't like the constraints of design of experiments. But when you see the results, they start buying in."

JMP is accessible to statistics-naïve biologists while still providing powerful analytics and visualization tools. With guidance from DNA Foundry staff, these biologists can be up and running with JMP within a few days, allowing them to build their research on standardized methodologies and achieve the type and quality of outcomes they require to grow their new ventures. Furthermore, the flexibility and visualization capabilities of JMP allow these scientists to understand what is happening with their data and rapidly adjust models based on new findings. "The visualization really helps. That's the key part," Chambers says. "Scientists can then start visualizing their experiments and understanding them." For biologists inexperienced with design of experiments and other statistical techniques, JMP gives them the tools to bring their ideas to life and show investors their potential.

Changing the conversation

While there is understandable reticence around some of these biological innovations, such as the release of genetically modified organisms, increased scientific rigor and reproducibility across the synthetic biology research landscape can only help change attitudes. Even more promising, as new products arrive on the market, advocates for synthetic biology can show real solutions to serious problems. "For example, the recent approval of Novartis' new gene therapy, Kymriah, which used engineered CAR T-cells to successfully treat childhood leukemia, is transforming the way that we treat cancers. Now we have real-life products that we can point to and appeal to people." These success stories help breed consumer and investor confidence.

Since their inception, SynbiCITE and the London DNA Foundry have shown a significant return on investment from public funding with economic benefits through the translation of research, the creation of new startups, jobs and private investment in synthetic biology companies. Going forward, SynbiCITE is now increasingly looking toward the private sector with industrial collaborators and customers, both large and small, who want to exploit the synthetic biology in the London DNA Foundry.

Online

Use QbD methods to optimize pharmaceutical production processes with this white paper: jmp.com/qbd

Learn more about SynbiCITE, the UK's national Innovation and Knowledge Centre: synbicite.com

Robots speed innovation in engineering biology

The London DNA Foundry unlocks the power of automated experimentation with design of experiments in JMP®


David McClymont, Head of Automation, London DNA Foundry

Despite his prolific automation engineering and assay development expertise, David McClymont, PhD, knows what it's like to juggle. As the head of automation at the London DNA Foundry, he runs multiple robotic systems in support of various synthetic biology projects, helping scientists prototype and commercialize innovative biological research. Specializing in the "design, assembly, verification and characterization of synthetic DNA," the Foundry gives scientists access to a suite of cutting-edge robotics, automated analytical equipment, mentorship and training – things that would otherwise be unavailable to them – with the aim of promoting the commercialization of synthetic biology research.

In his role as head of automation, McClymont acts as a consultant to the startup companies using the Foundry. "The goal of each startup is very different, but the standardized way of creating genes and genetic constructs are the same," he says. "So we built an essentially universal robotic platform for putting genes together that can be applied to almost any function. Using standardized DNA-building techniques, we think design of experiments can fit almost any need for many different types of biotech applications."

Automation enables reproducibility

Reproducibility and statistical power in biology research is a pervasive issue, one that stems from the twin problems of:
1) The time it takes to conduct and reproduce experiments by hand.
2) Limited statistical training and support for biologists, leading to a minimal application of robust statistics to biological questions.

With increasingly stringent statistical requirements with the aim of increasing reproducibility, McClymont says, "a number of funders and journals are really starting to crack down. Particularly with in vivo animal studies, there is a push to carry out power analysis and ensure robust experimental design." Not only are journals pushing for deeper statistical analysis, researchers are realizing the need to design and reproduce experiments in such a way as to extract data more tailored to answer complex questions.

With the state-of-the-art resources available at the Foundry, scientists can overcome the first issue through the automated execution of big experiments, removing the need for large numbers of staff to carry out experiments by hand which introduces potential bias and inconsistency. "Automation essentially gives you extra replicates for free," McClymont explains. Furthermore, automation allows users to pass control of monotonous tasks to robots, freeing up time and giving scientists the opportunity to focus on big ideas.

McClymont and his team at the Foundry are addressing the second issue by integrating statistical techniques like DOE directly into the design of the robotics and the systems that support them, lowering the threshold of knowledge required to use these analyses and helping scientists learn statistics while they conduct their research.

Building a powerful framework for statistical integration

Using DOE, researchers can evaluate multiple parameters in the same experiment, speeding the research process and generating consistent, high-quality data; this is particularly important for the entrepreneurial goals of the Foundry. "The data that [entrepreneurial biologists] generate must convince investors to invest in their company. There can't be any doubts about [the science]," stresses SynbiCITE CEO Stephen Chambers, PhD.

Instead of a scattershot approach, wherein researchers simply do as many experiments as possible, DOE adds a robust framework that guides study design and execution, allowing them to harness the power that end-to-end automation offers. As McClymont describes, "It's not just doing more replicates, but doing smarter replicates and building models which have much more value."

Online

How can you use design of experiments to increase knowledge of your process and design space?
jmp.com/pharmadoe

Read the full story:
jmp.com/dna

Driving the future of mobility

NXP Semiconductors accelerates chip technology for the auto industry

Challenge

Design, test and manufacture new semiconductor solutions that can power a rapidly changing automotive industry.

Solution

Use advanced analytics in JMP Pro to transform chip testing and manufacturing processes.

Results

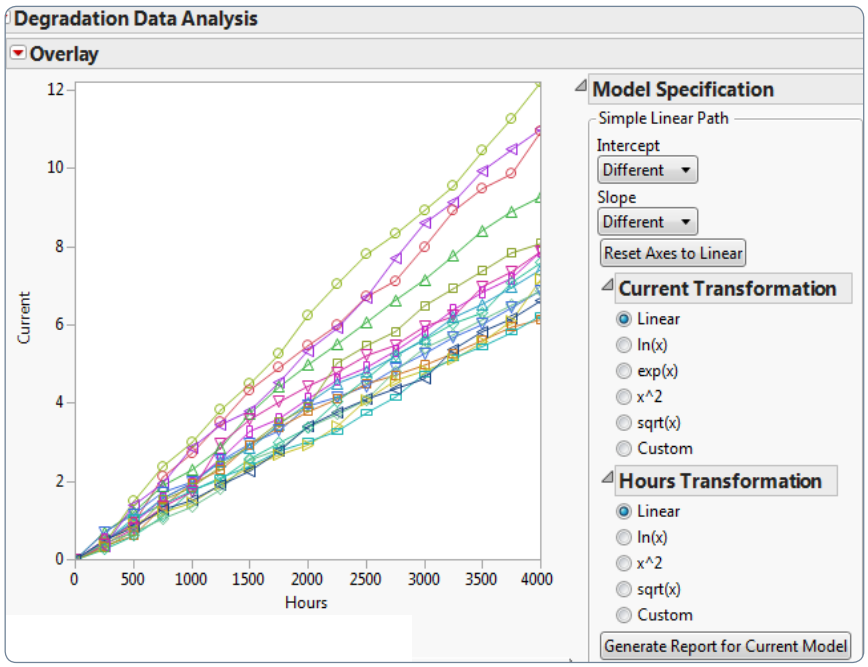
JMP Pro tangibly improves solution design and manufacturing, Six Sigma training and other process-improvement efforts that help keep NXP at the forefront of the market.

Cars and trucks will evolve faster – and more fundamentally – over the next 10 years than they have in the past 100. As the industry moves toward alternative power sources, expands connectivity options and embraces autonomous vehicles, it will require hardware and software that are both more diverse and more capable. At the core of those developments reside computer chips – and NXP Semiconductors.

NXP is the world leader in connectivity solutions for embedded applications such as connected vehicles. Based in Eindhoven, Netherlands, its 31,000 employees serve customers in more than 30 countries, generating annual revenue of US$9.5 billion. The company focuses on the application of semiconductors in three primary areas: automotive, industrial and Internet of Things (IoT). NXP claims the No. 1 market position in automotive semiconductors and microcontrollers, in-vehicle networking and entertainment, secure vehicle access and automotive safety features such as airbags and radar.

"Electronics are more and more necessary in a car," says Corinne Bergès, PhD, who manages risk assessment, statistical and safety analysis for NXP's Advanced Automotive Analog department and manages Six Sigma training across Europe, the Middle East and Africa. "It's not possible to even imagine a car now without semiconductors."

But to continue to lead the industry, NXP must design and make semiconductor solutions faster, more efficiently



Six Sigma practitioners at NXP use degradation analysis features in JMP Pro to model and predict drift on high-temperature operating life test results from a typical 2,000-hour test, as required by automotive standards, for a specific customer request of 4,000 hours.

and with greater precision than ever before. And to achieve that goal, the company must continually improve its manufacturing and testing processes.

Data-driven at the core

Central to the design and manufacture of semiconductors is data. For NXP and its partners, every aspect of the manufacturing cycle, as well as the continual improvement of that process, is data-driven. This data abundance represents both opportunities and challenges. "The more data we have, the more we can correlate that data, and the more information we can extract from that data," Bergès observes.

The challenge, Bergès says, lies in effectively managing the data volume – especially when it comes to testing. Chipmakers test their solutions with a variety of parameters and environmental conditions to identify the limits at which a device could fail or to control process variability. That process generates a vast number of data points. "Data is worthless if we don't have the right tools to work with it," Bergès notes. "So we need to be using the newest and most innovative means to manage this data. JMP Pro uses the most innovative and effective methods, and that's very important."

JMP Pro helps with both data visualization and data analysis. "Product engineers rely on data visualizations," Bergès says, because an image can be read and understood much more quickly than rows of data alone, thereby enabling engineers to quickly spot outliers, correlations or any issue. And interactive visualization functionalities in JMP she says, have done much more than that; they have helped engineers and operators across NXP grow their capability from univariate to multivariate analysis.

"In the past we did many tests, but we studied only a single test at a time, in visualization and univariate analysis," Bergès explains. "Today it's no longer possible to work with only one test." Instead, the company must correlate data for multivariate experimentation. "This kind of analytics requires advanced statistics and methods," she says – capabilities available in JMP Pro.

Cornering the market on 'corner lots'

While JMP software enables data analysis for a broad range of users, JMP Pro adds advanced features such as predictive modeling and cross-validation for scientists and engineers. One important practice for which NXP uses JMP Pro is in studying manufacturing-process variability for automotive semiconductor "corner lots."

In chipmaking, a "corner lot" is a design of experiments (DOE) technique that tests extreme fabrication parameters. To verify the quality of a circuit design, chipmakers fabricate corner lots, or sets of semiconductor wafers with process parameters set to extremes. They test devices made from these wafers under varying environmental conditions, such as voltage and temperature, to identify their operational limits. When NXP develops a new semiconductor, it uses the corner-lot approach to perform a comprehensive analysis of variabilities that could be encountered during manufacturing. This allows the company to identify and correct any technical weaknesses in the manufacturing process.

"The traditional corner-lot approach used many engineering lots, up to 75 and more for an automotive valve driver component, with each parameter assessed separately," Bergès points out. "It was limited in that it didn't evaluate interactions between parameters." JMP Pro enables the assessment of real corner points on single lots. As a result, NXP can control dispersion of class probe parameters. Its new statistical methodology also generates yield modeling and predictions.

The journey toward defect-free manufacturing

Bergès boasts Black Belt certification in Six Sigma, the well-known set of techniques and tools for process improvement Six Sigma is essential to Bergès – who provides Six Sigma training for NXP departments around the world – and to her company. Some 3,300 of NXP's employees are certified as White, Yellow, Green, Black or Master Black Belts. And at NXP, JMP Pro is essential to Six Sigma. "In the Yellow Belt curriculum, there are simple concepts such as distributions and standard deviations," Bergès says. "Between Green Belt and Black Belt training, there's much about statistics to learn. And the best trainings we have at NXP are performed in JMP Pro."

NXP uses JMP Pro in numerous Six Sigma trainings covering analysis of variance (ANOVA) and regression, custom DOE, modern screening design, measurement system analysis, sources of variation analysis, and reliability. "We designed these courses with JMP three years ago," Bergès says. "And these are now the main courses for the many people within NXP who are advancing to Green Belt and Black Belt levels."

The company's Six Sigma training also covers a data-driven approach to improving process quality, called e-DMAIC (elimination-Define, Measure, Analyze, Improve, Control). e-DMAIC lies at the heart of Six Sigma. "In fact, Green Belt certification requires an improvement project with an e-DMAIC approach," Bergès says. "This e-DMAIC project has to be validated by management and scorecarded. That ensures that the knowledge being gained is actually applied in a real-world situation." While DMAIC is a well-known, structured problem-solving framework, the letter e refers to the elimination of root cause(s) that can become a real mindset. This enhancement comes from a collaborative study between NXP and Toyota, and it continues to increase quality levels. Quality is everywhere in NXP; it's at the core of the company's "Total Quality" concept.

"It's simply no longer possible to do our work without JMP Pro."

Corinne Bergès

Benefiting users, driving industry success

NXP's use of JMP Pro has been an evolution. "We were using Minitab and then JMP," Bergès recalls. "And the engineers chose JMP for its compatibility and ease of use. Now I think we'll never give up JMP." Today, 1,600 of the company's data experts actively rely on JMP solutions.

One advantage of JMP Pro, she says, is that it can be used effectively by both highly advanced statisticians and engineers with only entry-level statistical training. "With Minitab, you have to have a certain knowledge of statistics to produce meaningful analytics," she says. "But because JMP is so intuitive, anyone can achieve useful analytics." Another clear benefit is the efficiency with which JMP Pro enables NXP to conduct testing and optimize processes. "With one or two clicks, we have the most complete and accurate statistical analysis possible. We use Python for high-volume data, but if we want the fastest result, without any need for programming, we use JMP Pro."

The ability of NXP to use JMP Pro to optimize semiconductor design and testing will help keep the company at the forefront of the market. As the pace of change in the automotive industry accelerates, NXP's capacity to rapidly and reliably design and manufacture new solutions will increasingly be a differentiator. Or, Bergès says, "It's simply no longer possible to do our work without JMP Pro."

Online

Want to transition from Minitab to JMP?
jmp.com/transition

See the power of JMP Pro:
jmp.com/pro

NXP Semiconductors is driving Internet of Things (IoT) innovation:
nxp.com

Mapping new frontiers in the infrared

Engineers at Lynred design high-performance technology for probes traveling deep into space on a voyage of discovery

Challenge

Lynred's fit-for-purpose, high-value-added spatial activity technologies – often engineered for a single use – require a significant investment in R&D. Costly prototyping and testing mean that engineers contend with unusually small sample sizes, motivating the need for robust statistical approaches.

Solution

JMP and JMP Pro help Lynred's engineers optimize quality, driving efficiencies from R&D to production and ensuring that each new model fulfills its brief.

Results

Lynred is the only space-qualified manufacturer of IR detectors to have delivered several dozen flight models to most of the major satellite and space missions over the past decade. The company has built more than 70 fit-to-purpose designs, including those that played a critical role in Sentinel 2, Tropomi, Hayabusa 2 and the ExoMars space programs.

In early 2018, the Hayabusa 2 space probe made history when it sent a lander to the surface of Ryugu, a primitive, diamond-shaped asteroid only half a mile wide. The probe was designed to gather data that will deepen scientists' understanding of the solar system's formation nearly 4.6 billion years ago.

This kind of science – unthinkable even just a few decades ago – is only possible today because of advances in engineering and the ingenuity of companies like French infrared (IR) technology pioneer Lynred, whose vast portfolio of IR detectors covers the entire electromagnetic spectrum from near to very far IR.

Lynred and its US-based subsidiary, Lynred USA, are global leaders in the design and manufacture of high-quality IR technologies for aerospace, defense and commercial markets. In addition to having products at the center of multiple military programs, Lynred's IR detectors are now the key component of many top brands in commercial thermal imaging equipment sold across Europe, Asia and North America. Furthermore, Lynred is today the leading European manufacturer for IR detectors deployed in space.

The origins of our solar system

On board Hayabusa's lander, a Lynred detector known as NEPTUNE can scan the rocky surface of Ryugu at the molecular level. "This IR hyperspectral map helps scientists identify the chemical composition of the sediments and therefore understand the history of how and when the asteroid was formed," explains Augustin Cathignol, Principal Engineer for Product Reliability and Data Science at Lynred. Like many of Lynred's technologies, NEPTUNE was

engineered specifically for MicrOmega, the IR microscope developed by IAS (Institut d'Astrophysique Spatiale at Orsay, France) with the support of CNES (Centre National d'Etudes Spatiales, the French space agency) that is transported by the Mascot lander of the Hayabusa 2 mission.

"Lynred's detectors are designed in a way which is really adapted to the mission. They're sur-mesure – bespoke," says Cathignol. Each detector or camera destined for a space mission is only produced once, which can mean years of prototyping and testing for a single end product. But these kinds of innovations open the door to new scientific and industrial applications, and the company doesn't shy away from a heavy investment in R&D.

Providing quantifiable evidence of reliability

With international space programs setting sights on ever more distant missions, Lynred's engineers face significant pressure to make IR detector technology smaller, lighter and more efficient. Product reliability is therefore a primary concern. "In our business, it's very important to be present at the beginning stages of research so that we can take reliability into account early on," says Cathignol, who leads a reliability team from the time the first prototype is developed in R&D through to when the final design leaves production.

If at any stage the product is not reliable, it could fail to meet critical deadlines or even fall short of a mission's narrow launch window, Cathignol explains. "We need to reach performance targets and ensure that everything is tuned and designed perfectly in order to achieve this level of performance." Providing quantifiable evidence of reliability to customers, he adds, requires complicated statistical analyses. "So I was very glad when I joined Lynred to see that JMP was in the walls for them."

JMP® Pro, from modeling to visualization

An engineer and semiconductor physicist by training, Cathignol had used JMP statistical discovery software in previous roles. When he came to Lynred, potential applications for the software were even more abundant. "Now I use JMP for various types of linear and sometimes nonlinear modeling," he says. "I also use recreations with either continuous or categorical variables, as well as distribution analysis and advanced techniques, some of which are only available in JMP Pro."

JMP makes it possible for Cathignol and his colleagues to quickly identify issues with testing or production that affect yield, scrap or opportunities for other cost efficiencies. By catching potential arisings early, the team is able to accelerate the cycle of engineering while keeping costs down and delays infrequent.

In one instance, when a high-value thermal process in the production line was showing signs of low yield, Cathignol and his team used signal processing in the Functional Data Explorer in JMP Pro to optimize the process. "The idea was to look for a correlation between the shape of the signals generated by the equipment itself (called 'log files')," he recalls. "For this, we used Functional Data Explorer and dynamic time warping, which generated some distances from signal to signal. Then, with clustering techniques, we were able to see the correlation we were looking for, which was very interesting and useful."

With the lengthy R&D cycles typical of highly specialized technology – and the immense cost of testing prototypes – Lynred's engineers often have a large number of data points but only a very small sample size. And this, Cathignol says, presents some obvious statistical challenges. "Compared to the semiconductor or automotive industry, the number of products we build is pretty

"The user experience, the ergonomics, the aesthetics of visualization, the mathematical rigor and the support – that's why I really love JMP."

Augustin Cathignol

Cathignol points to a case containing IR detector prototypes at the company's site in Veurey-Voroize, France. Lynred IR detectors are designed to image the IR fingerprints of molecules, which emit radiation at given frequencies. Each molecule or element has its own IR signature, allowing scientists to identify chemical compounds by matching IR data to the known spectrum.

small," he says. "This means we work on a small sample size and it's much more difficult to derive conclusions. We often have to adapt our methods to take this [reality] into account." JMP helps to mitigate these limitations, Cathignol says, by boosting the statistical robustness of every analysis.

Product lifespan in extreme conditions

Once a Lynred product reaches a more advanced stage in its development life cycle, Cathignol and his colleagues are responsible for quantifying product reliability in different mechanical and thermal environments. Whether Lynred's technology succeeds or fails is often related to whether it can function in extreme conditions. To provide customers with the guarantee that a detector will work, engineers must estimate the product's lifetime under an array of environments including cryogenic temperatures and space radiation.

The company is equipped with environmentally controlled labs that simulate these extreme conditions to test the robustness of new prototypes or components. In the development of their NEPTUNE model, for example, Lynred had to ensure the detector could not only survive the journey to Ryugu, but also withstand the conditions at the surface of the asteroid.

Dealing with big data, Cathignol says, "the higher the volume of data we collect, the greater the chance we will get some signals from them that add value and build knowledge. But some data are just not that meaningful, and it's a big challenge to sift through and focus only on what will maximize value…. It can be complicated to visualize. I really like JMP for this, especially Graph Builder, which is very practical and efficient."

Traversing the final frontier

"We watch our competitors to maintain an innovation rate which is high enough to keep being leaders," Cathignol says. "We're facing tough competition from other companies that are [increasingly achieving good] performance from their own detectors. So we need to keep innovating, and we do so.

"That's why I really love JMP! The user experience, the ergonomics, the aesthetics of visualization, the mathematical rigor and the support," he adds, noting that one of the most valuable things about JMP is the close collaborative relationship he maintains with JMP technical experts. JMP staff, Cathignol says, brings a human touch to the software, enabling his team to solve problems quickly and work smarter to drive the innovation the Lynred brand is known for.

"We are very proud of the products we build," Cathignol says. Just think: "At any given time, they are traveling in the universe around asteroids or Mars or other planets in our solar system." It's technology like this that will help humankind traverse the final frontier – space: an unconquered vastness of immense opportunity and discovery.

Online

Learn more about Functional Data Explorer in JMP Pro: jmp.com/pro

A leading innovator of infrared imaging detectors: lynred.com

Telefónica democratizes data-driven thinking

The telecommunications giant promotes creative decision making based on data and interdisciplinary collaboration

Challenge

As the explosion of big data continues to transform business, marketers must think creatively about what they can do with data to both monetize information and add value for the consumer.

Solution

Use JMP to extract actionable insight from any given data set – whether that be external data like weather reports or price information, or internal data like service usage patterns – to narrow the gap between data and decision across the organization.

Results

Telefónica Germany's Digital Command Center has adopted a transformative new system, based in JMP. Known as the JAM Sessions Methodology, it allows professionals from different sectors and backgrounds to collaborate on data insight generation.

"Jam musicians are masters of communication; every aspect of their music hinges on real-time collaboration," says Alfredo López Navarro, Knowledge Manager at Telefónica Germany's Digital Command Center. "A quartet is a musical metaphor for conversation; each group has its own rhythm, style and way of coming together as a whole, developing specific patterns of interaction that directly influence their music." This metaphor is at the core of Navarro's brainchild, what he calls the JAM Sessions Methodology. It's a guideline for working fluidly with data that aims to promote creativity and bridge the gap between available data and organizational knowledge.

When Telefónica, one of the world's largest multinational telecommunications providers, was recently recognized by Fortune magazine for its innovation, Navarro says, the JAM Methodology exemplified that spirit of evolution. "JAM is a fruit of that tree," he says. "It's capable of incorporating different levels of skills and a wide range of profiles into our data-driven decision-making culture."

Telefónica has implemented the JAM method across a wide swath of the organization, from product development and R&D to market research and marketing. And as a JAM champion and proponent, Navarro's remit is to explore data – and help others to do the same. "I'm very lucky!" he says. "I get to play with data, exploring it from multiple sources and shifting through all incoming data with the goal of discovering insights and turning them into competitive advantages that improve

business decisions and sustain operational and financial efficiency."

Communication in Industry 4.0

Over the past two decades Telefónica has evolved from its foundations as a traditional telephone company, providing landlines and mobile phone services, to a fully fledged internet giant, upon whose services millions across the world rely to navigate the interconnectivity of modern life. As the company prepares to launch mobile 5G networks, the scale of its operations has never been bigger, and equally so are its challenges. Today being a telecommunications company is about much more than providing the physical infrastructure and devices for internet browsing, but encompasses navigation, self-driving vehicles, smart city developments and much more. As ever more of daily life comes to rely on smart technology, companies like Telefónica, which provide consumers with access to that technology, are at the forefront of the new data revolution. And Navarro's efforts help Telefónica capitalize.

"Communication will explode, because everything is going to be connected to everything else," he says. "In some cases, there are technical barriers as with reliable and low-latency communications which will power brand new industries and services. Sometimes, as in Internet of Things applications, the ceiling is cultural, for smart cities or smart grids to be fully incorporated into our daily lives, society must change accordingly. Cultural transformation is, by the way, one of the biggest challenges for Industry 4.0. And the JAM Methodology pays special attention to it."

Freedom from old ways of thinking

For Navarro and his team, JAM Sessions are all about experimenting with data to generate insights into these new growth opportunities. "These days, as well as in the heads of the experts and professionals, knowledge comes in the form of data," Navarro says. "The 'data cyborg,' if you allow the metaphor, incorporates the best of human creativity and feelings together with the cold, fast, algorithmic logic of bots."

The JAM Sessions Methodology is a process Navarro has devised to stimulate collaboration between professionals with the explicit aim of creating outside-the-box thinking, which will lead to strategic marketing and business insights. Like the jam bands he admires, it's about experimenting with a range of tools to find solutions that work, taking creative, collaborative thinking to drive commercial success. The JAM Sessions Methodology motivates interdisciplinary collaboration

The JAM Sessions Methodology: An overview

The following list highlights the key elements of the JAM Methodology. These can be thought of as the building blocks of the overall methodology and process.

1. **Information retrieval.** The JMP Query Builder is a powerful tool that helps Navarro and his colleagues reduce complexity and speed processes.
2. **Quality control algorithmic checks.** Quality verification is critical, as inaccurate data could contribute to misleading insights.
3. **Automated data review.** Software agents (bots) aggregate the data in a prestructured way and note any significant differences. When appropriate, these bots create several types of automated alerts.
4. **Table creation.** Attributes relevant to each segment are calculated and stored for each specific table.
5. **Filtering to specific views of the market.** These filters are often dynamic, with the ultimate goal being to aggregate the data in a way that reasonably emulates the view of a consumer facing the market.
6. **JMP Journal documentation.** JMP Journals keep track of any analysis and the context for the data used in that analysis – every data transformation, from end to end.
7. **Visual storytelling.** Data visualizations, enhanced by statistically significant knowledge produced across the JAM data flow help convey insights.
8. **Culture transformation.** As more people come into contact with the JAM Methodology, the benefits of cross-pollination become more clear. There is a tipping point wherein JAM becomes the common way of working.

and facilitates communications between individuals at different statistical skill levels.

To ensure JAM would catch on, Navarro says he had to find a way to democratize data science so that people from different backgrounds could adopt the methodology in a meaningful way. "Data is said to be the 'new oil' in our age," he says. "But I think that we are in a transition period where some of the new possibilities are still constrained by old ways of thinking."

JMP® brings insights, promotes behavioral change

Though most at Telefónica were already avid users of Excel, Navarro realized that in order to drive a real cultural shift within the organization, they required a new tool - a software package with a more sophisticated analytics back end than Excel, Tableau or SPSS, but also a more user-friendly, interactive interface that wouldn't be daunting for statistics-naïve employees.

"I'm always in the scanning mode," Navarro says. "I'm very curious and always want to try new tools and see how they perform." When he first encountered JMP, Navarro recalls, the impression was immediate. "JMP can do everything that SPSS can, and plenty more things too. It's so interactive, so visual." JMP is now fundamental in bringing the insights that drive behavioral change – and with it, better business across Telefónica.

At the moment, Navarro and his team are working with pricing, looking at market data as well as external figures – including things like weather and macroeconomic data sets – to produce insights about internet usage and costs. For Navarro data visualization is the most important part of the process; "There is a magic effect in touching the data," he explains. "Sometimes visualization is enough, but storytelling with data is key. JMP Graph Builder is king here. We do a lot of modeling and multivariate analysis and the new Project feature in JMP 14 makes it very easy to document complex processes."

For those with less statistical training, features like JMP Projects help facilitate JAM Session conversations that would not have been previously possible. "The Project feature is a real game changer… People are excited," Navarro says. "We can now use JMP like a PowerPoint."

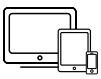
Even the more experienced data scientists in Telefónica's ranks see significant advantages in JMP, Navarro adds. Though statisticians or developers at Telefónica may regularly write scripts or contribute to an increasingly vast open source library, they still benefit considerably from the graphical interactivity of JMP. JMP Graph Builder is a cornerstone of data exploration, even when data scientists may choose to run their own scripts in R or Python with the output.

"Sometimes our data comes from a process that originated in R or in Python, but then we feed it into JMP... connect through the Query Builder to a data set and then do the exploration and gain insights," Navarro says. "You can be an open-source-minded data scientist and still be in love with JMP."

Business knowledge meets statistics

Applying a collaborative, creative approach to statistical analysis is at the very heart of Navarro's work and his JAM Sessions Methodology; he is passionate about demystifying the world of data both within Telefónica and beyond.

"As far as Telefónica Germany's Digital Command Center is concerned, conducting JAM Sessions is an open exercise of insight generation where business knowledge comes across statistics with visual interfaces facilitating team-ups between non-statistical experts and data scientists," explains Navarro. "Ideas flow spontaneously and unexpectedly. Nevertheless, the same way an impromptu concerto is possible only after years of practice and study, a systematic methodology is needed for this serendipity to be achieved in a useful way: insights and theories ought to be grounded."



Online

How can companies encourage their employees to adopt data visualization techniques and an exploratory mindset?
jmp.com/mindset

Telefónica uses telecommunications technology to connect people:
telefonica.com

A recipe fit for the modern consumer

Sensory analysis helps Kirin align its beverage portfolio to the evolving marketplace with the introduction of Hyoketsu, Japan's leading ready-to-drink cocktail

Challenge

Alcohol consumption in Japan is decreasing overall, with nationwide consumer data indicating that a growing segment of the population increasingly prefers liquor-based cocktails to beer. Kirin has turned to sensory analysis and other analytical tools to identify candidate drinks for the changing consumer marketplace.

Solution

Identify the specific properties of liquor-based drinks that make them more palatable to certain segments of the population by using JMP Pro and JMP to perform multivariate analyses.

Results

With the results of their advanced sensory analyses, scientists at Kirin developed Hyoketsu, a hugely popular ready-to-drink (RTD) cocktail that has since become the No. 1 RTD product in Japan.

Japanese consumers took their first sips of Kirin beer in 1888, when the forerunner of Kirin Company Ltd. brewed a malt beverage for local drinkers. In 1907, the Kirin Brewery Company formed with the intention of widening the market for beer to a nation mostly unfamiliar with the product. Since then, Kirin has consistently brewed the nation's No. 1 beer, in addition to offering an array of other alcoholic and non-alcoholic beverages.

Responding to new trends in consumer tastes

Though Kirin led the industry for nearly a hundred years, the company began losing market share in 1985. And the brand took an even bigger hit when it was replaced as market leader in 2001 by Asahi Breweries. Furthermore, Kirin and its competitors have had to contend with new challenges as Japan's demographics have evolved.

In light of both an increasingly competitive marketplace and an evolving consumer demographic, Kirin opted to meet these challenges by developing an expanding portfolio of ready-to-drink (RTD) vodka and fruit juice cocktails. The company did this strategically, with its research and development arm responding directly to consumer preference data.

"Unlike the US, Japan has a rapidly aging population, and we have fewer people drinking alcohol," says Hideharu Odai, Senior Research Scientist at Kirin Company. "Japanese consumer data has indicated that women and younger generations in particular are not drinking very much. This may be a common trend

in advanced countries. Overall, alcohol consumption by volume is declining. Before, consumption of beer was highest. However, for the last 10 years, there has been a decline in the consumption of beer by 1 or 2 percent every year. So at Kirin, we knew we needed to develop a strategy to fill this gap. For that, we focused on distilled types of spirits – the RTD drinks."

It was an investment that paid off and in 2001, Kirin saw a dramatic increase in share prices upon the introduction of its landmark RTD product, Hyoketsu. A cocktail that combines fruit juice, soda and vodka, Hyoketsu has a light, fruity flavor – a recipe Odai and his colleagues at Kirin developed in direct response to younger consumers' preference for beverages that are "drinkable" and "refreshing." And since its launch in 2001, Hyoketsu has been the No.1 RTD brand in Japan, with more than 10 billion cans sold since 2015.

Sensory testing data helps scientists identify new candidate beverages

With the marketplace growing steadily more competitive as other beverage companies release their own RTD products, Kirin continues to innovate and improve upon Hyoketsu. Armed with the sensory profile used to optimize the composition of the major flavor compounds in the original Hyoketsu recipe, Odai and his colleagues at Kirin have focused their efforts on refining the sweet spot for RTD beverages. For this, they use a combination of taste tests and analytics, using JMP and JMP Pro, among other things, to review potential sensory profile combinations to establish candidate drinks for further testing.

New drink recipes include vodka, not-from-concentrate fruit juice and rheological agents – items that affect the flow of the liquid and fine-tune the drink for consumers. Traditional testing of the different drink combinations at

Hideharu Odai stands in front of a display case mapping Kirin's factories across Japan.

distinct and variable combination levels made cost a concern, but with the applications available in JMP and JMP Pro, the analysis could be done quickly and efficiently.

"People are very sensitive about how they feel when they swallow their drinks and how the liquid feels as it passes through the throat," Odai says. "That's very important. But even so, historically, there has not been much texture-related research when it comes to alcoholic beverages."

Odai believes that for both beverages and food, including alcohol, instrumental analysis and sensory evaluation are mandatory. Advanced equipment along with methods of sensory evaluation now make it possible to obtain vast amounts of data. Odai and his colleagues at Kirin have used JMP and JMP Pro to gain new knowledge by integrating and analyzing these mountains of data more closely.

"Before, in the R&D department, they had another analysis methodology, a general type of analysis that people used," Odai says. "To introduce a new methodology, we implemented multivariate analysis together with risk analysis in order to make it possible to match the basic research. We also started doing proficiency analysis and validations. We did not have that analysis before."

A more robust tool supports multivariate analysis

Odai has not always used JMP, he says. "Kirin already had a software package in place for this kind of analysis. However, in order to execute the multivariate analysis we needed, the existing software wasn't good enough. That's why we implemented JMP."

Kirin needed to understand how to put the right products on the market. "Having multivariate analysis available helped to get the information or the insight in terms of what the flavor should be or what sort of concept we should follow in a new type of alcoholic beverage."

Odai was the first person to use JMP at Kirin. "Around 2010, when I was first transferred to alcoholic beverage research and development. That's when I needed to have statistical analysis software," he recalls. After learning that JMP was the industry standard for sensory analysis in the consumer packaged goods industry – and used by companies like Kraft Foods – Odai decided to try the

"Having multivariate analysis available helped to get the information or the insight in terms of what the flavor should be or what sort of concept we should follow in a new type of alcoholic beverage."

Hideharu Odai

software for himself. In the wake of Odai's early successes, other researchers at Kirin also started using JMP. "I showed other employees how to use JMP, and then through that process, my colleagues started to understand the benefits that they can get out of JMP. That's how we increased the number of users here at Kirin – they found that it was so easy to use."

Advanced analytics help adapt beverage thickness and viscosity

After experimenting with data mining, predictive modeling and a variety of nonlinear regression models in JMP, Odai decided to upgrade to JMP Pro in order to take full advantage of the software's more advanced functionalities. One of the added features of JMP Pro, the PLS analysis of primary, squared and interaction terms, yielded good regression models for beverage palatability.

One study led by Odai investigated the thickness of alcoholic beverages by comparing the sensory scores and rheological properties. Thirty-three adults participated in a preference investigation. Using pectin and xanthan gum as thickening agents for Kirin's designed beverages, researchers conducted a sensory evaluation of thickness with the selected panel using magnitude investigation. Odai and his colleagues then subjected the data they obtained to statistical analysis in JMP Pro.

The results indicated a proportional relationship between the logarithms of both the sensory score and the rheological properties, confirming that Stevens Power Law applies to the thickness of suspension. Odai tabulated the Power Score of thickness at 0.9 – higher than coffee aroma but lower than salty tastes.

Odai also found that the addition of sugar, acid or ethanol varied the beverage prototype's sensory score and rheological properties. By using stepwise partial least-squares regression with variable importance in projection (PLS-VIP), the sensory scores of projections were successfully predicted. PLS analysis of primary, squared and interaction terms yielded good regression models for beverage palatability models.

Such findings allow Odai and his colleagues to understand how the optimum properties of thickening agents in designed beverages can improve the recipe and give consumers a more appealing product. "I think we are getting there, to good solutions. We're getting good results with JMP," Odai says.

Hyoketsu, Japan's leading ready-to-drink cocktail, continues to top the charts

Tests like these allow Odai and his colleagues at Kirin to select candidate drinks at substantially lower costs. "That's how we are able to reduce the cost of finding the candidate idea for the new type of beverage," he says. And, they are able to create drinks that cater to Japanese consumers' shifting taste preferences.

In the wake of Hyoketsu's widespread success, Odai says, Kirin is on its way to getting even more impactful results with sensory analysis. In Japan's highly competitive RTD market, Kirin is elevating the science of taste with an innovative approach to analytics.

Online

Learn about preference mapping in the food and beverage industry: jmp.com/consumer

Find out which product features are most important to your customers: jmp.com/choice

For more information on Kirin: https://www.kirin.co.jp/company/english/

Statistics at the center of the evolving pharma industry

NNE and Ferring Pharmaceuticals navigate compliance and regulatory processes with JMP® Clinical

Challenge

High costs, changing regulatory guidelines and an increasingly competitive marketplace pose a serious hurdle to pharmaceutical companies aiming to get new therapies from the development pipeline to market.

Solution

Pharmaceutical consultancy NNE partners with leading clients like Ferring Pharmaceuticals to design and implement custom centralized statistical monitoring programs that meet the unique requirements of each of the company's new Phase II or III drugs. JMP Clinical enables NNE to roll out scripted applications and dashboards that streamline workflows.

Results

By instituting a sophisticated statistical approach that systematizes data quality improvements, NNE has helped clients like Ferring to navigate compliance and regulatory processes and, ultimately, bring new products to market more efficiently.

As the drug development industry evolves, so do the regulations that guide it, compelling pharmaceutical companies to repeatedly adapt their processes to comply and compete in the marketplace. Addressing these challenges has made the pharmaceutical industry particularly receptive to new statistical approaches; not only are methods like centralized or risk-based monitoring increasingly required by governing bodies like the FDA and EMA, these approaches have the potential to streamline the drug development process and even reduce costs.

To facilitate the transition away from traditional monitoring, many pharmaceutical companies have sought partnerships with experts in this growing field who can tailor solutions to their exact needs. Ferring Pharmaceuticals is one such company; Ferring's biostatistics group has developed a partnership with NNE, a leading pharmaceutical engineering consultancy with vast experience in the drug development process both through manufacturing optimization and statistical support for clinical trials.

"Our core competence is building factories in the pharmaceutical industry – alongside this we have a team of 200 consultants working to optimize and drive the best production, products, quality levels and, for clinical trials, GCP (good clinical practice) levels," explains Kasper Munck, Principal Consultant at NNE and an expert in statistical monitoring and operations research. "And once their factories [are built], we can also fill all the other needs [a client] might have."

Munck and his colleagues at NNE contribute across all stages of the drug development process. "I'm called in

when [our clients] have a need for either automation or clinical trial analysis," he says. With the highly specific data collection requirements set forth by regulatory authorities in markets around the world, having targeted statistical support at all phases of development is essential, particularly at companies that lack high-level statistical systems infrastructure in-house.

High-quality data, regulatory compliance

Because much of their clinical trial collaboration is on products in Phases II and III – which are large, expensive and essential to gaining regulatory approvals – ensuring high-quality data and regulatory compliance is of utmost importance. To that end, Munck and his team work with Ferring's biostatisticians to develop central and risk-based monitoring approaches, tailoring analyses to any areas deemed critical by the study team in view of regulatory standards. This includes scripting specific dashboards to allow users to monitor study progress and assess ongoing trends including fraud detection, giving them tools to inform decision making.

In contrast to traditional, full monitoring – whereby all data collected for each patient is reviewed – a risk-based approach identifies the key data points to the study and lets study teams focus primarily on this data. "[By instituting a risk-based approach,] we can free up so much time, enabling people to actually perform real tasks instead of just looking at data," explains Munck. "That's so beneficial for the company. It's a question of optimizing time, and at the end of the day, the employees can focus on the fun part of their work instead of tedious tasks."

When Ferring initially approached NNE, the company's clinical trial managers had an urgent need to automate their system for constructing patient profiles and narratives. Munck demonstrated how JMP Clinical could easily address this need, and although Ferring's existing software was adept at distributing results, it lacked the statistical robustness that JMP Clinical offered. "That really got them hooked," he recalls.

With the arrival of a series of new study opportunities shortly thereafter, Ferring was primed to take full advantage of the comprehensive statistical capabilities in JMP Clinical. NNE was tasked with building a centralized monitoring tool that incorporated the parameters set forth by Ferring and that allowed for flexible, on-the-fly statistical evaluations. "The tool needed to be dynamic because each therapeutic area could have a very different [target endpoint]," explains Munck. While Ferring had worked with experienced coders before, NNE's statistical expertise allowed them to go beyond simply coding what Ferring's monitors dictated; instead NNE engaged with its client to develop the ideal model collaboratively. "We had the capability of understanding both how to code, and the purpose and requirements for the right statistical model," he says.

Working with a dedicated specialist from NNE was a great experience, says Egbert A. Van Der Meulen, Senior Director of Biostatistics at Ferring Pharmaceuticals. "We jointly decided to develop interactively with writing and modifying the specs, only starting from two basic requirements: first, being able to perform site performance using statistical inference and clinically relevant endpoint metrics (based on volcano plots). And second, recruitment quality as opposed to recruitment speed (the latter being only too often the limited short-term focus). All, of course, based on a dynamic click-and-play, slice-and-dice JMP Clinical interface tailored for central statistical monitoring."

Even statisticians can become number blind

Like many experienced consultancies, NNE remains officially software-agnostic. However, with the freedom to use the tools best suited to their clients' specific challenges, Munck says he has found that JMP Clinical serves his portfolio of pharmaceutical R&D clients particularly well. "JMP Clinical has definitely been the easiest tool to both teach and learn," Munck affirms. "That is why we tell our customers that if you don't already have another preference, you should use JMP."

"JMP Clinical is fully based on CDISC ADaM and SDTM data, which Ferring has had in place for over a decade," adds Van Der Meulen. "As such, it has the potential to, frankly, do anything needed for conducting and reporting clinical trials, whether it is central statistical monitoring, signal detection, narrative creation, building clinical data warehouses or, indeed, full-blown clinical trial reporting with close to a single push on a button. This great potential is not fully realized yet, but it will make room for statistical programmers to move on to the next, likely more challenging, level of adding value to clinical trial reporting and data explorations."

Beyond its reasonable learning curve, JMP Clinical offers interactive visualization capabilities that make it invaluable in a clinical setting. With such high volumes of data that must be analyzed, even statisticians can become "number blind" and struggle to see the big picture, Munck says. "And that's the good thing about JMP: It always pairs statistics with graphs. When you have a lot of information, looking at graphs is so much easier than looking at a lot of numbers." Take, for example, the volcano plot. Munck says this visualization is especially useful in identifying statistically significant differences in reporting between clinical trial sites. And that's crucial to understanding potential data quality issues and outliers.

Functionality is not the only consideration for NNE and its clients, however. In a highly regulated environment like

"JMP Clinical is fully based on CDISC ADaM and SDTM data, which Ferring has had in place for over a decade. As such, it has the potential to, frankly, do anything needed for conducting and reporting clinical trials, whether it is central statistical monitoring, signal detection, narrative creation, building clinical data warehouses or, indeed, full-blown clinical trial reporting with close to a single push on a button."

Egbert A. Van Der Meulen

pharmaceuticals, tools and processes must meet certain standards to be accepted by regulatory bodies. While a variety of statistical tools are available, including inexpensive open source options, companies must be careful. "[Programming languages like R and Python] are definitely a direction a lot of companies are moving toward," Munck points out. "But you need to trust that the things you use… have actually been developed in the right way, and that you can use them for validation purposes." Otherwise, companies conducting clinical trials risk the rejection of their regulatory submissions and could be forced to implement other costly changes to their studies.

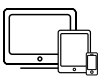
Saving time in a time-sensitive process

With a combination of standard functionalities, like the data quality and fraud applications, and customized scripts in JMP Clinical, NNE and Ferring can assess variation in critical variables within and between sites to pinpoint risks and determine where additional monitoring or oversight is required, all within a shorter timeframe than before. "What we've done is adapted some of the filtering options in JMP Clinical, implementing JMP scripting on top of the existing customer functionality," explains Munck. Working together over the last year and a half, NNE and Ferring have honed the central monitoring model and now can integrate new trials into the framework within just a week.

Data analysis and review are ongoing throughout each trial, and using JMP Clinical, project teams can review data, quickly evaluate study progress and respond to stakeholder questions immediately by using the adjustable statistical interface. "They have weekly meetings where they look at the trial," Munck says. "You don't need to prepare analyses, you don't need to print anything, you just bring in the data, put it into the tool, and you can answer all the questions directly at the meeting." In the past, Ferring's teams would spend a day or two before each meeting running ad hoc analyses and preparing slides, and even that didn't guarantee that they could answer all potential questions. Having the right tool and the right statistical approach has not only saved time in the analysis and preparation stage, but made meetings more productive.

As they implement their central monitoring approach on a wider array of projects, NNE and Ferring continue to explore new ways to streamline study operations and stay on the leading edge of statistical innovation. For Ferring, Van Der Meulen adds, "We are presently exploring with NNE – keeping the JMP Clinical business unit in the loop – the possibility of auto-generating statistical reports and further down the line, building clinical data warehouses using JMP Clinical and our ADaM-based repository databases. We also plan to explore how JMP Clinical can be put to use in the design of experiments applications of more complex adaptive and possibly Bayesian Seamless Phase IIa/IIb or II/III trial designs. JMP Clinical would be preferred over a stand-alone special purpose software package – which would likely be rather more expensive.

"Having a wonderful tool in place for central statistical monitoring is not the end of our collaboration with NNE – far from it."



Online

Modernize your medical and statistical clinical trial monitoring:
jmp.com/monitoring

Developing medicines to improve patient care:
ferring.com

End-to-end engineering services tailored to pharma:
nne.com

When it comes to unstructured data, Text Explorer does the heavy lifting

Oshkosh Corporation uses free-form service call records to evaluate past performance and prioritize engineering improvements

Challenge

Optimize equipment performance and minimize machinery downtime by providing rapid, proficient technical support over the phone to customers around the world.

Solution

Use JMP Pro to mine unstructured text data from support calls, identifying common problems and matching sources of malfunction with verified solutions that have proven successful in the past. With an ever-growing repository of maintenance data that can be deployed to troubleshoot recurring issues, Oshkosh's 24/7 global service providers are now able to proactively resolve service calls as they come in.

Results

A dramatic reduction in incident resolution time has not only saved on labor costs and improved customer satisfaction, insights gained from data analysis now guide future engineering efforts to preempt mechanical problems in products currently in development.

Every piece of equipment, from the smallest chip to the biggest processor, will require diligent maintenance throughout the course of its useful life. A smartphone might need nothing more than a software upgrade or battery replacement, amounting to fairly inexpensive care. But when a heavy-duty piece of equipment like a military grade tactical truck or a heavy equipment transport provider goes down, the costs of troubleshooting quickly escalate. Wisconsin-based Oshkosh Corporation is committed to strengthening customer support operations using state-of-the-art text analytics to minimize machinery downtime and preempt future mechanical shortcomings with a data-driven design process.

Oshkosh is one of the world's leading manufacturers of specialty trucks and access equipment. In business since 1917, the company has worked tirelessly to design and deliver tactical vehicles critical for defense, access, rescue and commercial missions around the globe. It's a competitive market and Oshkosh maintains its leadership first by providing superior system performance but also by offering exemplary post-purchase care through a 24/7 global customer service and maintenance network. "We have to really work hard to gain new business, satisfy our current customers and make our product even more competitive in the market," says Oshkosh Senior Chief Engineer Ron Zhang.

Service calls provide a wealth of free-form text data

Zhang works in Oshkosh's research and development group, leading a team that is currently pioneering a new use of statistical methods to improve product efficiencies, boost performance and

inform preventative maintenance deployment. "How can we make our design even better, even more efficient, better-performing and cost-competitive? Can we help customers improve their operational efficiencies?" Zhang asks. "We can. By collecting data from the vehicles [and operators themselves]. That way, we can make the product better, help our customers to improve business operation and gain more market share as well."

One of the group's key areas of focus is data analytics; unstructured data, a historically underutilized resource, is now playing a key role in helping the company understand and evaluate past performance and drive future product design. "Whenever a machine has problems," he explains, "the customer running the machine calls our service center. [Oshkosh] technicians pick up the call and tell the customer what to do – and what not to do – in terms of structuring procedures." Service calls are conversational, and Oshkosh technicians take notes to ensure that the particulars of any incident are recorded. While unstructured text data provides a level of detail unparalleled by more quantitative alternatives, it has long been seen as impractical for use in aggregate analyses. The problem? Free-form text data is just too messy.

"We might have three or four months of service data with tens of thousands of entries," Zhang explains. "Different individuals using different terms are creating these records. Sometimes there are typos and other problems. There are a lot of challenges [stemming from] people using different terms to mean the same thing, for example." Regardless, Zhang and his team have now adopted new statistically driven methods with which to make sense of all the noise. With text mining, free-form fields of text can be parsed to identify frequently used words and phrases, and then to analyze these phrases in aggregate to elicit enhanced meaning and other practical inferences. This process would be excessively tedious if it weren't for JMP.

From free-form text to useable data clusters

Zhang and his team "clean" free-form text by grouping similar terms together and excising words lacking in specificity or relevance. "We use tools in JMP like recoding where you have similar meanings for different terms," Zhang says. "We can easily combine those [entries] to make a single data point. Then we also look at the tokenizations – getting rid of stock words or punctuation. Once we clean the data set, then we have some information that we can start to work with."

There are many ways to cut the data, Zhang explains: model categories, components, parts, etc. But Zhang's team chose to focus on fault codes – that is, predefined service problems and solutions. With JMP Text Explorer, Zhang performs latent class analysis to group records into clusters. This process requires some level of interpretative decision making, and the interactive nature of Text Explorer is an asset in enabling users to organize words visually, create summaries and extract terms.

By combining data insights with the engineering know-how the Oshkosh team already has at hand, Zhang's team has generated intelligence to shape future troubleshooting service calls, and in particular to make the customer support process more consistent and efficient. For one, an analysis of fault codes has helped deepen engineers' understanding of mechanical issues and advance the company's growing incident resolution knowledge library. Perhaps even more importantly, Zhang was able to characterize solutions by order of likelihood.

"We identified a lot of improvement opportunities for the folks who do the service calls and who keep the records of these calls," he says. "We made the process better. I hope that in the next round of data, when it comes in, we will have much more refined granularity and better recorded terms that we can easily use."

By combining data insights with the engineering know-how, Zhang's team has generated intelligence to shape future troubleshooting service calls, and in particular to make the customer support process more consistent and efficient.

Improving customer service, informing R&D priorities

Oshkosh has also been able to use the insights gained from service call data to redefine priorities for engineering improvements in such a way as to prioritize the most significant issues. "When you look at compact records combined with the cluster terms or cluster tokens, you can generate several categories for future troubleshooting," Zhang explains. "That really intrigued us, so we continued doing more analysis and many other knowledge-base expansion exercises based on the information contained within our customer service records."

They are already able to generate valuable insights that are shaping Oshkosh Corporation's customer-service response, depending on the location, product, length of service and so on; the data will preempt the solutions based on

Ron Zhang

statistical probability. The next step for the team is to "expand this knowledge base for future debugging and troubleshooting," says Zhang. "There have already been some new design ideas generated from our insights."

Intuitive data visualization

While Zhang's team once relied on a combination of Excel for data processing and other tools for scripting, Zhang has seen a dramatic increase in capability with JMP. "With JMP you don't need to know how to script. You can use the role selections, and some of the really convenient features that you can sort in, do descriptions, analysis, then pick up that one you want to extract," he says. "Another JMP feature that we use a lot is Graph Builder. If you wanted to use Excel or something else to generate the report in multidimensional graphics, it would be too hard. But with Graph Builder, you can just drag and drop and then push a couple of buttons and the data appears for you. Those are some of the best features in JMP."

For Zhang's team, the return on investment of upgrading to JMP Pro was also tangible: "Our data sets are not huge at the moment, but we expect we (will) have to scale up. We use the presentation and visualization tools to help present data to customers and other functions in the corporation by leveraging the full potential of text mining to help with the business." Zhang is already proving that his text-based insights can be used proactively to preempt certain types of incoming calls with more pertinent information available to operators, as well as guide future engineering efforts to mitigate common machine problems in the first place. And when time is of the essence – as it almost always is in the tactical vehicle industry – even seemingly small improvements can really move the working world.

Online

Tackle your unstructured data with text exploration:
jmp.com/text

Oshkosh Corporation designs and builds some of the world's toughest specialty trucks and access equipment:
oshkoshcorp.com

Learn JMP®

JMP training and books are designed to strengthen your JMP skills.
Supercharge your potential.

Grow with us
jmp.com/training

sas®

100 SAS Campus Drive
Cary, NC 27513 USA

DISCOVERY SUMMIT

EXPLORING DATA
INSPIRING INNOVATION

jmp Statistical Discovery™ From SAS.

Discovery is universal

No matter where you are in the world, the thrill of discovery is the same, as is the desire to celebrate data exploration and innovation. We're in Europe every March, in China in April, in the United States in October and in Japan in November. Tell your international friends and colleagues to find a Summit near them. Or contemplate submitting content for consideration by one of these steering committees. We believe in sharing analytic best practices – everywhere in the world.

Learn more about the Discovery Summit
discoverysummit.jmp