

Death by Statistics

“Statistics cannot substitute for clear thinking. It can’t do the job of human inductive inference.”

. . .

I’ll admit it up front, statistics has never been my forte. I belong to a generation that was poorly educated on the topic. The courses I took focused on frequentist probability. The lectures were aimed at giving us tools for generating publication worthy p-values rather than interpreting data to understand natural phenomena. This cookbook approach came across as unsatisfactory and counter-intuitive to the budding scientist I was, especially once I started generating my own experimental data and came to realize how messy biological experimentation and data can be.

These days, I rely primarily on expert colleagues for their guidance. But I can deal with data much better than I used to. I also understand better what my job is about. My goal as a scientist is to produce knowledge that yields predictable outcomes, and my obsession isn’t with p-values but with reproducibility. Nothing beats controls and replication, especially when orthogonal replication with a different method independently validates a finding. I’m not going to build my reserach program based on a single experiment with borderline p-values. I keep steering my lab away from shaky findings, and over and over again, I have resisted the temptation of becoming enamoured with weak models no matter how exciting they were—or how significant the p-value is. I can now confidently report that this approach has served our research team quite well.



“Data don’t make any sense,
we will have to resort to statistics.”

Via The LoveStats Blog

The Frequentist vs. Bayesian statistics non-debate

I only became aware of the frequentist vs. Bayesian debate in statistics late in my career. When I finally learned about Bayesian statistics, I was shocked that I have never been exposed to it in my statistics courses, not just the mathematical expression of Bayes theorem but also the philosophy of Bayesian reasoning. This is perhaps due to the hold that frequentists have had on applied statistics or the philosophical vacuity of modern science teaching.



Sophien Kamoun @KamounLab · Jan 29

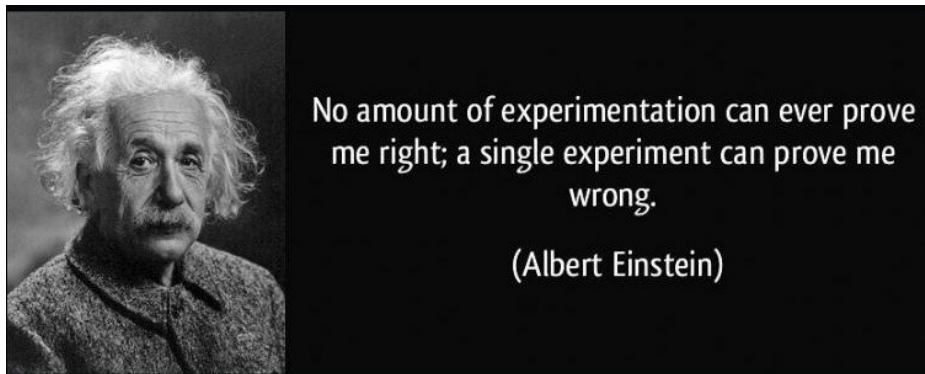


I once coauthored a Bayesian **statistics** paper
bmcsystbiol.biomedcentral.com/articles/10.11...



Yours truly among Bayesian statisticians.

I can't claim that I fully understand Bayesian statistics, but I understand enough to say that its reasoning aligns with the ideals that I want to reach as a scientist. The Bayesian view—adjusting our beliefs as new information comes to light—makes more sense to me than frequentism, with its focus on the significance of single experiments in the absence of context. More than anything, it's this philosophical approach to certainty that wins me over—the view that one single experiment can't reveal the truth, but that we need to continuously build and challenge prior knowledge to unravel the reality of the natural world. Indeed, the toughest challenge we can apply to current knowledge is to make predictions and test them. Only then, can we start converging towards some sort of scientific consensus (see the discussion below of the concept of Bayesian convergence).



Modern science training needs more philosophy of science.

In recent years, there has been calls to ditch teaching frequentism to non-statisticians. Data dredging or p-hacking—the practice of manipulating data until it becomes statistically significant—is rife. Authors and journals alike have no shame in endorsing these dodgy practices. Nature Springer (aka [#NatureRipoffs](#)) journals like Nature Communications charge hefty publictaion fees for their supposedly stringent editorial and peer-review service, but can't filter out papers in which authors proudly claim: *“we continuously increased the number of animals until statistical significance was reached to support our conclusions”*.

Sophien Kamoun
@KamounLab · [Follow](#)

This is the peer-review service you get for \$5200. **#NatureRipoffs #PredatoryJournals @NatureComms**

Daniel Padfield @padpadpadpad

Wait. Wat. Wut. 🤔🤔🤔🤔🤔🤔🤔 Published in @NatureComms... [nature.com/articles/s4146...](https://www.nature.com/articles/s4146...)

This is the peer-review service you get for \$5200. #NatureRipoffs #PredatoryJournals

At least these authors reported their questionable practice. For many others, it's probably such a routine practice that they don't feel the need to report it. Can we blame them when statisticians themselves have been

teaching frequentism with quasi-religious zeal. As [William M. Briggs wrote in a widely debated 2012 arXiv article](#):

“We should cease teaching frequentist statistics to undergraduates and switch to Bayes. Doing so will reduce the amount of confusion and over-certainty rife among users of statistics.”

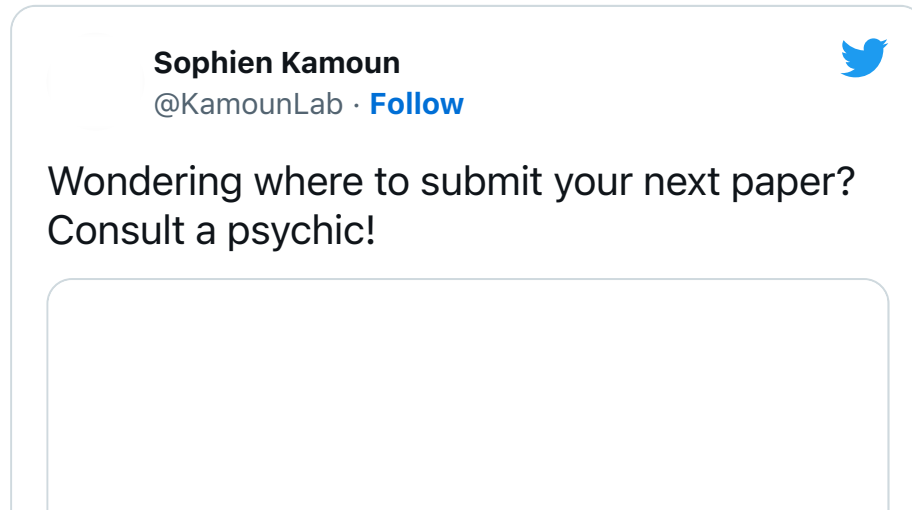
Psychic powers

Predictably, frequentist probability ended up being abused in many quarters of science, sometimes to comical effects. There is plenty of nonsense hiding as serious science behind “statistically significant” p-values. According to [a 2010 article in the Journal of Zoology](#) (2020 Impact Factor = 2.322), the common toad *Bufo bufo* is said to have pre-seismic anticipatory behavior, or in plain English to predict earthquakes. This isn’t science.

And according to the scientific literature, psychics are for real. In 2013, [Science Magazine](#) discussed the reaction to the publication in a “*top psychology journal*” of a paper claiming extrasensory perception (ESP)—psychic powers in street English. [As the Science commentary states](#):

“[The publication of the ESP paper] has rekindled a long-running debate about whether the statistical tools commonly used in psychology—and most other areas of science—too often lead researchers astray. “The real lesson to be learned from this is not that ESP exists, it’s that the methods we’re using aren’t protecting us against spurious results,” says David Krantz, a statistician at Columbia University.”

Think about this for a moment. *“The statistical tools too often lead reserachers astray...”* Have scientists become so naive that they blindly follow a set of tools, statistical or otherwise, without much critical thinking and without digging deep to challenge their findings with independent methods and approaches? As David Kranz is quoted as saying in the [Science piece](#), *“no statistical method can safeguard completely against erroneous results, and none can substitute for clear thinking. One can’t expect statistics to do the job of human inductive inference”*.



Consult a psychic! It works according to the *Journal of Personality and Social Psychology (JPSP)*, 2020 Impact Factor: 7.673.

Statistics—the convenient alibi

The issue is at the essence of what we want to achieve as scientists. For too many of us, we simply lost the plot. The goal somehow has drifted from producing robust knowledge that stands the test of time to getting whatever piece of work published in the best possible journal. This cynical attitude is at the heart of many of the chronic problems of academia. Statistics becomes a convenient alibi to convince editors and reviewers that a paper should be published. [#DeathByStatistics](#) endorsed by academics.

You know we have a serious problem when the *Journal of Personality and Social Psychology (JPSP)* publishes hocus-pocus as science, and yet is still viewed as a “top journal”. Oh wait, I get it, JPSP’s 2020 Journal Impact Factor (JIF) is 7.673. That’s what’s defines a top scientific journal in the eyes of many. Whether or not it publishes tabloid level nonsense is irrelevant.

And what’s the Journal Impact Factor by the way? Just another flawed statistical metric enthusiastically endorsed by a number of scientists and academic institutions. As Imperial College Professor Stephen Curry wrote in 2012, “The impact factor is a statistically indefensible indicator of journal performance; it flatters to deceive, distributing credit that has been earned by only a small fraction of its published papers.”



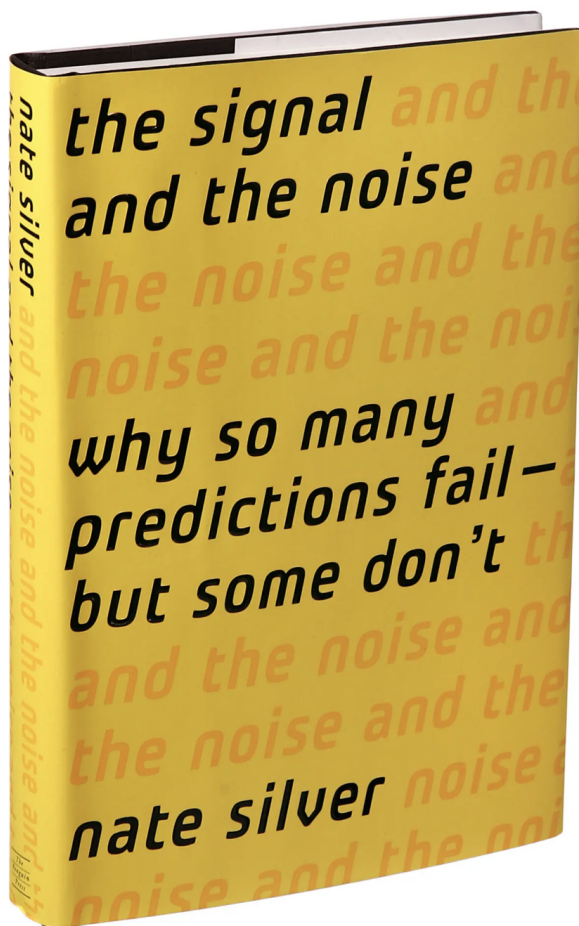
Death to the impact factor

Crude statistically flawed metric
Skewed by few highly cited papers
Misused and abused

Death to the impact factor — a meaningless and statistically flawed metric. (image source: the conversation)

Bayesian convergence—to be less and less and less wrong

My enthusiasm for Bayesian statistics was boosted by reading Nate Silver's superb book *The Signal and the Noise: Why So Many Predictions Fail—but Some Don't*. Nate, the founder of the popular website FiveThirtyEight, has made a name for himself in forecasting, from baseball to US elections. His book goes at the heart of what I consider the most useful contribution of science, *its capacity to predict outcomes*. This is what differentiates scientists from charlatans. What best defines the boundary between science and pseudoscience.



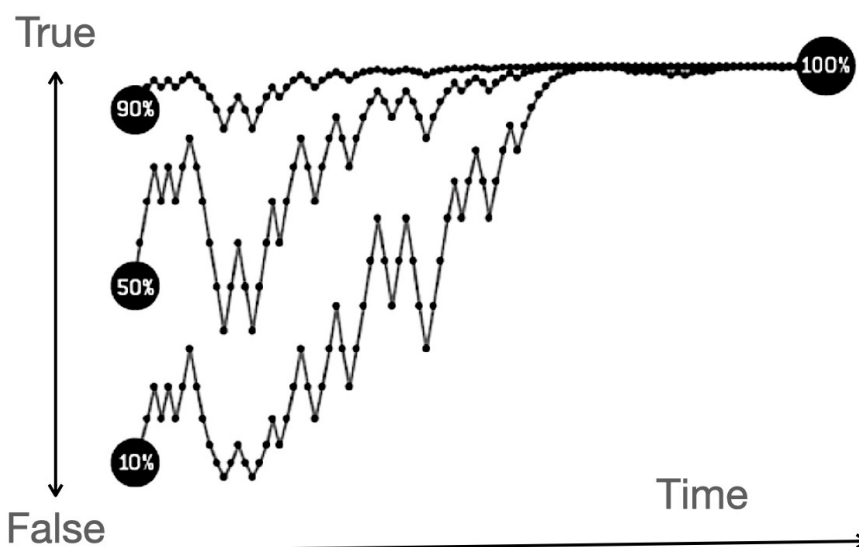
Alessandra Montalto/The New York Times

The Signal and the Noise. (source: New York Times)

One concept I intuitively relate to is Bayesian convergence, which in practical terms means that no matter where our prior probabilities stand, they will get revised up or down once new information becomes available. This is how good science should work in practice. What matters most is that, as evidence accumulates through experimentation and replication, we converge towards a scientific consensus. This is another way of saying good science stands the test of time. That new knowledge forces us to revise our view of the world.

For example, scientists may have had different levels of skepticism when the CRISPR gene editing methodology was first reported in 2012 by Nobelists Emmanuelle Charpentier and Jennifer Doudna, but by now there has been enough replications for a wide consensus to emerge—the method does work as advertised (nearing the 100% certainly in the graph below). This is Bayesian convergence. No matter where we stood with our posterior beliefs (10%, 250% or 90% in the graph below), at

some point we start to converge towards a consensus. Indeed, it is now widely accepted that CRISPR is a robust method and that the associated knowledge is solid—a scientific consensus. Indeed, you can design a CRISPR experiment with quasi-mathematical certainty. You target a particular sequence of DNA, and bingo the CRISPR method will generate mutations at precisely that location in the organism’s genome. The prediction works over and over, and has been validated by hundreds of laboratories throughout the world. This is science at its best.



Bayesian convergence as illustrated by Nate Silver in the Signal and the Noise. Adapted from the lecture notes of Danilo Freire.

What is beautiful about the concept of Bayesian convergence is how aligned it is with the scientific method, where the idea of unravelling the reality of the natural world through observation and experimentation is more often than not a “path to less wrongness”, as Nate Silver wrote, rather than through a revelation of the truth. The journey towards scientific truth is a path to being less and less and less wrong.

Nate Silver quotes in his book a beautiful poem by Danish mathematician Piet Hein: “Err and err and err again but less and less and less.” THIS is our fate as scientists.

„Err and err and err again but less and less and less.“

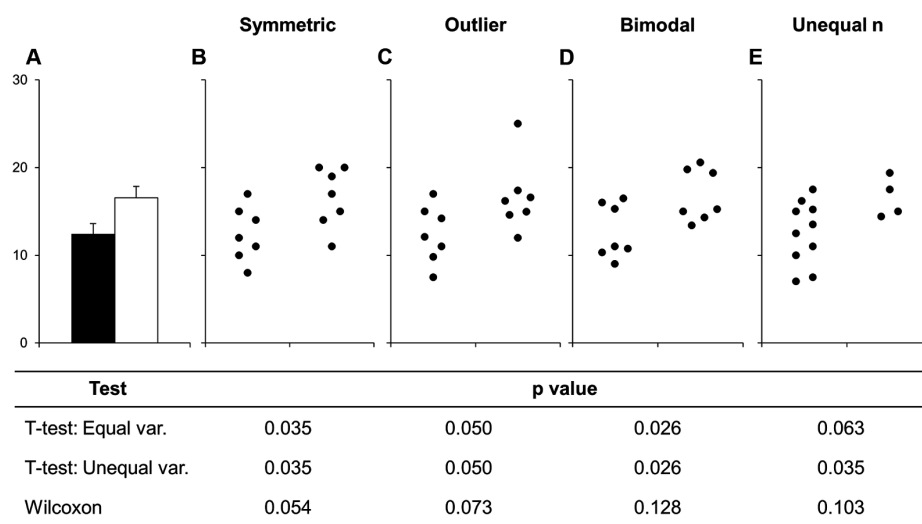
– Piet Hein

The Road to Wisdom?

Piet Hein. The Road to Wisdom?

#DeathByHistograms

There are praiseworthy efforts in the community to improve data reporting and analysis. A popular effort calls for ditching bar and line graphs. Its manifesto is a [PLOS Biology article by Tracey Weissgerber and colleagues](#) “*Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm*”. The article has gathered >400,000 page views since its publication in 2015. By all odds, you probably have read it by now. But if you haven’t, please read it now! I mean after finishing reading my post of course 😊.



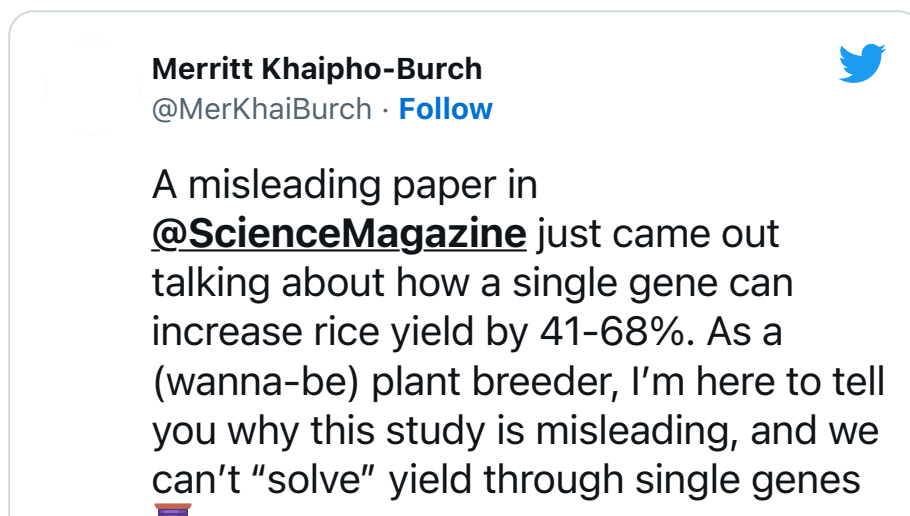
Different datasets. One bar graph. This clearly shows how histograms do not reveal the nature of the underlying data.

The Weissgerber et al. article should be required reading for all scientists, from undergraduates to established Professors. And you’ll be

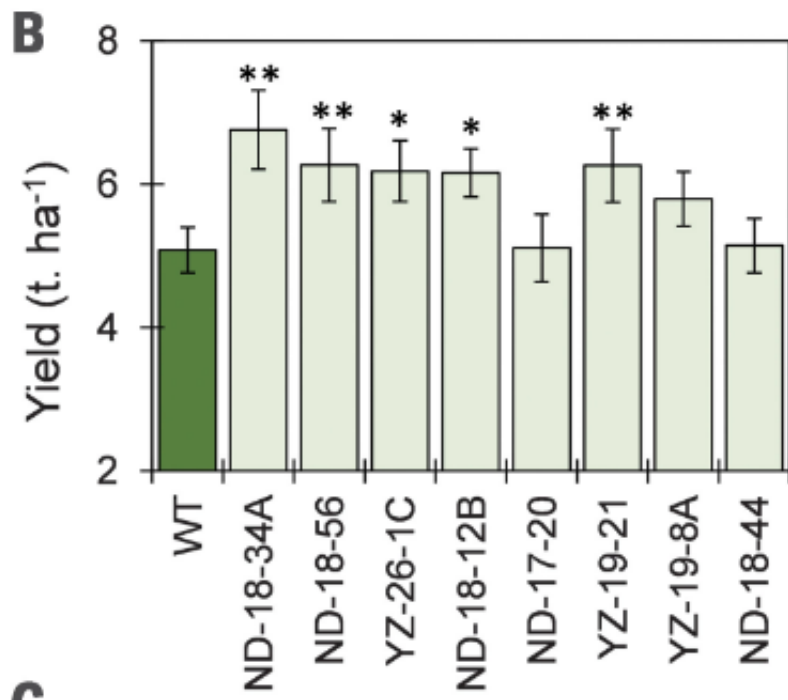
forgiven to think that the editors of the most prestigious journals, like the infamous Cell/Nature/Science (CNS) triad of glam-mags, must be among the >400,000 readers. Surely, the news of misleading histograms and the paradigm shift towards scatterplots must have reached the gatekeepers of the high-profile articles that show up in these journals, alter scientist careers, and frequently end up widely publicized in the popular press.

Apparently not. Science Magazine, presumably the leading light of the scientific community as the flagship journal of the AAAS—the American Association for the Advancement of Science—still regularly publishes histograms when scatterplots would be far more appropriate.

A pair of Science articles recently gained much attention with their sensational claims about massive increases—in the 30 to 40% range—in crop yields. My aim here isn't to address technical issues about this work. This has been expertly peer-reviewed after publication by Cornell University Ph.D. student Merritt Khaipho-Burch. You can read her pertinent Twitter threads about the two papers here and here.

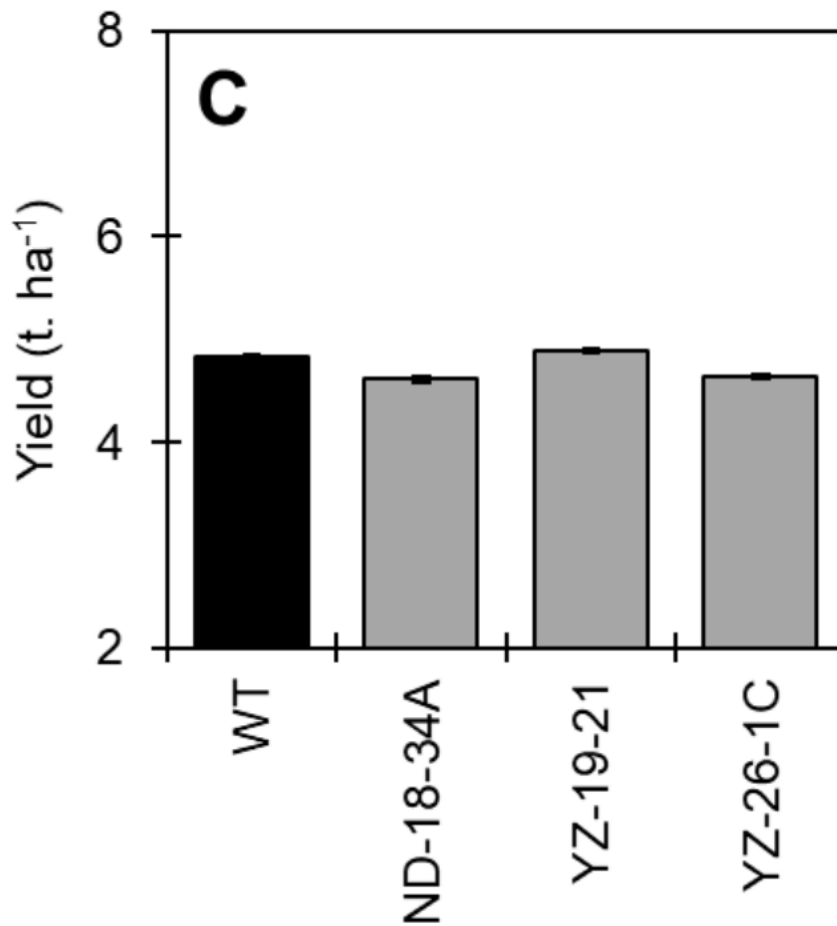


What struck me with these two papers is the repeated use of bar charts (aka histograms) when the data could have easily been displayed as scatterplots—the state-of-the-art norm these days in many labs and journals. In addition, in this one example, the results were only marginally significant (** = $p < 0.05$) as Merritt tweeted. You would think that, we the readers, deserve better from authors and editors alike. I mean, we're talking 2022, 7 (Seven!) years after the Weissgerber et al. call for a paradigm shift in data presentation.



** = p < 0.05.

I could stop there, but let me share with you another beauty. Notwithstanding any biological interpretation of the data displayed in this figure, I think, we the readers, once again deserve much more respect than having our intelligence insulted with this underwhelming histogram. A better display of the data underpinning the bar chart is badly lacking. Is the data of the symmetric, outlier, bimodal or unequal type? Is it acceptable to publish such bar graphs in 2022?



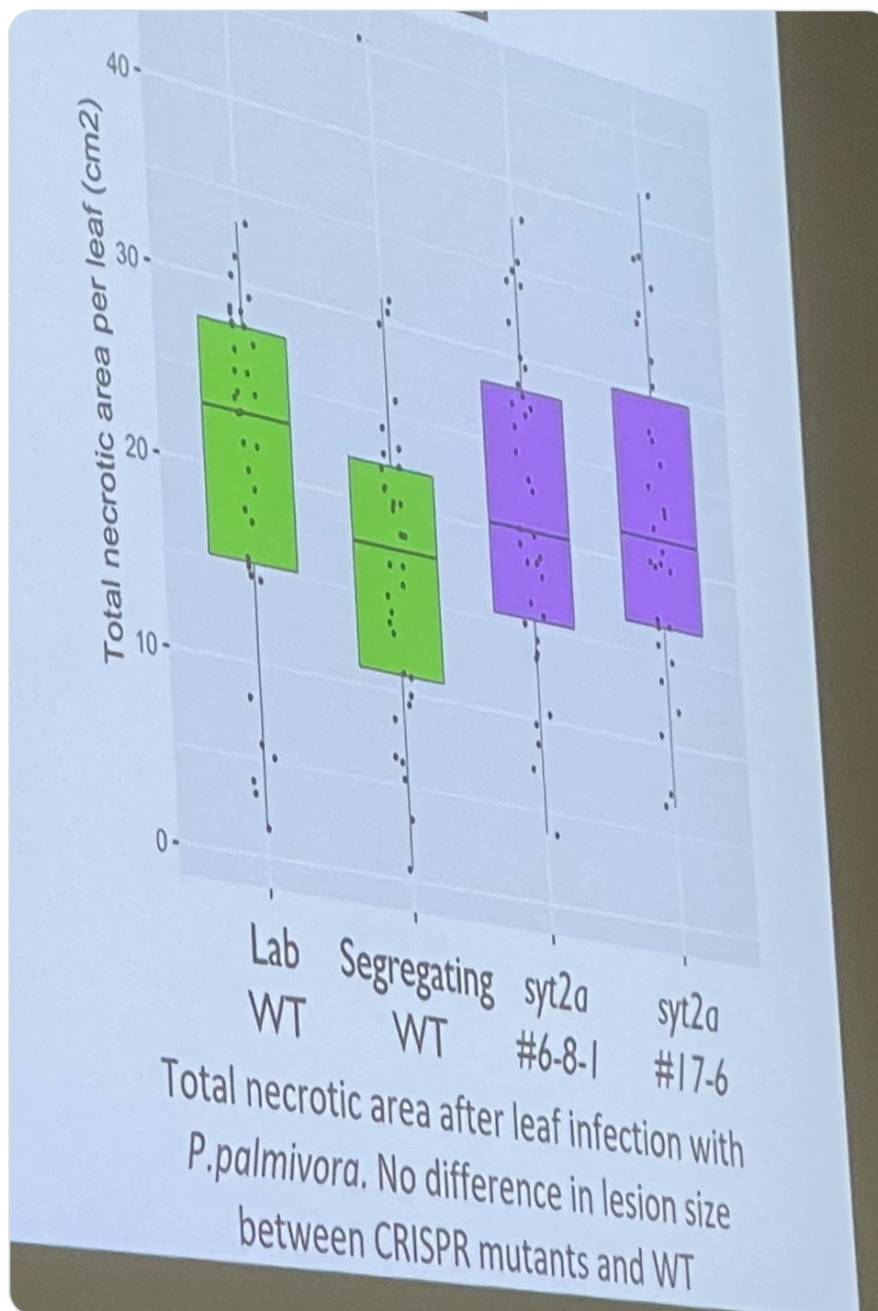
Death by histogram.

Contrast the Science bar graphs with the scatterplots displayed by Cambridge PhD student [Alex Guyon](#) at a recent conference. Not only did [Alex and colleagues nicely illustrated the variation they noted in their pathogen virulence data](#), but they also included more than one negative control avoiding any experimental bias due to an outlier control treatment. Alex's conclusion was simply that the results were inconclusive. But the scatterplot immediately conveys to us the type of variation they observe with this type of experiment. Alex and colleagues could have easily fooled us with a histogram. They chose science over unethical gamesmanship.

How data should be displayed. And multiple control treatments too. Don't lie with statistics.

#deathtohistograms #inconclusiveresults

Well done @AlexWGuyon @dromius #OMGN22



How data should be displayed.

As Imperial College plant biologist [Pietro Spanu](#) tweeted: “Anyone who has done these sorts of experiments knows what they look like. Displaying data like this, reflects reality.” This goes at the heart of the matter.

Uncovering reality. What the business of science should be all about.

Pietro Spanu

@PietroSpanu · [Follow](#)



(Leaving aside controversies about yield)

This is an excellent example of REAL LIFE variation seen in pathogenicity/virulence/symptom assays.

Anyone who has done these sorts of

Develop your data presentation skills

I call all of you to support [Weissgerber et al. call to ditch bar graphs/histograms and embrace a new data presentation paradigm.](#)

Your first step is to develop basic skills in R, and explore tools like ggplot that help you analyze and display your data. Here is some useful links on this topic from the training that [Dan MacLean](#) and team offer to newcomers to The Sainsbury Laboratory:

[ggplot2 scatter plots : Quick start guide.](#)

[Using ggplot2 for producing plots.](#)

[A cookbook on creating many types of plots in R—ggplot-centric.](#)

[The BBC's data journalism cookbook.](#)

[A general theory book on data visualisation.](#)

[A package for creating figures from multiple plots.](#)

[A set of palettes that are colour-blind friendly and perceptually uniform.](#)

[A set of palettes inspired by the outfits of a popular musician.](#)

[Rstudio cheatsheets.](#)

Acknowledgements

I'm thankful to many colleagues for discussions and insights on this topic. Thanks to Alex Guyong for inspiring me to write the post.