What Does it Mean?

Susan Kemper, Roberts Distinguished Professor of Psychology, University of Kansas

Web of Science has weighted, measured, counted, and aggregated, and … I am a 31! What does it mean? According to ResearchGate, I am 34.83 (note that illusion of measurement precision) – I got a real boost by acquiring a well-connected colleague as a "follower" since followers with high scores impact your score. I might also be a 39 or maybe a 91-- Google has its own system of metrics.

What does it mean? It all depends…but on what? There is now a vast literature on these impact statistics, critiques, applications, revisions, alternatives: they are often criticized for their "age" bias favoring "old" but it is more accurately a "time since degree or time in profession" bias; most of the indices assess quantity over quality because critiques and refutations can contribute to high impact scores; they are plagued with citation 'gaps,' inaccuracies, and distortions; they are affected by publication practices such as page limits, citation limits, and publication lags; they reflect peer review's confirmation bias; they can be difficult to compare across scientific fields due to field-wide practices affecting citation density and recency; the expanding inventory of journals, the emergence of ejournals, pay-to-publish journals, and "predatory open-access journals" have introduced new complications; and most of the statistics suffer from various distributional biases since they overlook or ignore skew, variability, and kutosis.

But these critiques notwithstanding, we have now ratcheted up counting, measuring, and weighing, aggregating faculty scores into scores for departments, schools, and universities. These inherit all of the flaws of individual-level impact statistics and add a few more: the weights assigned to variables that are aggregated; the variables themselves and whether, which, and how books, grants, and academic/scientific honors are counted and weighted; the unit's mix of undergraduate, masters, doctoral, and assorted professional degrees; the balance of full-time vs. adjunct faculty, and teaching vs. research faculty; limitations on the available data for interdisciplinary or multidisciplinary programs; and the lack of relevant outcome measures to gauge the actual "impact" of research and graduates.

Access to the "gold standard" of "Academic Analytics" is limited by a number of confidentiality agreements but *The Chronicle of Higher Education* (2007) did once offer a peak at some of its data. I looked at psychology programs. Well, first I looked at the "general psychology" classification – Stanford is on top, but note only 7 faculty members contributed and the department website lists 32. But there's also a "various psychology" classification – IU leads that list. And there's

a "clinical psychology" list – here KU shows up with 30 clinical faculty so I guess they were considering me to be clinical. And then there's also a "cognitive science" list where IU shows up again, now with 56 faculty members.
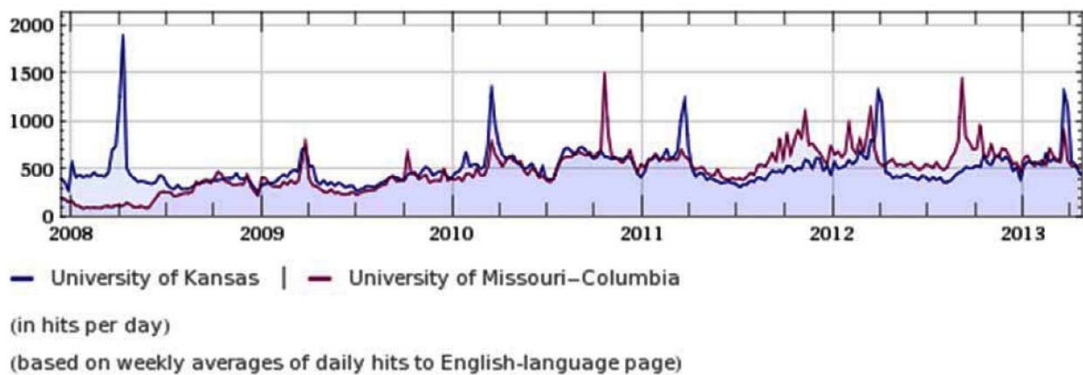
So what does it mean for a program to be 1.36 or a 2.06? Well, I constructed a box and whisker plot, leaving aside the cognitive science data. And here we are: KU falls just outside the box defined by the interquartile range, bearing in mind all of these programs were at least 1 SD above the overall mean for all programs assessed.

Psychology Programs
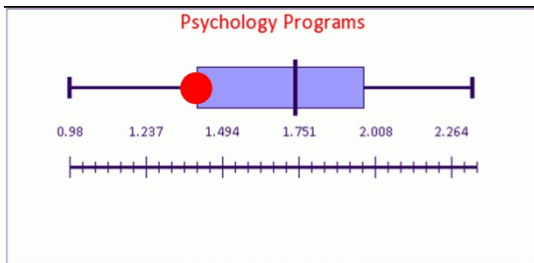
What does it mean? The great promise of analytics is that benchmarking – faculty members, departments, universities, - will lead to wise strategic decision-making. My question is "what does it mean" to see "every variable in each academic discipline …[and] national quartile, quintile, decile, and vingtile summaries.." (Academic Analytics, 2013)?"

Wolfram Alpha (2013) trolls lots of electronic data and lets one compare, well, nearly anything and everyone including universities. So here's a comparison of KU and MU using Wikipedia hits per day. What does it mean? The answer probably has something to do with football and basketball but that's just my guess, in this case derived from the seasonal periodicity of the spikes. That is, given a theory of what determines Wikipedia hits, it tested that theory against this data by, e.g., looking at win/loss records in football and basketball.

Wolfram himself published detailed visualizations of his email history (Wolfram, 2012). What do they mean? In his cases, it seems obvious: When he is sleeping, he is not emailing – and his email vs. sleep cycle was affected by a 2009 trip to Europe! And his use of email is increasing. But I would argue that these data and the graphic displays of this data itself provide few insights into Wolfram's personal history. Surely he knew he was unlikely to be sending emails in the middle of the night; surely he knew he was sending more and more emails every day? The graphs provide a visual confirmation and don't seem to themselves to trigger new insights.

University of Kansas | University of Missouri–Columbia

(in hits per day)

(based on weekly averages of daily hits to English-language page)

Indeed, Wolfram produced a distribution of the number of emails per day only to conclude "What is this distribution? Is there a simple model of it? I don't know. Wolfram Alpha Pro tells us that the best fit it finds is to a geometric distribution. But it officially rejects that fit. Still, at least the tail seems – as so often – to follow a power law. And perhaps that's telling me something about myself though I have to say I don't know what." (Wolfram, 2012).

Perhaps that is because appropriate benchmarks are lacking – would greater insight into Wolfram's life be provided by knowing where he stacks up in terms of "quartiles, quintiles, deciles, and vigintiles" of all email users? Probably not. The real challenge is to move beyond descriptive analytics. Even comparative analytics don't really answer the right questions. In Wolfram's case, the visualizations at any level of aggregation don't suggest how he might more effectively manage his email correspondence, whether the volume of email is negatively (or positively) impacting his productivity, or even affecting his sleep cycles. This data might provide a baseline against which to compare interventions but the data themselves do not tell him whether or how to intervene.

What does it mean? Suppose it is an ex-Gaussian distribution? Ex-Gaussian distributions result from a convolution of a normal distribution and an exponential function. They can be modeled as 3 parameters: one for the mean (*mu*) (the peak), one for the standard deviation (*sigma*) (the variability or spread), and an exponential (*tau*) (for the tail). Ratcliff (1979) has mapped these 3 parameters onto specific cognitive processes that determine the speed (but not accuracy) of decision making and he and others have investigated lifespan developmental differences in decision making: children tend to more variable, affecting the *sigma parameter*, than young adults, older adults tend to be slower (*mu*), more variable (*sigma*), and more extreme (*tau*).

So if Wolfram's email distribution is an ex-Gaussian, what interpretation might we attach to these 3 parameters, *mu*, *sigma*, *tau*? Would knowing his average daily burden of emails (*mu*), the variability of his email traffic (*sigma*), or its extremes (*tau*) affect strategic investments in, e.g., network speed? My point is not that the data and its visualization are irrelevant but that they must be coupled with an explanatory theory –of reaction times, of emailing, or of faculty productivity.

Knowing how individual faculty members, departments, or universities stack up on various metrics – those "quartile, quintile, decile, and vigintile" comparisons - doesn't really provide answers to how productivity can be enhanced or sustained. And I think we are distracted by the logistics of compiling all this data and generating the fancy graphics, apps, visualizations.

What does it mean? One guy who does seem to be able to answer this question is Ed Tufte. Tufte is an emeritus professor of political science from Yale who founded his own publishing company to produce a series of books on graphical design and analysis. Along with others like Stephen Few, Nigel Holmes, and Nathan Yan, he has created a new discipline

of data visualization. Tufte's scathing critique of PowerPoint (Tufte, 2003) should be mandatory; its low resolution leads to over-generalizations, imprecise statements, slogans, lightweight evidence, and thinly-argued claims; bullet outlines make us stupid by omitting critical relationships in favor of a 1-dimensional ordering; and the reliance on projected slides reduces information transmission to a few words, lots of "phluff" – white space, cartoons, bullets, frames, data-thin graphics.

Tufte coined the term 'chartjunk.' He has also suggested some general principles of data visualization (Tufte, 2001). Tufte has provided few general principles: avoid chart junk, maximize the data ink to total ink ratio, and employ small multiples. Tufte offers other principles for informativeness by creating multiple layers of information. Tufte is known for 2 types of graphics he introduced. Both exemplify his principle of maximizing data-ink. One graphic Tufte developed is the slopegraph – Tufte's riff on scatter plots – to relate to scalar variables.
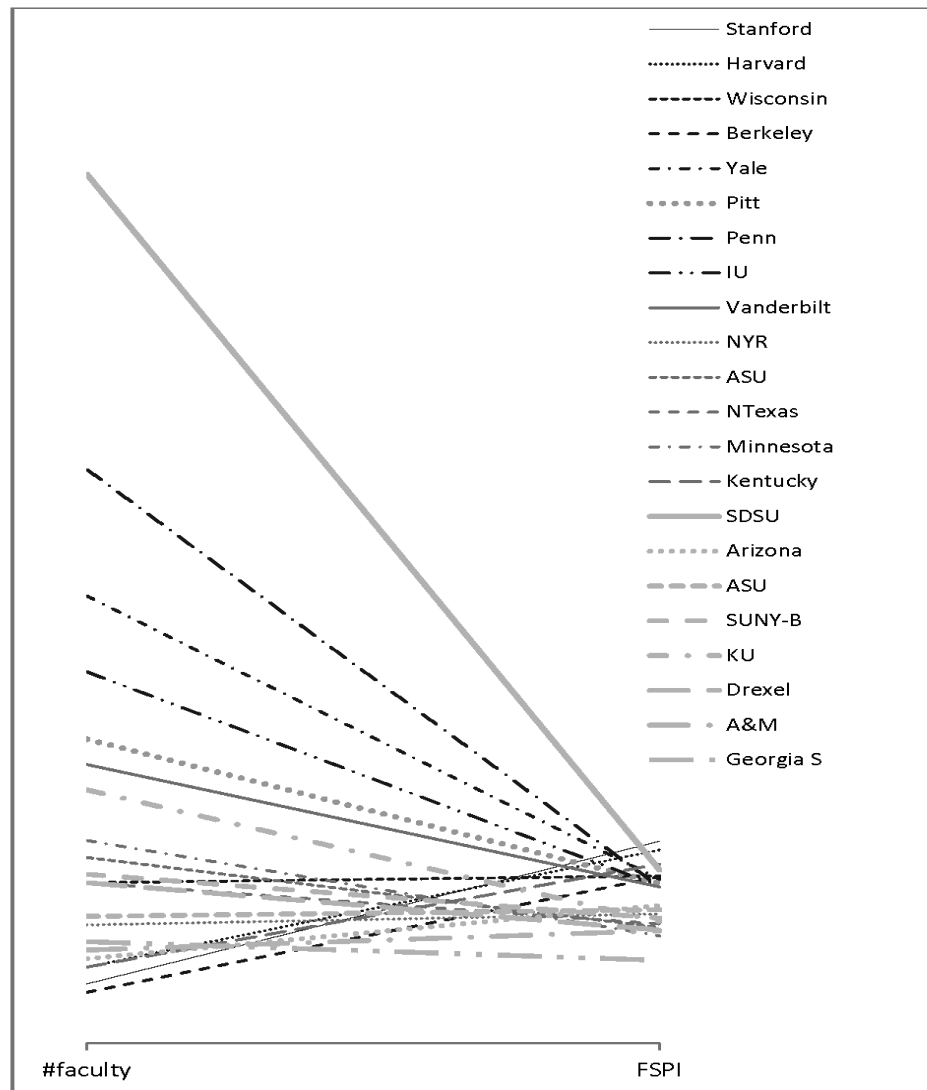
I constructed a slopegraph using the *Chronicle* FSPI data. Although the Index is supposed to be normalized for faculty size, we can clearly see that the overall trend is for larger departments to have somewhat more productive faculty – suggesting size confers some "synergies" perhaps by distributing administrative duties more widely. The slopegraph reveals 2 patterns embedded in the overall pattern: a cluster of programs showing diminishing returns with increasing faculty size and a cluster that seems to indicate "smaller IS better." So the challenge would be to understand whether there is a critical or optimal size for an administrative unit- one that promotes productivity of individual faculty.

more years. I think we can identify a number of factors that contribute to early career peaks, factors that favored my cohort and disadvantaged those from earlier and later cohorts: a higher level of state support and annual merit salary increases that rewarded research productivity; an expanding university, one creating many new interdisciplinary programs built around research areas; smaller class sizes, more GTA support, and fewer "ancillary" obligations such as building and maintaining websites, supervising students engaged in service learning, and devising ways to incubate and transfer technologies.

At the 2001 Merrill Retreat on "evaluating research productivity," I turned to

impact / date of degree
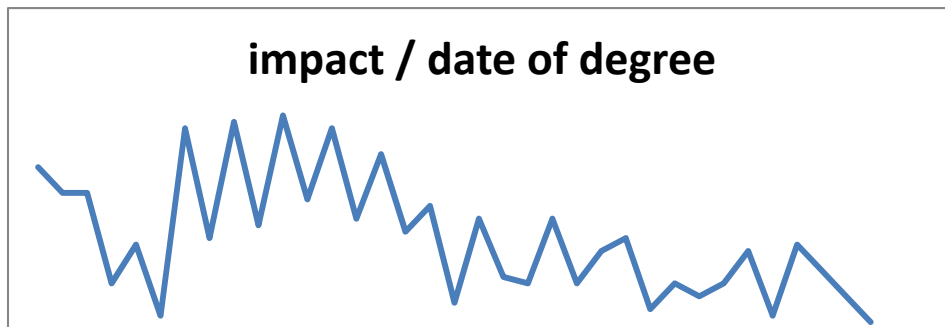
The other Tufte graphic is the sparkline. A sparkline is a simple trace of one variable against a second, usually on a scale like time. I also constructed a sparkgraph for my colleagues in Psychology – plotting Web of Science impact scores over time since degree. It reveals 3 clusters or cohorts, defined by their impact scores. My hope is that this sparkline might spark some ideas about the factors that contribute to these 3 clusters. In my 2010 Merrill talk, I pointed out that highly productive faculty tend to peak early in their careers and to sustain that level of peak performance throughout their career, even careers that span 30 or

some sage advice from 1897: Cajal (1999) recognized 6 impediments to faculty productivity – what he termed "diseases of the will:" the dilettantes or contemplators; the erudite or bibliophiles; the instrument addicts; the megalomaniacs; the misfits; and the theory builders (p. 75)." He is most dismissive of the contemplators as "likeable for their juvenile enthusiasm and piquant and winning speech as they are ineffective in making any real scientific progress" (p. 77) and he recognizes that "cold-hearted instrumental addicts cannot make themselves useful"(p. 82) and he labeled the misfits, who occupy a professorship "simply to collect

the salary, and to enjoy the incidental pleasure of excluding the competent" (p. 82-83), as "hopelessly ill" (p. 82). For the rest, Cajal has some recommendations regarding promoting research productivity that ring as true today as they did in 1897 or in 2001. Cajal reminds the bibliophile that "We render a tribute of respect to those who add original work to a library, and withhold it from those who carry a library around in their head" (p. 78). He advices the megalomaniac to "tackle small problems first …[an approach which] may not always lead to fame but [to] the esteem of the learned and the respect and consideration of our colleagues" (p. 80). He notes that rather than bemoaning the lack of able assistants, or laboratory equipment, or government funding, that "dreamers do not work hard enough" (p. 80). And he reminds the theorist that "Theories desert us, while data defend us" (p. 86).

Cajal cautions that independent judgment, intellectual curiosity, perseverance, and concentration the keys to productivity. Beyond these prerequisites, Cajal emphasizes that research productivity results from a "passion for reputation, for approval and applause," and a "taste for originality, the gratification associated with the act of discovery itself". These are the real determinates of faculty productivity. Analytics, no matter how aesthetically plotted as "quartile, quintile, decile, and vigintile summaries" do not assess this "passion for reputation" and this "taste for originality." That's what it means – to be productive, to have an impact.

References:

Academic Analytics. (2013). Retrieved from http://www.academicanalytics.com/Public/WhatWeDo

Cajal, S. R. y (1999). (Translated by N. Swanson and L. W. Swanson). *Advice for a Young Investigator.* Cambridge, MA: MIT Press.

Chronicle of Higher Education (2007, July 22). Faculty Scholarly Productivity Index. Retrieved from https://chronicle.com/stats/productivity/page.php?bycat=true&primary=8&secondary=65&year=2006

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86*, 446-461.

Tufte, E. (2001). The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

Tufte, E. (2003). The Cognitive Style of Powerpoint: Pitching out corrupts within. Cheshire, CT: Graphics Press.

Wolfram Alpha. (2013). Retrieved from http://www.wolframalpha.com/

Wolfram, S. (2012, March 8). The personal analytics of my life. Retrieved from http://blog.stephenwolfram.com/2012/03/the-personal-analytics-of-my-life/