Down the Victorian data mine

We tend to think of data mining as having originated in the late twentieth century, prompted in large part by e-commerce. But, here, **James Hanley** and **Elizabeth Turner** tell the story of an eighteenth-century customer loyalty programme later "quarried" for a 1884 article whose statistical ingenuity and overreach won it international media reaction

Keeping track of your weight is easy today, but it hasn't always been.

In 1698, the "Widow Bourne" (her first name is unknown) opened an upscale grocery and coffee house at 3 St. James's Street in London. In 1755, the then owners began a customer loyalty programme that capitalised on Londoners' inability to easily weigh themselves. A century later, data mined from this programme was subjected to a statistical analysis that was published in *Nature*.

chrisdorney/iStockPhoto.com

The quarry: Two centuries of data

Coffee shop customers were offered an opportunity to be weighed on a grocer's balance-beam weighing scale, with a resolution of a quarter of a pound. Their weights were carefully and clearly entered into indexed ledgers that contained the customer's name, the date, and notes as to any extra clothes or shoes worn during weighing. Since entries were meant for personal use, the person's age or date of birth was not recorded.

The business later expanded to wine and spirits when Berry Bros. (later Berry Bros. & Rudd) took it over in the 1840s and a separate ledger was established for women: it contains over 300 entries. As is seen in the lower half of Figure 1, repeat weighings of women became more common from 1885 onwards. The actual measurements for the two women with the longest traces ("Lady" A and "Mrs" B) are shown in the upper half of the figure.

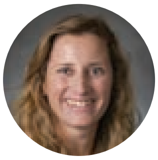
The 1880s also saw an important expansion and democratisation in the self-tracking of body weight, with the rapid spread of publicly placed weighing-machines ("penny scales") that, when a person who had stepped on them inserted a coin, displayed – and even printed – the person's weight. Domestic bathroom scales began to be available in the early 1900s, and made self-monitoring more private.

1884: Quarrying and data-analyses

Francis Galton was a pioneer in the development of ways to measure statistical associations and relationships. (But he also endorsed "eugenics", a term he coined. Its widespread practice after the Victorian era has since been widely condemned.)

In addition to crowdsourcing information and making direct measurements in his Anthropometric Laboratory, Galton also

James Hanley is professor of biostatistics in the Faculty of Medicine at McGill University.



Elizabeth Turner is an associate professor of biostatistics and global health at Duke University.

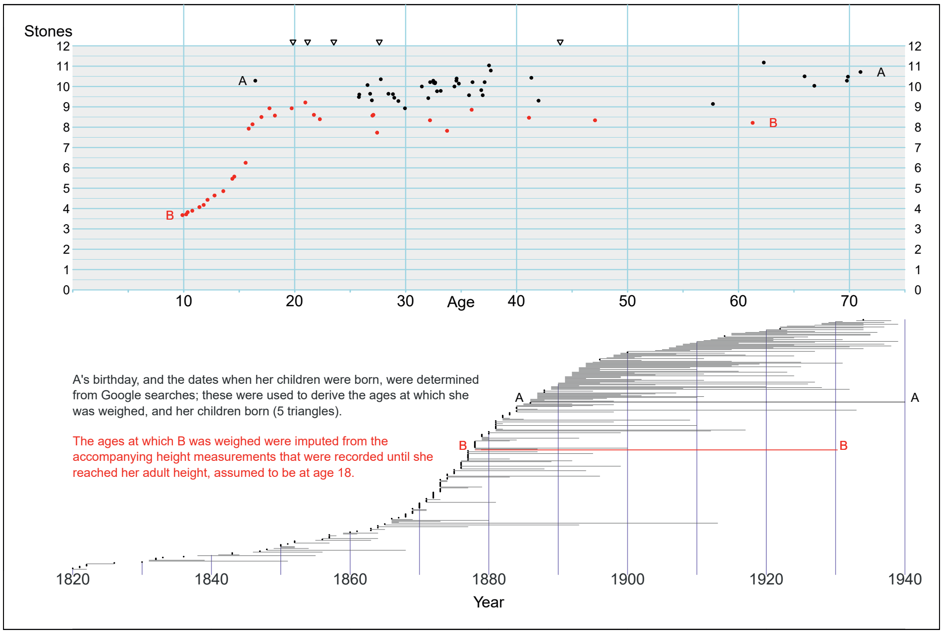


Figure 1: Bottom: horizontal lines (or single dots) showing connecting dates of Berry Bros. & Rudd's first and last weighings of upwards of 300 women in their ledgers. Top: the two longest age–weight traces from these. The authors thank Simon Berry for graciously permitting us to consult and photograph the material in these ledgers.

The dissolute life led by the upper classes has left its mark on their age-weight traces

mined non-traditional sources of data. He was keen "to know in an exact way the amount of change that may have occurred in our race during recent generations".[1] For an answer, he looked to the data "quarry" (his word) located at what was then Berry Bros., where "some of the customers who had town houses had been weighed year after year during the Parliamentary season for the whole period of their adult lives".

After convincing himself that the measurements were "genuine and accurate", he selected the nobility for study. He selected this (in Karl Pearson's words,[2] "well-rounded") class because the dates of their births could be easily learned (presumably from a publication such as Debrett's or Burke's Peerage), which had to be done in order to collect the years in which they were weighted with their ages at the time, and also because they "formed a more homogeneous group than one that included younger brothers and men about town, who marry late and lead less regular lives".

Given today's concerns with data privacy and reproducibility, his data-abstraction procedure is of note:

> I therefore begged Messrs. Berry to find a clerk for me who should make the required abstracts under their direction in an anonymous form for statistical purposes. I also asked to be furnished with an alphabetical list of the persons weighed, that I might know generally with whom I was dealing, and that each schedule should bear a reference to the folio whence it was extracted, so that, whenever verification was needed, the original might be referred to. All this was done, and I am in possession of 139 schedules referring to as many different persons, namely, 109 peers, 29 baronets (who were added as makeweights), and 1 eldest son of a peer …. Each schedule gives the age and year of the several weighings, the highest and lowest weights recorded in that year, and a copy of such remarks as were entered at the time about the dress. An age-weight trace … was plotted on a large scale on each schedule.

He divided the individual age traces into three year-of-birth divisions or "generations", each covering a period of 30 years, as shown in Figure 2.

Galton first addressed the *within-person* fluctuations. Using his own numerical measure (which we will return to below), he found that larger fluctuations in weight were more common in the earlier generations (annual fluctuation of about 7 lbs, or about 3.2 kg) than in the latest group (5 lbs, or about 2.3 kg), reflecting "an irregularity in the mode of life that was greater two or three generations back than now" and leaving no doubt that "the dissolute life led by the upper classes about the beginning of the nineteenth century has left its mark on their age-weight traces".

As for the *inter-generation* differences in the mean age–weight curves, his three smoothed curves (see Figure 2) "show with great distinctness that the noblemen of the generation which flourished about the beginning of this [nineteenth] century attained their meridian and declined much earlier than those of the generation 60 years their juniors". (In his biography of Galton, Pearson[2] interpreted the curves as showing that the "primes" for weight were earlier in age for the older generations, being hardly discoverable at all in those born in the first third of the 19th century".)

Reacting to the report in *Nature*, a *New York Times* editorial dated 13 March 1884 remarked that the patterns were "certainly somewhat strange" and speculated on possible reasons for "the physical deterioration of the younger British aristocracy". It even predicted that the House of Lords might "abolish itself as its members grew thin and fragile as the cultured Bostonian swell because of Mr Gladstone's dastardly plan to bring French wines into England and thus to render thin and impoverished the blood of British noblemen".

Today, in retrospect, it may well be that it was Galton's and Pearson's concern with the physical and moral decline of the British race that led them to overinterpret these data. Although he acknowledged that Galton's paper was "rather amusing" and

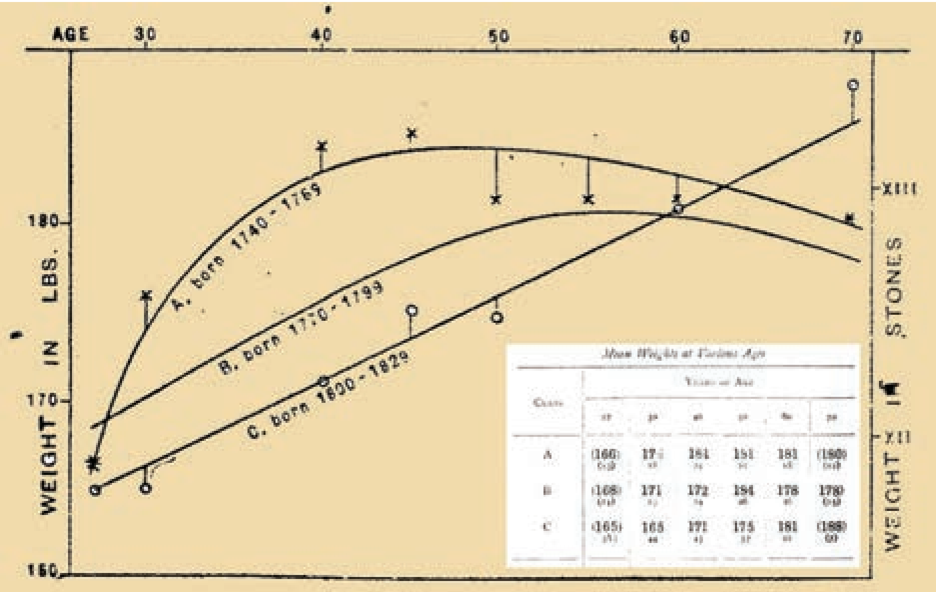


Figure 2: Galton's[1] "translat[ion]" of mean weights at various ages (means for A shown by crosses, for C by small circles and for B omitted for clearness' sake) into "smoothed curves"; the figure was entitled "Mean Age-Weight of British Noblemen in three successive generations". Inset: data table from the same 1884 article: note in particular the small sample sizes at the (high-leverage) age of 70 in Generation C: some may not even have reached 70 by the data-extraction date. "The figures in parentheses are doubly meaned results, those to the left being derived from observations made at the ages of 26 and 28, and those to the right at 68 and 72."

his "data were somewhat scanty", Pearson nevertheless suggested that "his general conclusions appear to be justified". Today, if the data analysis were evaluated for its statistical soundness, Galton's "smoothing" of these scanty and imbalanced longitudinal data would probably receive a failing grade in most modern data-analysis courses.

However, in the spirit of talking about the statistically good as well as the statistically bad, we end with a Galton-developed index that we hope would deserve an A grade were it to be developed in a modern data-analysis project, or used to communicate variability to lay people.

As noted above, Galton began by investigating how irregular the individual trajectories were, so he quantified the amount of irregularity in each of the 139 traces. He then sorted them from "least" irregular to "most" irregular, and his report included plots of the traces representing the lower and upper quartiles (see the top of Figure 3).

By 1884, the root mean square deviation was already being used to quantify variation, and ten years later Karl Pearson renamed it the "standard" deviation. Instead, Galton usually ranked the observations into an "array" (a favourite phrase of his) and just took half the interquartile range as a "probable error". But working with deviations from sloping lines or curves was cumbersome. Instead, for each trace, he devised an ingenious and intuitive measure that involved very little hand-calculation:

> I calculated a rough measure of the irregularity of each trace on the same principle that one might adopt to measure the discursiveness of a rambling path, in comparison with that of a straight turnpike road between the same points, namely by finding the proportion that the length of one bore to the other. I measured the trace and also the general sweep of the trace with a mapmaker's "perambulator," and divided one by the other. I corrected each result on the principle that a fluctuation of 12 lbs. in a man of 16 stone should not count more than one of 9 lbs. in a man of 12 stone.

Just as in 1884, data mining continues to be used not just in the service of people, but to promote ideological causes

If you find his description difficult to follow, you might instead imagine using Google Maps to gauge how much longer a local road that weaves back and forth beside a motorway is than the motorway itself. You might also replace the word "proportion" with the word "ratio".

Figure 3 shows our rendition of the "traces" of two of the longest – and one of the shortest but most high-resolution – series we found at Berry Bros. & Rudd. Before looking at the numerical measures of the irregularity of these series, you may wish to judge "by eye" which of the two adult-lifetime series seems more irregular. The rankings provided by our test viewers agree with those based on Galton's numerical measure of irregularity: the trajectory in black is 2% longer than the direct (smooth) curve, while the trajectory in red is 5% longer than the corresponding direct curve. His index might be more easily explained to laypersons than the standard deviation or the coefficient of variation – indices often used to quantify volatility in (say) returns on financial investments, or to measure the degree of weaving while driving after consuming cannabis.

Galton had hoped the publication of his paper would "suggest methods of treating the data" in Berry Bros.'s quarry "that have not occurred to [him]self". He could hardly have foreseen just how much recreational products, turnpike travel, computing facilities, statistical tools and practice – and the size of data – would change over the next centuries. Nor could the *New York Times* commentators have imagined just how much the technological, electronic and communication revolutions would change societies, the weight of nations, or how citizens self-track their own weights and their other activities.

Just as in 1884, data mining continues to be used not just in the service of people, but to promote ideological causes. So we need to be wary. We need to be particularly sceptical of "bigger" data, where it is possible to be even more precisely wrong.

But we should also encourage and welcome the creativity that leads to intuitive

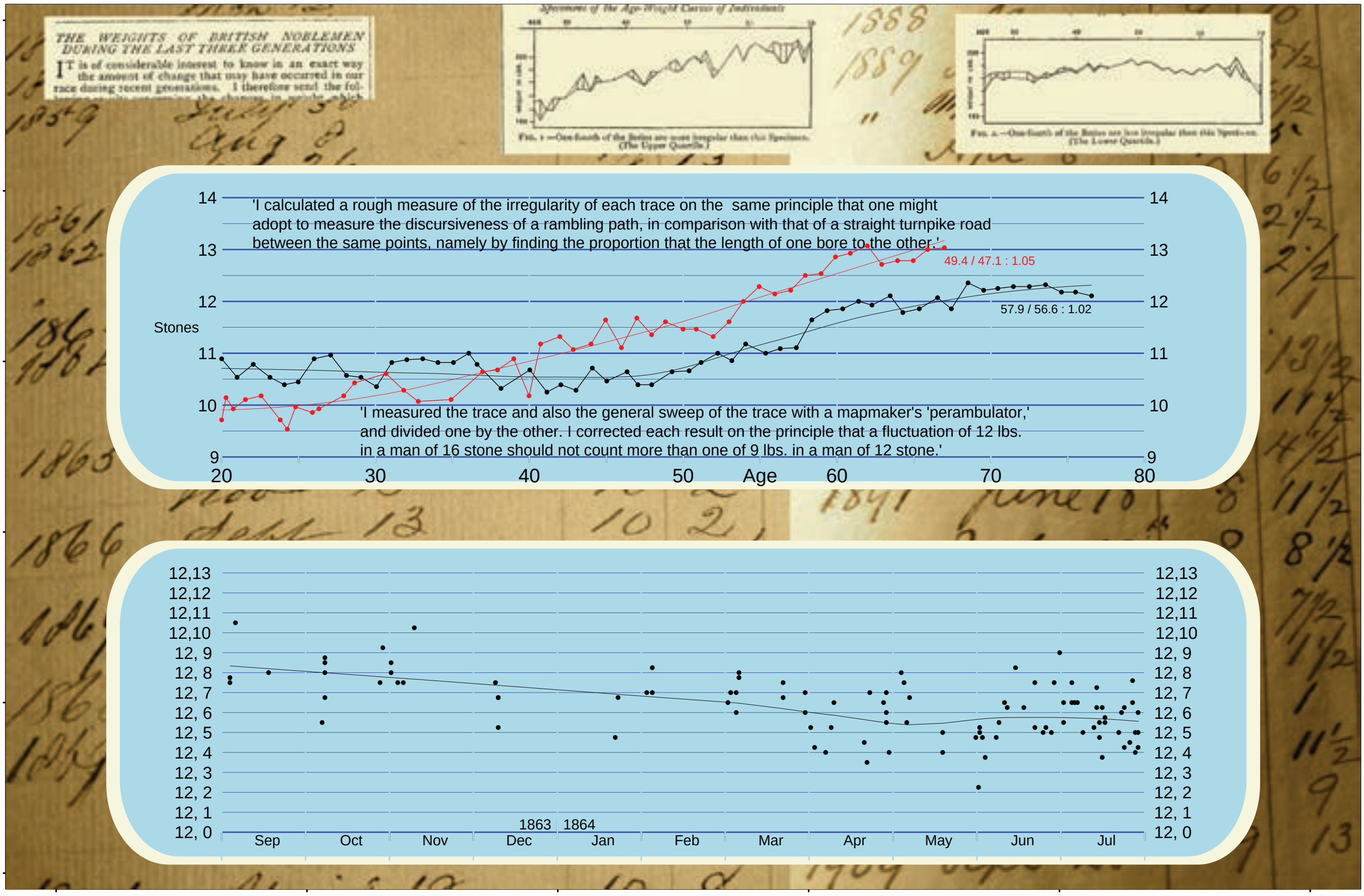


Figure 3: A modern visualisation (mock-up of a smartphone app) of the "traces" of three noblemen whose weight records we extracted from Galton's "data-quarry" at Berry Bros. & Rudd, along with the description and application of his measure of irregularity. The two series in the upper panel, running from 1819 to 1875 (in black) and from 1852 to 1899 (in red), are among the longest we found. The series in the lower panel is based on just over 100 weighings of one nobleman over 11 months, many of them several hours apart on the same day. The smooth lines, fitted by "loess", show the "general sweep" of each trace. The ratio of the pair of numbers shown at the right end of each trace measures how meandering the trace is: the meandering road in red is 49.4/47.1 or 1.05 times longer than the red turnpike. The meandering road in black is only 57.9/56.6 or 1.02 times longer than the black turnpike. The partial records of two Berry Bros. & Rudd clients, and the traces for the noblemen in Galton's data set who were at the first and third quantiles in terms of "irregularity," are used as background.

statistical measures. The "winding road versus turnpike ratio" may not endear itself to mathematical statisticians, or become the standard for measuring deviations from the straight and narrow, but it is a lot easier to explain than a "standard" deviation.

Epilogue

In 2009, Simon Berry, then CEO of Berry Bros. & Rudd, told us: "We still weigh our customers, but only on very rare occasions. The practice slowed down drastically after the Second World War, probably as more and more people bought bathroom scales at home."

This practice took a little longer to reach less affluent segments of society. In the 1950s, when one of the authors (James Hanley) was growing up in an offshore Irish island, few homes had such scales, although his family did have a handheld spring scale for weighing grain, infants, and so on. When he and his siblings accompanied their parents to the mainland a few times a year, the merchants would (as part of their customer service) measure the children on their beam scales and congratulate the parents on how well they were growing.

We also shared the *New York Times*'s 1884 editorial with Mr Berry. He was intrigued by the newspaper's fears for the British nobility and was happy to note that its prediction of societal collapse had, thus far, proved unfounded. ■

References

1. Galton, F. (1884) The weights of British noblemen during the last three generations. *Nature*, **29**, 266–268.

2. Pearson, K. (1930) *The Life, Letters and Labours of Francis Galton*, Vol. IIIA, Chap. XIV, pp 136–137. Cambridge: Cambridge University Press. galton.org/pearson/

Note

Additional material, including a link to the full text of the *New York Times* editorial of 13 March 1884 mentioned in the text, can be found at jhanley.biostat.mcgill.ca/BBR/