

University League Tables

Brian D. Ripley

Professor of Applied Statistics

Scores and league tables are the spirit of the age: we have seen them for schools and hospitals, and league tables for universities were inevitable. Perhaps equally inevitably they would reveal depths of innumeracy unsuspected by a working scientist, let alone a trained statistician.

The *Financial Times* produced a league table on 29 April 1998, the *Times* on 15 May 1998, the *Daily Telegraph* one on 29 July 1998 and the *Sunday Times* had a league table in its *University Guide* (aimed at prospective students) on 8 November 1998. Last summer *Research Fortnight* produced a table mainly based on research funding.

All of these tables place Cambridge at the top and have been accompanied by critical (if not always accurate) headlines and comment, for example

‘Cracks open in ivory towers.’

‘Cambridge may sit at the apex but the FT’s rankings, report . . . , suggest that Oxbridge supremacy belongs to times past.’

‘Dreaming spires forced to wake up to reality.’

‘Oxford forced to settle for a third.’

‘Cambridge and York top for teaching. The 1960s universities are setting standards that rival the league leaders.’

Data

The data used has been collected from various areas:

Research assessment based on the 1996 RAE ratings (usually with 5* = 6, possibly even on a 7-point scale).

Teaching Two elements, one based on scores from QAA subject assessments. The *Times* thresholded the English QAA assessments at 22/24.

Student:staff ratios

Admissions The A-level points needed for entry.

The number of applicants per place.

Degree class Various measures, including the proportion of firsts awarded, and the proportion of firsts or 2:1s.

Employability Some measure of the destination of graduate, either as the % employed or including where they are employed.

Facilities usually library and computer expenditure, but also social and recreational facilities.

Research income either in total or by source (industry / research council).

A critical appraisal

- *Comparability?* Most of the measures are comparable across universities, and externally assessed. The use of degree results stands out as the exception. Cambridge's graduates may deserve their 28.1% of firsts, but only Cambridge says so.

Some of the units in the assessment are not universities, and although this may be defensible for UCL and the colleges of the University of Wales, it is a lot less clear for UMIST, IC and LSE, and even less for SOAS.

- *Subject biases?* The measures are definitely not independent of subject mix at a university. Science and Engineering subjects attract (and spend) much more research income, have more graduate students, more dispersed sets of A-level scores and degree results, (probably) have more employable graduates and are less over-subscribed at university admissions.

I believe that the RAE scores are quite explicitly designed for the division of funds within a unit of assessment, and are not audited to be comparable across units of assessment. (They certainly seem not to be.) Although there are generic definitions, these apply to a 'group' and it is much easier to achieve the criteria if the groups are large. Inspection of the unit of assessment scores bears this out: 45% of assessments in computer science were four or higher, against 26% in business and administration. This *might* genuinely reflect national research strengths, but would still be a subject bias.

The net effect would appear to be to favour technological institutions, e.g. ICSTM.

- *What is good?* The *FT* used student:staff ratio and the *Times* used staff:student ratio. The *Sunday Times* said they used student/staff ratio but appear to have actually used $\min(10, (1000 \times \text{staff})/\text{students})$. Amazing!
- *Apples or Oranges?* The measures which are used to compute the tables are a mixture of output measures (research and teaching assessments, degree class, employability) and input measures (spending on facilities, research income, admissions). Should a university be rated highly because it is expensive to run, because it is run efficiently, because it makes the best of its students and staff, because it has a strong public image, ...? The current league tables seem to be a mixture of the first and last.
- *Scaling and weighting.* All the tables have effectively scaled each measure to a score out of 100, and then applied a weighting factor, for example 2.5 for teaching quality, 2 for research quality, and so on. This is a great deal less objective than it is made to appear.
 - It depends on what is measured. *The Times* and *Sunday Times* measured % not unemployed, the *FT* measured % unemployed amongst graduates. The % 'not unemployed' varies from 85 to 97%. If this were treated as the % unemployed, it would range from 3 to 15% and be treated as $100/15 \approx 7$ times as important!
 - It depends on what institutions are included. Taking the employability issue again, the *FT* methodology would give this even more weight if the former polytechnics were excluded. For research income and facilities excluding Oxbridge would change the ranking amongst the remaining institutions.
 - There is a ceiling effect. If the maximum RAE score is 6 and the maximum QAA score is 24 (itself the sum of six elements out of 4), the differences are compressed at the high end. Similarly with facilities: if one institution has amazingly good (costly?) facilities, the effect is merely to down-weight that factor for everyone else.

- Many measures, including all the financial ones, depend on the size of the institution. The *FT* scaled by the (full-time equivalent) number of students. Why? The students do not generate (let alone spend) the research income, and in top universities they are minor users of research facilities. It might be better to scale by the number of staff. Academic staff? Academic and research staff? Employed by whom? The size of Oxford is many-dimensional.
- *Volume effects.* The research and teaching assessments are unweighted averages. More precisely, they are weighted averages with weights one or zero (if a subject is not assessed). Thus low ratings in small subjects are as damaging on these measures as those in large subjects. This both makes small subjects nationally disproportionately important, and gives undue weight to minor subjects at the institution (education at Oxford?)

Management information?

Are the league tables completely useless? For students choosing a university, they are probably only useful for honing numeracy skills. Figure 1 illustrates how poor the journalists' perception of numbers can be. My colleagues' perception that the differences between old and new universities were being blurred are not reflected in the overall scores (a point the journalists seem to have been blinded to).

I found an article on the Web¹ by Kingston University's VC to be thoughtful. He says:

One way to look at league tables is as free performance indicators and free market research. Private companies have to pay large sums of money for comparative information about how they are doing compared with their competitors and how they are perceived in the marketplace. Universities get this information provided by newspapers.

Whether the free information is worth its price is worthy of consideration, as is whether it is in fact provided by newspapers (rather than HESA, QAA, ...).

For the universities the information can be viewed as useful multi-dimensional information on relative performance. The information should already be known, but it does seem not to be widely appreciated in Oxford. Figures 2, 3 and 4 show some views of the *Times* and *Financial Times* data, scaled in the soundest ways we could find. These may help to show where Oxford scores well and where it is falling behind.

We are probably all aware of the RAE ratings (although these particular summaries have many flaws). Fewer of us are likely to be aware of the pattern shown in Figure 3, which has Cambridge and York substantially ahead² on teaching assessments. Note the relative positions of Warwick in two 1998 summaries of teaching assessment.

The league tables conflate input and output measures, and output measures seem more important.

Output measures

For teaching and research output measures we can only credibly use a weighted form of the RAE and QAA assessments. That Oxford is comfortably beaten by Cambridge and York (neither of which uses the treasured tutorial system) must be of great concern. Either Oxford is failing to

¹http://www.kingston.ac.uk/~webmast/nf_004466.html

²remember the ceiling effect; not even Cambridge can exceed 100%.

make clear to external assessors the great advantages of its expensive system of undergraduate teaching or it is deluding itself.

Note though that both of these measures are of output in terms of perceptions, not of fact. In particular, for teaching quality not attempt is made to measure value-added, how much the students have learnt or even how much more they have learnt than they would have at a middle-ranking university.

Other output measures would be possible. For example, a more credible measure of excellence of library facilities than spending would be an output measure such as 'what proportion of book/journal requests could be met at an institution, and in how many minutes (hours?)', or a measure of customer satisfaction. In the commercial world service facilities are not rated as excellent because they are very costly!

Input measures

Looking at Figure 4, where Oxford shows up well, we see that there is something very special about Heriot-Watt and Robert Gordon (petroleum engineering, I believe). Oxford's strengths are input measures, its spending on facilities and its research income per FTE.

For better input measures we have been investigating data on research funding produced by HEMS³ for HESA⁴ for the 1996/7 financial year. This does attempt to adjust for the subject spread of an institution, but in a way that favours institutions which specialize in unusual subjects (not just petroleum engineering), and has considerable problems with 'size'. As Cambridge came out 50% above any other institution on 'average market share' but below Oxford and UCL on total market share, it is not credible.

For our own purposes we need a rather more limited basis for comparison, one that is not dominated by the teaching quality assessment of Bournemouth University (as in Figure 3).

³Higher Education Management Studies Group

⁴Higher Education Statistics Agency

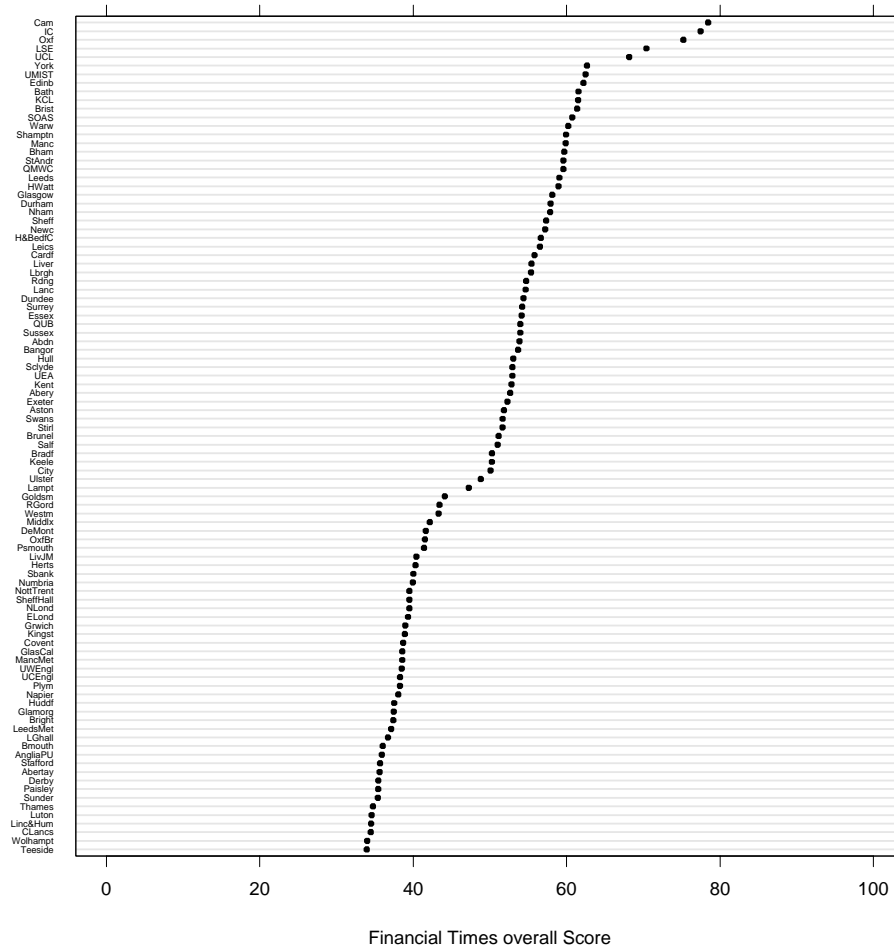


Figure 1: Dot chart of the overall scores in the *Financial Times* analysis, which says ‘Oxford and Cambridge are starting to see their crown threatened by young pretenders ... York, founded 35 years ago, emerges as the top provincial university.’ Notice just how ahead the top five are of York (number 6). Numbers 57 to 97 are the ‘new universities’.

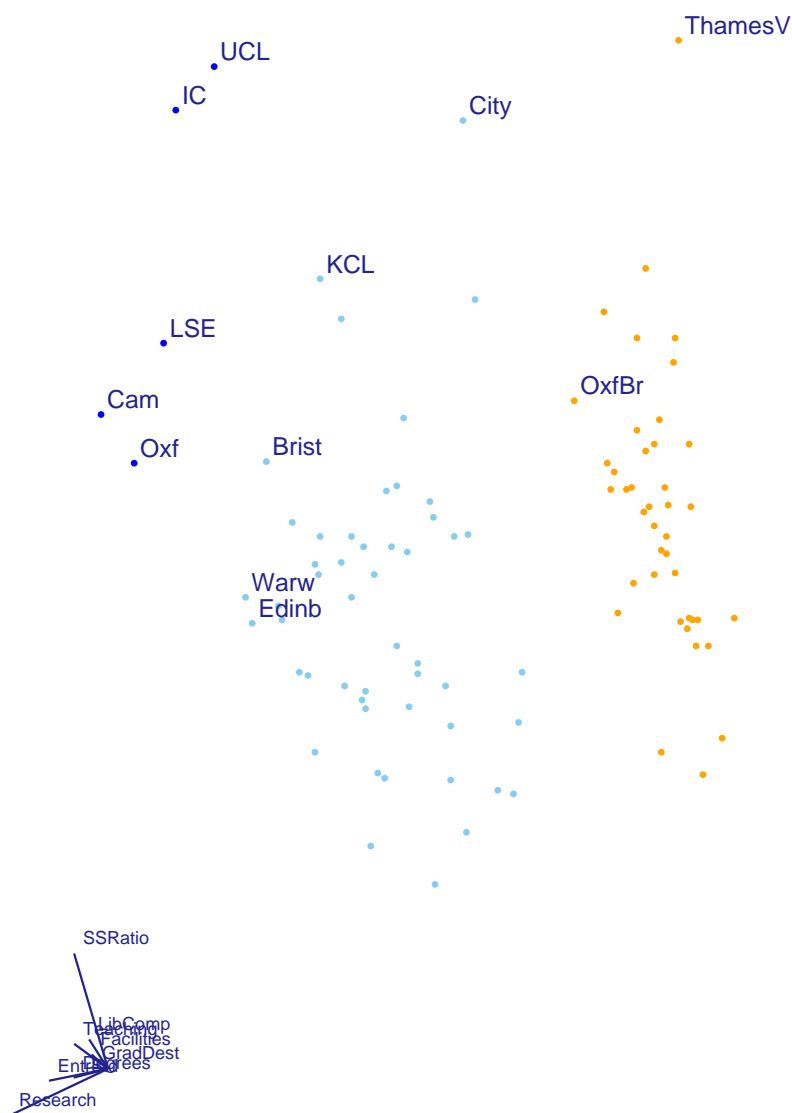


Figure 2: Plot of the first two principal components of the *Times* data, dominated by the research ratings (left–right) and student:staff ratio (top–bottom). Old universities are shown in light blue, new universities in orange, the top five (all old) in dark blue.

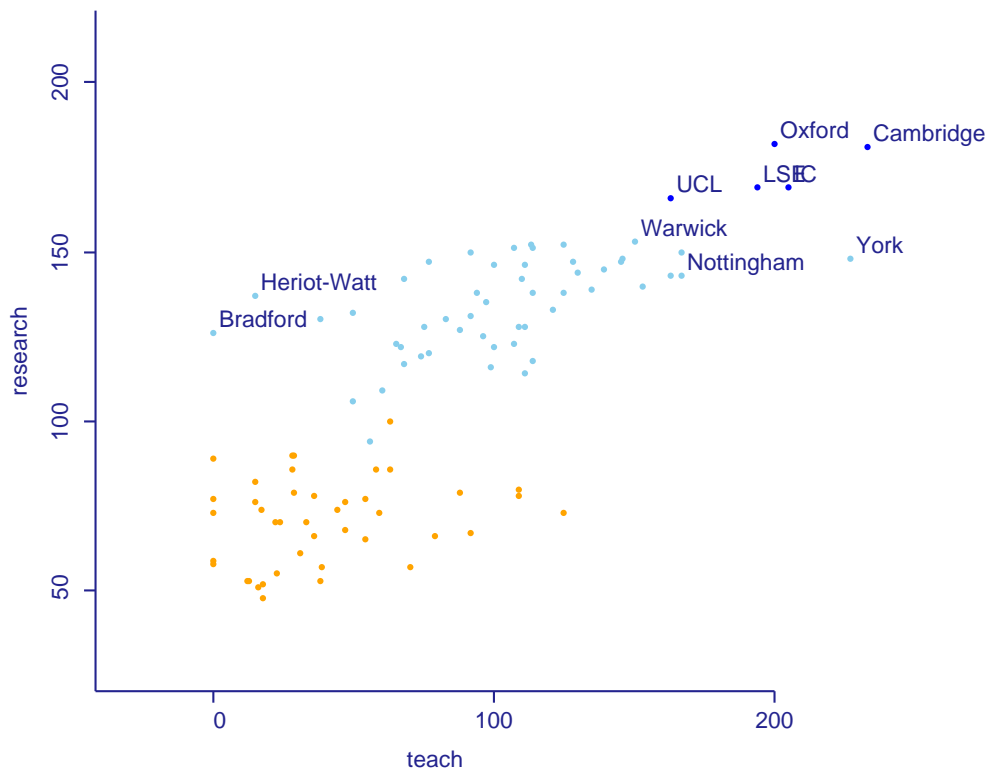
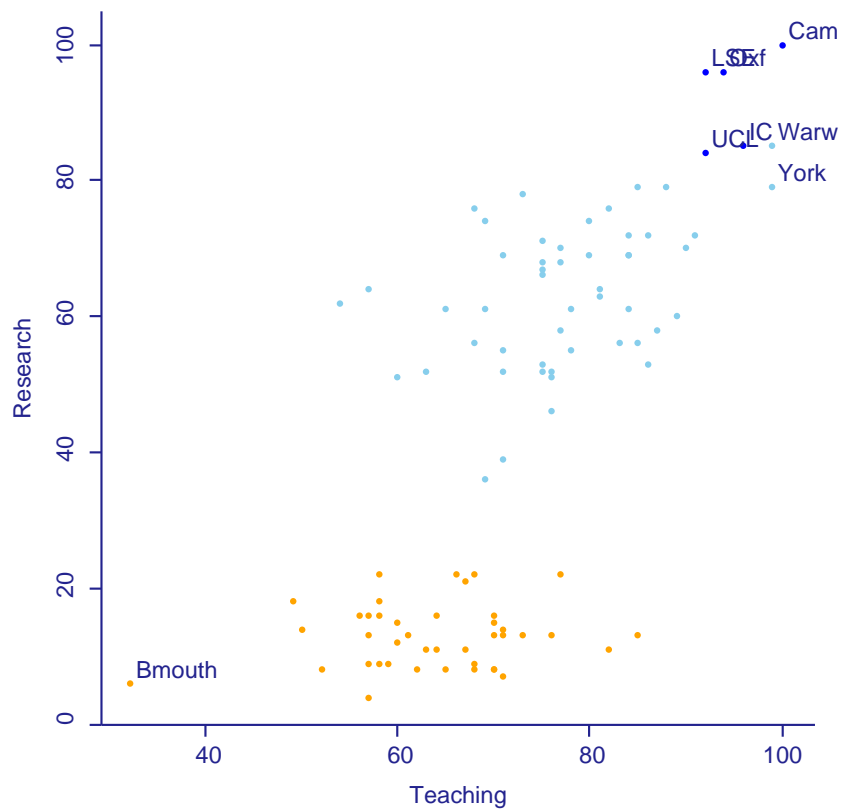


Figure 3: The two most important measures on the *Times* (top) and *Sunday Times* (bottom) scores, research and teaching assessments.

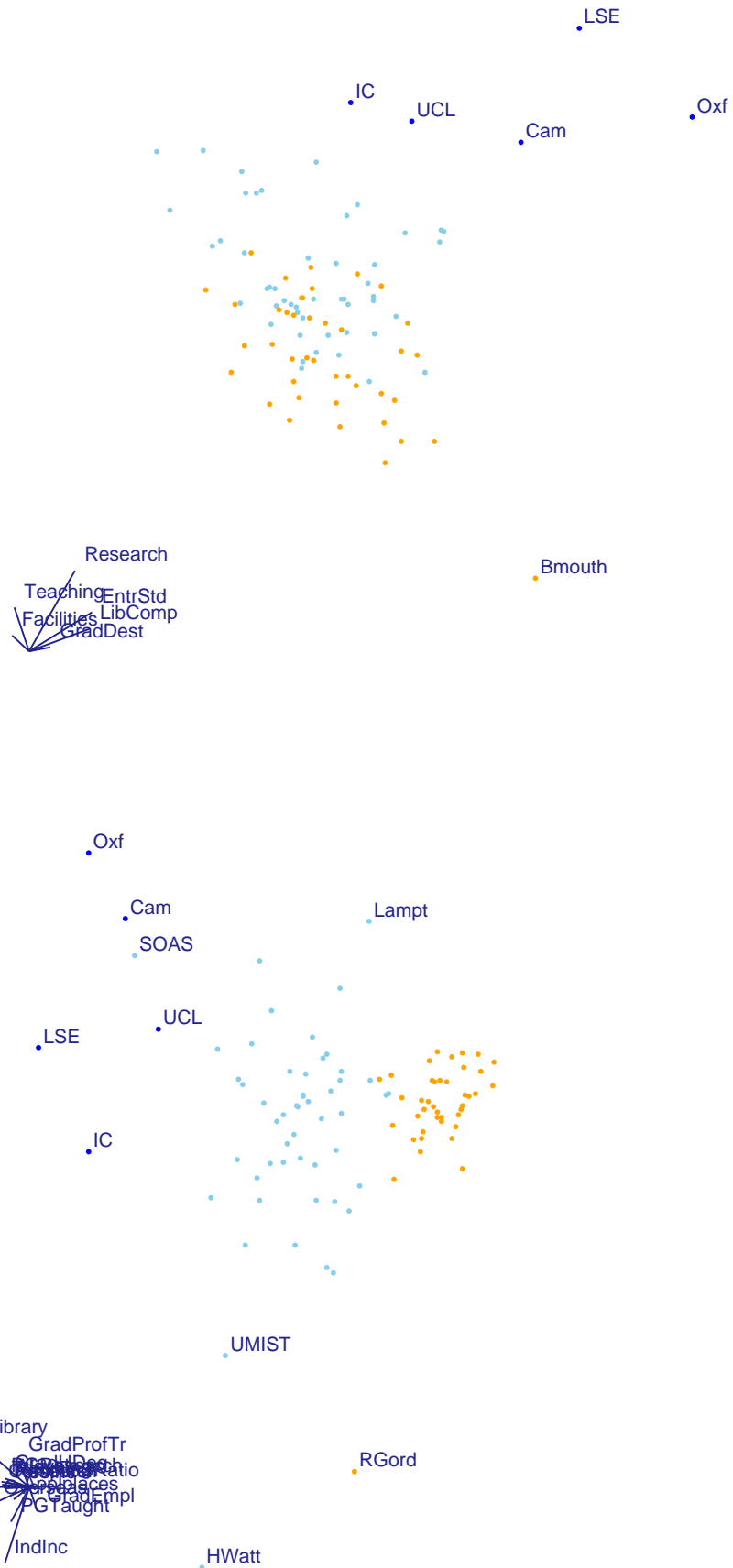


Figure 4: The sort of views of the *Times* (top) and *Financial Times* (bottom) data that Oxford would like to see!