Learning from data science

Mikołaj Barczentewicz explains what data science can teach us about litigation before the Supreme Court

Mikołaj Barczentewicz

Mikołaj Barczentewicz is a law lecturer at the University of Surrey School of Law
barczentewicz.com

> This gender disparity is probably explained by the fact that there are relatively few women among the senior litigators

Despite the woefully inadequate state of open access to case law data in the UK, I decided to use some of what's available (texts of House of Lords and Supreme Court judgments) to show what opportunities open when we're able to use modern data science techniques on case law.

This is how my project at courts.barczentewicz.com was born.

So far, I have revealed the hard data on the worrying gender disparities among lawyers who appear before the Supreme Court. I have also ranked the counsel, the instructing solicitors and the barristers' chambers by numbers of cases in which they appeared in the Supreme Court.

However interesting, these results give merely a small window into what will be possible with further analysis and with greater open access to case law data. But I have faced challenges in applying data science to case law and in what I have been able to learn so far.

NO FULL DATABASE

The most fundamental limitation for any project like mine is the frankly embarrassing situation that there is no online database of all UK case law, openly accessible for machine analysis. Even the closest thing we have, the charity-run BAILII website (bailii.org), explicitly prohibits making copies of significant parts of its database – which is precisely what any data science project has to do.

Though it is rumoured that the government may be inclined to improve access to case law data in the future, I know of no plans to provide access to data even from recent decades. And without such access, we may never be able to understand our law as well as we could using data science.

Given the problems with access to data about other courts, I chose to focus first on judgments of the Supreme Court (which are all available as pdf files on its website); and second, on the last decades of the Appellate Committee of the House of Lords (available on parliament's website). Unfortunately, extracting valuable information from those judgments cannot be entirely automated by an algorithm.

Usually, for example, the names of counsel follow the case heading with the status of the party (eg appellant). However, sometimes the text embedded in a pdf file contains the name of counsel before the

heading or elsewhere. Other typical problems are typos and misspellings of names. Those problems typically require algorithmic workarounds or manual interventions, but they can be managed.

What is more difficult is coping with missing information. Names of judges and lawyers (and of course litigants) are a good example. The issue is that a person's given name and surname are not a unique identifier. Hence, any analysis based on texts of judgments will treat people with identical names as the same person. Conversely, if one person uses different names in different cases they will be seen as several different people.

There is no other database linking lawyers or judges with cases, against which someone could check a dataset created from texts of judgments. One could try to reconstruct some of that data based on the information given on the websites of, for example, barristers' chambers, but those are often incomplete and out of date. Also, the websites don't follow a unified standard of displaying such data and thus cannot be easily mined for information.

WHAT I LEARNED

Despite all the difficulties, I was able to create and analyse a dataset about litigation before the Supreme Court, which yielded a number of valuable insights. I first ranked all lawyers (barristers, solicitor advocates and Scottish advocates) who represented clients in cases before the Supreme Court by the number of such cases they did.

I then created similar rankings for instructing solicitors and barristers' chambers (and stables of Scottish advocates). I was able to do the latter by linking my dataset with data from the Barristers' Register and with data kindly provided by the Scottish Faculty of Advocates.

Unfortunately, I didn't yet have access to similar data on solicitor advocates, so I wasn't able to give as much information about them as about the other groups of counsel.

The ranking of instructing solicitors showed the government's numerical dominance, with the Treasury Solicitor (108 cases); the Government Legal Department (101); HM Revenue & Customs (HMRC) (47); and the Crown Prosecution Service (CPR) (37) taking the first four spots.

The top ten law firms who acted before the Supreme Court are:

- Leigh Day (29 cases)
- Freshfields Bruckhaus Derringer (25)
- Bindmans (22)
- Allen and Overy (21)
- Irwin Mitchell (19)
- Herbert Smith Freehills (17)
- Birnberg Peirce (15)
- Dawson Cornwell (13)
- Hill Dickinson (13), and
- Pinsent Masons (13)

Interestingly, Freshfields and non-governmental organisation Liberty make the overall top five by the number of times they represented interveners or interested parties. This shows Freshfields' strong pro bono practice.

Looking at the instructed counsel, the data showed how significant is the gender disparity at the top of the profession. Only 21 out of the top 128 counsel (ie everyone with six or more Supreme Court appearances) and only two in the top ten are women.

This gender disparity is probably explained by the fact that there are relatively few women among the senior litigators, ie the pool from which the solicitors and their clients choose counsel to appear in the Supreme Court.

After all, according to the Bar Standards Board women constitute only 16.2 per cent of Queen's Counsel.

What may be harder to explain this way is the fact that only 20 per cent of female counsel led teams of counsel; whereas 39 per cent of the male counsel did. I also found that when women lead teams of lawyers, they are significantly more likely than men to have at least one other woman on their team.

FUTURE PLANS

Given that the Supreme Court has been functioning only since late 2009, the picture of top-level litigation we can get from looking at this court is limited in the crucial respect that it does not cover the full span of careers of even one generation of lawyers. Therefore, it can say little about longer trends.

This is why I'm now working on extending the dataset to litigation before the Appellate Committee of the House of Lords since 1973. That date is, unsurprisingly, dictated by the availability of data.

With the House of Lords data I will be able, for example, to check my hypothesis that there is a significant trend that the proportion of women among the most junior counsel instructed in the highest court of appeal is increasing.

I will also look at possible trends in overall gender balance. And I will, of course, rank the top advocates and law firms – by number of cases before the country's top court – for the last nearly fifty years. **SJ**

> ❝ The most fundamental limitation... is the frankly embarrassing situation that there is no online database of all UK case law, openly accessible for machine analysis