**Ch2. What can data do for Investigative Journalism**

Why are we even doing this? If you look at Wikipedia’s answer you will find this “[Data-driven journalism](https://en.wikipedia.org/wiki/Data-driven_journalism), often shortened to "ddj", a term in use since 2009, is a journalistic process based on analyzing and filtering large data sets for the purpose of creating or elevating a news story. Many data-driven stories begin with newly available resources such as open source software, open access publishing and open data, while others are products of public records requests or leaked materials.”

The Wikipedia article has recently been merged with the one on DataJournalism. The entry warns the reader that DDJ is “not to be confused with [database journalism](https://en.wikipedia.org/wiki/Database_journalism)”.

More frequently datajournalism is often confused with journalists “doing statistics” – some critics would say “journalists doing statistics badly”. I think it’s quite simple – we take datasets and use software to help us find facts or patterns which lead us to stories.

For example, we might take a large dataset and filter it to find companies with large government contracts which are owned by politicians, their families or friends. This isn’t going to result in a story full of numbers and statistics, it’s going to be a story with some names, and some eye-watering sums of money!

My favourite definition of news is “news is something which someone, somewhere, doesn’t want you to know. Everything else is advertising”. I think it was Lord Northcliffe who said that, in the late 19th or early 20th century – but I can’t nail it down. (For an interesting investigation of the origins of this thought, pithily expressed by editors and proprietors over the ages, read [this online article](https://quoteinvestigator.com/2013/01/20/news-suppress/)).

For centuries, looking for that kind of story – the something which someone doesn’t want the world to know, investigative journalists have concentrated on looking for scoops – stories which nobody else has. To do that they needed great sources, whistleblowers, and luck. You might think that in a world of instant communication it’s harder to get information which nobody else has. Maybe.

But that doesn’t matter – now you can get “[conceptual scoops](https://www.poynter.org/archive/2004/the-conceptual-scoop/)”. Governments and others are now publishing so much data that the scoops are hiding in plain sight. We just need to be the ones who know how to join the dots.

[For more on this topic you could read – “[Does Open Data Journalism](https://reutersinstitute.politics.ox.ac.uk/our-research/does-open-data-need-journalism)?”, a paper I wrote in 2015. ]