How statistics lost their power – and why we should fear what comes next

The ability of statistics to accurately represent the world is declining. In its wake, a new age of big data controlled by private companies is taking over – and putting democracy in peril

by William Davies

William Davies is a sociologist and political economist. His books include The Limits of Neoliberalism and The Happiness Industry. Twitter: @davies_will

In theory, statistics should help settle arguments. They ought to provide stable reference points that everyone – no matter what their politics – can agree on. Yet in recent years, divergent levels of trust in statistics has become one of the key schisms that have opened up in western liberal democracies. Shortly before the November presidential election, a study in the US discovered that 68% of Trump supporters distrusted the economic data published by the federal government. In the UK, a research project by Cambridge University and YouGov looking at conspiracy theories discovered that 55% of the population believes that the government "is hiding the truth about the number of immigrants living here".

Rather than diffusing controversy and polarisation, it seems as if statistics are actually stoking them. Antipathy to statistics has become one of the hallmarks of the populist right, with statisticians and economists chief among the various "experts" that were ostensibly rejected by voters in 2016. Not only are statistics viewed by many as untrustworthy, there appears to be something almost insulting or arrogant about them. Reducing social and economic issues

to numerical aggregates and averages seems to violate some people's sense of political decency.

Nowhere is this more vividly manifest than with immigration. The thinktank British Future has studied how best to win arguments in favour of immigration and multiculturalism. One of its main findings is that people often respond warmly to qualitative evidence, such as the stories of individual migrants and photographs of diverse communities. But statistics – especially regarding alleged benefits of migration to Britain's economy – elicit quite the opposite reaction. People assume that the numbers are manipulated and dislike the elitism of resorting to quantitative evidence. Presented with official estimates of how many immigrants are in the country illegally, a common response is to scoff. Far from increasing support for immigration, British Future found, pointing to its positive effect on GDP can actually make people more hostile to it. GDP itself has come to seem like a Trojan horse for an elitist liberal agenda. Sensing this, politicians have now largely abandoned discussing immigration in economic terms.

All of this presents a serious challenge for liberal democracy. Put bluntly, the British government – its officials, experts, advisers and many of its politicians – does believe that immigration is on balance good for the economy. The British government did believe that Brexit was the wrong choice. The problem is that the government is now engaged in self-censorship, for fear of provoking people further.

This is an unwelcome dilemma. Either the state continues to make claims that it believes to be valid and is accused by sceptics of propaganda, or else, politicians and officials are confined to saying what feels plausible and intuitively true, but may ultimately be inaccurate. Either way, politics becomes mired in accusations of lies and cover-ups.

The declining authority of statistics is at the heart of the crisis that has become known as "post-truth" politics

The declining authority of statistics – and the experts who analyse them – is at the heart of the crisis that has become known as "post-truth" politics. And in this uncertain new world, attitudes towards quantitative expertise have become increasingly divided. From one perspective, grounding politics in statistics is elitist, undemocratic and oblivious to people's emotional investments in their community and nation. It is just one more way that privileged people in London, Washington DC or Brussels seek to impose their worldview on everybody else. From the opposite perspective, statistics are quite the opposite of elitist. They enable journalists, citizens and politicians to discuss society as

a whole, not on the basis of anecdote, sentiment or prejudice, but in ways that can be validated. The alternative to quantitative expertise is less likely to be democracy than an unleashing of tabloid editors and demagogues to provide their own "truth" of what is going on across society.

Is there a way out of this polarisation? Must we simply choose between a politics of facts and one of emotions, or is there another way of looking at this situation? One way is to view statistics through the lens of their history. We need to try and see them for what they are: neither unquestionable truths nor elite conspiracies, but rather as tools designed to simplify the job of government, for better or worse. Viewed historically, we can see what a crucial role statistics have played in our understanding of nation states and their progress. This raises the alarming question of how – if at all – we will continue to have common ideas of society and collective progress, should statistics fall by the wayside.

In the second half of the 17th century, in the aftermath of prolonged and bloody conflicts, European rulers adopted an entirely new perspective on the task of government, focused upon demographic trends – an approach made possible by the birth of modern statistics. Since ancient times, censuses had been used to track population size, but these were costly and laborious to carry out and focused on citizens who were considered politically important (property-owning men), rather than society as a whole. Statistics offered something quite different, transforming the nature of politics in the process.

Statistics were designed to give an understanding of a population in its entirety, rather than simply to pinpoint strategically valuable sources of power and wealth. In the early days, this didn't always involve producing numbers. In Germany, for example (from where we get the term Statistik) the challenge was to map disparate customs, institutions and laws across an empire of hundreds of micro-states. What characterised this knowledge as statistical was its holistic nature: it aimed to produce a picture of the nation as a whole. Statistics would do for populations what cartography did for territory.

Equally significant was the inspiration of the natural sciences. Thanks to standardised measures and mathematical techniques, statistical knowledge could be presented as objective, in much the same way as astronomy. Pioneering English demographers such as William Petty and John Graunt adapted mathematical techniques to estimate population changes, for which they were hired by Oliver Cromwell and Charles II.

The emergence in the late 17th century of government advisers claiming scientific authority, rather than political or military acumen, represents the origins of the "expert" culture now so reviled by populists. These path-breaking individuals were neither pure scholars nor government officials, but hovered somewhere between the two. They were enthusiastic amateurs who offered a new way of thinking about populations that privileged aggregates and objective facts. Thanks to their mathematical prowess, they believed they could calculate what would otherwise require a vast census to discover.

Blindness to local cultural variability is precisely what makes statistics vulgar and potentially offensive

There was initially only one client for this type of expertise, and the clue is in the word "statistics". Only centralised nation states had the capacity to collect data across large populations in a standardised fashion and only states had any need for such data in the first place. Over the second half of the 18th century, European states began to collect more statistics of the sort that would appear familiar to us today. Casting an eye over national populations, states became focused upon a range of quantities: births, deaths, baptisms, marriages, harvests, imports, exports, price fluctuations. Things that would previously have been registered locally and variously at parish level became aggregated at a national level.

New techniques were developed to represent these indicators, which exploited both the vertical and horizontal dimensions of the page, laying out data in matrices and tables, just as merchants had done with the development of standardised book-keeping techniques in the late 15th century. Organising numbers into rows and columns offered a powerful new way of displaying the attributes of a given society. Large, complex issues could now be surveyed simply by scanning the data laid out geometrically across a single page.

These innovations carried extraordinary potential for governments. By simplifying diverse populations down to specific indicators, and displaying them in suitable tables, governments could circumvent the need to acquire broader detailed local and historical insight. Of course, viewed from a different perspective, this blindness to local cultural variability is precisely what makes statistics vulgar and potentially offensive. Regardless of whether a given nation had any common cultural identity, statisticians would assume some standard uniformity or, some might argue, impose that uniformity upon it.

Not every aspect of a given population can be captured by statistics. There is always an implicit choice in what is included and what is excluded, and this choice can become a political issue in its own right. The fact that GDP only captures the value of paid work, thereby excluding the work traditionally done by women in the domestic sphere, has made it a target of feminist critique since the 1960s. In France, it has been illegal to collect census data on ethnicity since 1978, on the basis that such data could be used for racist political purposes. (This has the side-effect of making systemic racism in the labour market much harder to quantify.)

Despite these criticisms, the aspiration to depict a society in its entirety, and to do so in an objective fashion, has meant that various progressive ideals have been attached to statistics. The image of statistics as a dispassionate science of society is only one part of the story. The other part is about how powerful political ideals became invested in these techniques: ideals of "evidence-based

policy", rationality, progress and nationhood grounded in facts, rather than in romanticised stories.

 Since the high-point of the Enlightenment in the late 18th century, liberals and republicans have invested great hope that national measurement frameworks could produce a more rational politics, organised around demonstrable improvements in social and economic life. The great theorist of nationalism, Benedict Anderson, famously described nations as "imagined communities", but statistics offer the promise of anchoring this imagination in something tangible. Equally, they promise to reveal what historical path the nation is on: what kind of progress is occurring? How rapidly? For Enlightenment liberals, who saw nations as moving in a single historical direction, this question was crucial.

 The potential of statistics to reveal the state of the nation was seized in post-revolutionary France. The Jacobin state set about imposing a whole new framework of national measurement and national data collection. The world's first official bureau of statistics was opened in Paris in 1800. Uniformity of data collection, overseen by a centralised cadre of highly educated experts, was an integral part of the ideal of a centrally governed republic, which sought to establish a unified, egalitarian society.

 From the Enlightenment onwards, statistics played an increasingly important role in the public sphere, informing debate in the media, providing social movements with evidence they could use. Over time, the production and analysis of such data became less dominated by the state. Academic social scientists began to analyse data for their own purposes, often entirely unconnected to government policy goals. By the late 19th century, reformers such as Charles Booth in London and WEB Du Bois in Philadelphia were conducting their own surveys to understand urban poverty.

 To recognise how statistics have been entangled in notions of national progress, consider the case of GDP. GDP is an estimate of the sum total of a nation's consumer spending, government spending, investments and trade balance (exports minus imports), which is represented in a single number. This is fiendishly difficult to get right, and efforts to calculate this figure began, like so many mathematical techniques, as a matter of marginal, somewhat nerdish interest during the 1930s. It was only elevated to a matter of national political urgency by the second world war, when governments needed to know whether the national population was producing enough to keep up the war effort. In the decades that followed, this single indicator, though never without its critics, took on a hallowed political status, as the ultimate barometer of a government's competence. Whether GDP is rising or falling is now virtually a proxy for whether society is moving forwards or backwards.

 Or take the example of opinion polling, an early instance of statistical innovation occurring in the private sector. During the 1920s, statisticians developed methods for identifying a representative sample of survey respondents, so as to glean the attitudes of the public as a whole. This breakthrough, which was first seized upon by market researchers, soon led to

the birth of the opinion polling. This new industry immediately became the object of public and political fascination, as the media reported on what this new science told us about what "women" or "Americans" or "manual labourers" thought about the world.

Nowadays, the flaws of polling are endlessly picked apart. But this is partly due to the tremendous hopes that have been invested in polling since its origins. It is only to the extent that we believe in mass democracy that we are so fascinated or concerned by what the public thinks. But for the most part it is thanks to statistics, and not to democratic institutions as such, that we can know what the public thinks about specific issues. We underestimate how much of our sense of "the public interest" is rooted in expert calculation, as opposed to democratic institutions.

As indicators of health, prosperity, equality, opinion and quality of life have come to tell us who we are collectively and whether things are getting better or worse, politicians have leaned heavily on statistics to buttress their authority. Often, they lean too heavily, stretching evidence too far, interpreting data too loosely, to serve their cause. But that is an inevitable hazard of the prevalence of numbers in public life, and need not necessarily trigger the type of wholehearted rejections of expertise that we have witnessed recently.

In many ways, the contemporary populist attack on "experts" is born out of the same resentment as the attack on elected representatives. In talking of society as a whole, in seeking to govern the economy as a whole, both politicians and technocrats are believed to have "lost touch" with how it feels to be a single citizen in particular. Both statisticians and politicians have fallen into the trap of "seeing like a state", to use a phrase from the anarchist political thinker James C Scott. Speaking scientifically about the nation – for instance in terms of macroeconomics – is an insult to those who would prefer to rely on memory and narrative for their sense of nationhood, and are sick of being told that their "imagined community" does not exist.

On the other hand, statistics (together with elected representatives) performed an adequate job of supporting a credible public discourse for decades if not centuries. What changed?

The crisis of statistics is not quite as sudden as it might seem. For roughly 450 years, the great achievement of statisticians has been to reduce the complexity and fluidity of national populations into manageable, comprehensible facts and figures. Yet in recent decades, the world has changed dramatically, thanks to the cultural politics that emerged in the 1960s and the reshaping of the global economy that began soon after. It is not clear that the statisticians have always kept pace with these changes. Traditional forms of statistical classification and definition are coming under strain from more fluid identities, attitudes and economic pathways. Efforts to represent demographic, social and economic changes in terms of simple, well-recognised indicators are losing legitimacy.

Consider the changing political and economic geography of nation states over the past 40 years. The statistics that dominate political debate are largely

national in character: poverty levels, unemployment, GDP, net migration. But the geography of capitalism has been pulling in somewhat different directions. Plainly globalisation has not rendered geography irrelevant. In many cases it has made the location of economic activity far more important, exacerbating the inequality between successful locations (such as London or San Francisco) and less successful locations (such as north-east England or the US rust belt). The key geographic units involved are no longer nation states. Rather, it is cities, regions or individual urban neighbourhoods that are rising and falling.

 The Enlightenment ideal of the nation as a single community, bound together by a common measurement framework, is harder and harder to sustain. If you live in one of the towns in the Welsh valleys that was once dependent on steel manufacturing or mining for jobs, politicians talking of how "the economy" is "doing well" are likely to breed additional resentment. From that standpoint, the term "GDP" fails to capture anything meaningful or credible.

 When macroeconomics is used to make a political argument, this implies that the losses in one part of the country are offset by gains somewhere else. Headline-grabbing national indicators, such as GDP and inflation, conceal all sorts of localised gains and losses that are less commonly discussed by national politicians. Immigration may be good for the economy overall, but this does not mean that there are no local costs at all. So when politicians use national indicators to make their case, they implicitly assume some spirit of patriotic mutual sacrifice on the part of voters: you might be the loser on this occasion, but next time you might be the beneficiary. But what if the tables are never turned? What if the same city or region wins over and over again, while others always lose? On what principle of give and take is that justified?

 In Europe, the currency union has exacerbated this problem. The indicators that matter to the European Central Bank (ECB), for example, are those representing half a billion people. The ECB is concerned with the inflation or unemployment rate across the eurozone as if it were a single homogeneous territory, at the same time as the economic fate of European citizens is splintering in different directions, depending on which region, city or neighbourhood they happen to live in. Official knowledge becomes ever more abstracted from lived experience, until that knowledge simply ceases to be relevant or credible.

 The privileging of the nation as the natural scale of analysis is one of the inbuilt biases of statistics that years of economic change has eaten away at. Another inbuilt bias that is coming under increasing strain is classification. Part of the job of statisticians is to classify people by putting them into a range of boxes that the statistician has created: employed or unemployed, married or unmarried, pro-Europe or anti-Europe. So long as people can be placed into categories in this way, it becomes possible to discern how far a given classification extends across the population.

 This can involve somewhat reductive choices. To count as unemployed, for example, a person has to report to a survey that they are involuntarily out of work, even if it may be more complicated than that in reality. Many people

move in and out of work all the time, for reasons that might have as much to do with health and family needs as labour market conditions. But thanks to this simplification, it becomes possible to identify the rate of unemployment across the population as a whole.

Statistical data is only credible if people will accept the limited range of demographic categories that are on offer

Here's a problem, though. What if many of the defining questions of our age are not answerable in terms of the extent of people encompassed, but the intensity with which people are affected? Unemployment is one example. The fact that Britain got through the Great Recession of 2008-13 without unemployment rising substantially is generally viewed as a positive achievement. But the focus on "unemployment" masked the rise of underemployment, that is, people not getting a sufficient amount of work or being employed at a level below that which they are qualified for. This currently accounts for around 6% of the "employed" labour force. Then there is the rise of the self-employed workforce, where the divide between "employed" and "involuntarily unemployed" makes little sense.

This is not a criticism of bodies such as the Office for National Statistics (ONS), which does now produce data on underemployment. But so long as politicians continue to deflect criticism by pointing to the unemployment rate, the experiences of those struggling to get enough work or to live on their wages go unrepresented in public debate. It wouldn't be all that surprising if these same people became suspicious of policy experts and the use of statistics in political debate, given the mismatch between what politicians say about the labour market and the lived reality.

The rise of identity politics since the 1960s has put additional strain on such systems of classification. Statistical data is only credible if people will accept the limited range of demographic categories that are on offer, which are selected by the expert not the respondent. But where identity becomes a political issue, people demand to define themselves on their own terms, where gender, sexuality, race or class is concerned.

Opinion polling may be suffering for similar reasons. Polls have traditionally captured people's attitudes and preferences, on the reasonable assumption that people will behave accordingly. But in an age of declining political

participation, it is not enough simply to know which box someone would prefer to put an "X" in. One also needs to know whether they feel strongly enough about doing so to bother. And when it comes to capturing such fluctuations in emotional intensity, polling is a clumsy tool.

Statistics have faced criticism regularly over their long history. The challenges that identity politics and globalisation present to them are not new either. Why then do the events of the past year feel quite so damaging to the ideal of quantitative expertise and its role in political debate?

In recent years, a new way of quantifying and visualising populations has emerged that potentially pushes statistics to the margins, ushering in a different era altogether. Statistics, collected and compiled by technical experts, are giving way to data that accumulates by default, as a consequence of sweeping digitisation. Traditionally, statisticians have known which questions they wanted to ask regarding which population, then set out to answer them. By contrast, data is automatically produced whenever we swipe a loyalty card, comment on Facebook or search for something on Google. As our cities, cars, homes and household objects become digitally connected, the amount of data we leave in our trail will grow even greater. In this new world, data is captured first and research questions come later.

In the long term, the implications of this will probably be as profound as the invention of statistics was in the late 17th century. The rise of "big data" provides far greater opportunities for quantitative analysis than any amount of polling or statistical modelling. But it is not just the quantity of data that is different. It represents an entirely different type of knowledge, accompanied by a new mode of expertise.

First, there is no fixed scale of analysis (such as the nation) nor any settled categories (such as "unemployed"). These vast new data sets can be mined in search of patterns, trends, correlations and emergent moods. It becomes a way of tracking the identities that people bestow upon themselves (such as "#ImwithCorbyn" or "entrepreneur") rather than imposing classifications upon them. This is a form of aggregation suitable to a more fluid political age, in which not everything can be reliably referred back to some Enlightenment ideal of the nation state as guardian of the public interest.

Second, the majority of us are entirely oblivious to what all this data says about us, either individually or collectively. There is no equivalent of an Office for National Statistics for commercially collected big data. We live in an age in which our feelings, identities and affiliations can be tracked and analysed with unprecedented speed and sensitivity – but there is nothing that anchors this new capacity in the public interest or public debate. There are data analysts who work for Google and Facebook, but they are not "experts" of the sort who generate statistics and who are now so widely condemned. The anonymity and secrecy of the new analysts potentially makes them far more politically powerful than any social scientist.

A company such as Facebook has the capacity to carry quantitative social science on hundreds of millions of people, at very low cost. But it has very little incentive to reveal the results. In 2014, when Facebook researchers published results of a study of "emotional contagion" that they had carried out on their users – in which they altered news feeds to see how it affected the content that users then shared in response – there was an outcry that people were being unwittingly experimented on. So, from Facebook's point of view, why go to all the hassle of publishing? Why not just do the study and keep quiet?

What is most politically significant about this shift from a logic of statistics to one of data is how comfortably it sits with the rise of populism. Populist leaders can heap scorn upon traditional experts, such as economists and pollsters, while trusting in a different form of numerical analysis altogether. Such politicians rely on a new, less visible elite, who seek out patterns from vast data banks, but rarely make any public pronouncements, let alone publish any evidence. These data analysts are often physicists or mathematicians, whose skills are not developed for the study of society at all. This, for example, is the worldview propagated by Dominic Cummings, former adviser to Michael Gove and campaign director of Vote Leave. "Physics, mathematics and computer science are domains in which there are real experts, unlike macro-economic forecasting," Cummings has argued.

Figures close to Donald Trump, such as his chief strategist Steve Bannon and the Silicon Valley billionaire Peter Thiel, are closely acquainted with cutting-edge data analytics techniques, via companies such as Cambridge Analytica, on whose board Bannon sits. During the presidential election campaign, Cambridge Analytica drew on various data sources to develop psychological profiles of millions of Americans, which it then used to help Trump target voters with tailored messaging.

This ability to develop and refine psychological insights across large populations is one of the most innovative and controversial features of the new data analysis. As techniques of "sentiment analysis", which detect the mood of large numbers of people by tracking indicators such as word usage on social media, become incorporated into political campaigns, the emotional allure of figures such as Trump will become amenable to scientific scrutiny. In a world where the political feelings of the general public are becoming this traceable, who needs pollsters?

Few social findings arising from this kind of data analytics ever end up in the public domain. This means that it does very little to help anchor political narrative in any shared reality. With the authority of statistics waning, and nothing stepping into the public sphere to replace it, people can live in whatever imagined community they feel most aligned to and willing to believe in. Where statistics can be used to correct faulty claims about the economy or society or population, in an age of data analytics there are few mechanisms to prevent people from giving way to their instinctive reactions or emotional prejudices. On the contrary, companies such as Cambridge Analytica treat those feelings as things to be tracked.

But even if there were an Office for Data Analytics, acting on behalf of the public and government as the ONS does, it is not clear that it would offer the kind of neutral perspective that liberals today are struggling to defend. The new apparatus of number-crunching is well suited to detecting trends, sensing the mood and spotting things as they bubble up. It serves campaign managers and marketers very well. It is less well suited to making the kinds of unambiguous, objective, potentially consensus-forming claims about society that statisticians and economists are paid for.

In this new technical and political climate, it will fall to the new digital elite to identify the facts, projections and truth amid the rushing stream of data that results. Whether indicators such as GDP and unemployment continue to carry political clout remains to be seen, but if they don't, it won't necessarily herald the end of experts, less still the end of truth. The question to be taken more seriously, now that numbers are being constantly generated behind our backs and beyond our knowledge, is where the crisis of statistics leaves representative democracy.

On the one hand, it is worth recognising the capacity of long-standing political institutions to fight back. Just as "sharing economy" platforms such as Uber and Airbnb have recently been thwarted by legal rulings (Uber being compelled to recognise drivers as employees, Airbnb being banned altogether by some municipal authorities), privacy and human rights law represents a potential obstacle to the extension of data analytics. What is less clear is how the benefits of digital analytics might ever be offered to the public, in the way that many statistical data sets are. Bodies such as the Open Data Institute, co-founded by Tim Berners-Lee, campaign to make data publicly available, but have little leverage over the corporations where so much of our data now accumulates. Statistics began life as a tool through which the state could view society, but gradually developed into something that academics, civic reformers and businesses had a stake in. But for many data analytics firms, secrecy surrounding methods and sources of data is a competitive advantage that they will not give up voluntarily.

A post-statistical society is a potentially frightening proposition, not because it would lack any forms of truth or expertise altogether, but because it would drastically privatise them. Statistics are one of many pillars of liberalism, indeed of Enlightenment. The experts who produce and use them have become painted as arrogant and oblivious to the emotional and local dimensions of politics. No doubt there are ways in which data collection could be adapted to reflect lived experiences better. But the battle that will need to be waged in the long term is not between an elite-led politics of facts versus a populist politics of feeling. It is between those still committed to public knowledge and public argument and those who profit from the ongoing disintegration of those things.

✳✳✳✳✳