Developing students' confidence in using data in their journalism

By Chris Frost, Liverpool John Moores University

Abstract

Anyone who has been paying attention over the past two years will be aware of the importance of data. The SARS-CoV-2 Pandemic has often been portrayed in figures, whether with the daily death toll or the imminent collapse of the health service and local hospitals. The wide transmission of misinformation has also often been played out in figures regularly showing a complete ignorance of the use of data.

Journalists, even working for major publishers such as the BBC or national newspapers, are not immune to failings. FullFact.org, the factchecking campaign group, have had their hands full during the pandemic correcting misinformation in newspapers that should (and possibly do) know better. Readers end up ill-informed either because of poor journalism or unethical journalism. Training in some basic statistical skills and the confidence to handle data can only help improve our students' employability, ethics and their journalism.

Introduction

Data in the form of numbers or statistics trouble many students. Would-be journalists who

would never admit to being nervous about interviewing a senior politician or major celebrity have no trouble admitting to being scared of numbers.

Partly this is cultural – it is acceptable to admit you can't hack maths, occasionally even a matter of pride as being confident with figures seems far from fashionable in an era that sees an appearance on a reality TV show as the pinnacle of success.

However, nerdy or not, statistics, business accounts and other forms of number-based news are the day to day bread and butter of much modern news journalism and no journalism student can afford to ignore this. It is, therefore, our task as educators, to ensure they are fully confident with the basics.

Setting out to produce a full course on professional level statistics is unlikely to prove popular with many students, though, so it's important to see what can be reasonably taught in a few sessions, alongside other forms of writing; a few lecture and seminar aimed at giving students confidence building practice and a better understanding of what they should be attempting to do in reporting statistics.

Using data

Numbers are just a form of information and therefore evidence and hold no other mystical powers. It is only when they are combined with other statistics and comparisons are made that some students flounder. Hence the difference between percentages and percentage points often confuses and the different types of average are a mystery to many. Understanding the relative relationships between big and small numbers is another difficulty for all of us. Take money, for instance. Our daily lives tend to be measured in £10 or £20 notes. Even card payments rarely stray into the hundreds of pounds. A grocery till payment of £200 may no longer be rare but for most people it is still a suck-your-teeth moment and when was the last time one of your students made a payment in four figures or more? So how are we supposed to understand payments in millions or even billions?

Good journalism using figures and statistics as data requires a clear understanding of numbers and their relationships and calls for imaginative ways of explaining them to the public that are clear and easy to understand. Graphs, charts and diagrams are tried and tested methods of making statistics clearer but modern technology offers a much wider range of tools with video, audio, animation and even games on hand to explain issues alongside a clearly written text.

Common problems

There are a number of common problems with working with statistics. One identified by several authors of books that I would happily recommend for both lecturers and students is the question "Is this a big number?". Blastland and Dilnot point out that "zeros on the end of a number are often flaunted with bravado to impress or alarm." (2007, p6) The Chivers point out that there's no such thing as a big number "the bigness or otherwise of a number depends entirely on its context." (Chivers and Chivers, 2021, 63).

Big numbers

Governments are fond of offering apparently large sums of money in order to convince us that they take a problem seriously. Telling the public that the government is so concerned about maternity care, for instance, that it is putting an additional £12m into hospitals to improve the service care makes it sound like a lot of money. £12m is a large sum of money for us as individuals, a big lottery win perhaps, but it is small beer for government, covered by a one-off 25p tax on income. We've all lost more than that down the back of the sofa. The reality is that the Office of National Statistics tells us there were 615,557 live births in 2020 (ONS2021a) and so £12m is only £19.49 extra per birth in that year. If the government then points out that that funding is supposed to cover the next ten years of maternity care, then it's now only £1.95 per birth, a miniscule increase hardly likely to buy any significant additional care or equipment.

These are problems of scale and we need to get students to understand this if they are to explain the is-

sues involved to their readers. We've seen the action of scale often during the pandemic. Much was made of the enormous cost of test and trace. The government budgeted £37bn to be spent over the first two years of the pandemic and opponents condemned this huge cost for something that was widely considered to be a failure. However, it was a budget and only a fraction of the money, £5.7bn to November 2020 was spent and that was only expected to rise to £20bn by April 2021. So is this a big number? In government terms, no. To convert to the kind of figures we understand, the original £37bn budget over two years would equate to the typical worker being present with a £640 bill each year for two years. Very unwelcome, certainly but probably not life changing. In reality, of course, it was only half that.

Looking at the government's spending for 2019/20 the track and trace bill was about the same as that for housing, or culture or government administration. (UK Government 2021)

To take another example, according to a Credit Suisse report, the top one percent of households globally own 43 percent, or just under half, of all personal wealth. The bottom 50 percent only own one percent between them. One popular You Tube video also circulating on TikTok shows a novel way to bring this message home by representing the wealth of multi-billionaire Jeff Bezos as rice. Placing one grain of rice on a table as worth $100,000, the common millionaire only had 10 grains whilst Bezos, at $122billion, had a huge bagful of rice (Youtube 2021). Bezos could easily afford the cup of rice needed for his space ambitions whilst most people in the world don't even have one grain of rice. As a piece of journalism explaining inequality around the world it was an excellent demonstration.

Averages

It's difficult to talk about statistics without talking averages but averages are actually one of the most difficult things to imagine. Take average height, for instance. The average height of a man in the UK is five foot nine inches (175.3cm) according to the ONS and an average woman is five foot three inches. If we were to chart all the heights of people in the UK, we would find that the graph is bell shaped, with those averages at the peak. That type of average is called a mean average, the figure given when all values are added together and then divided by the number of values. But it is also the median (the middle figure in a range) and the mode (the most common figure). Men can be up to two foot shorter or two foot taller, but those extremes are very uncommon.

However other types of average are not so easy. The average household income in the UK is one that often vexes politicians. This is because incomes do not follow a bell curve. A total of 2.4m people earn below £10,000pa whilst the top one percent earn 8.3 percent of total incomes according to the ONS. This is what is known as equivalised individual household income, a figure made up of the total earnings of a household and then formulaically allocated to each member of the household to produce an equivalised individual income. (Europa 2021). The ONS chart shown below shows that the mean disposable household income is £36.900. That is the average after totalling all the incomes and dividing by the total number of values. The median income, the true middle point is £29,900 whilst the mode, the most commonly earned amount is £24,000. The mean average is dragged up by a relatively small number of people earning large incomes (ONS 2021b). The chart below ends at £80,000 simply because to add in all incomes would require a very long tail (at least 120 more columns) more than doubling the length of the chart. More than 22,500 households have an income of £674,000 or more, for instance.

Averages can be tricky things and often don't help explain reality. Most people earn less than the mean average and more earn less than the median than earn more. In this case, the mode, or most common number best describes typical UK income, although even then, where you live in the country would be one of several other factors affecting your lifestyle.

Qualifiers

Accuracy is an important element of journalism and leads the list of complaints to editors and regulators. When dealing with figures it is important, as with all journalism, to be clear about what it is one is talking about and only to match like with like when making comparisons.

The data for income above also has some important qualifiers that any report should make clear. First the income specified is income rather than just wages and includes interest and other unearned income.

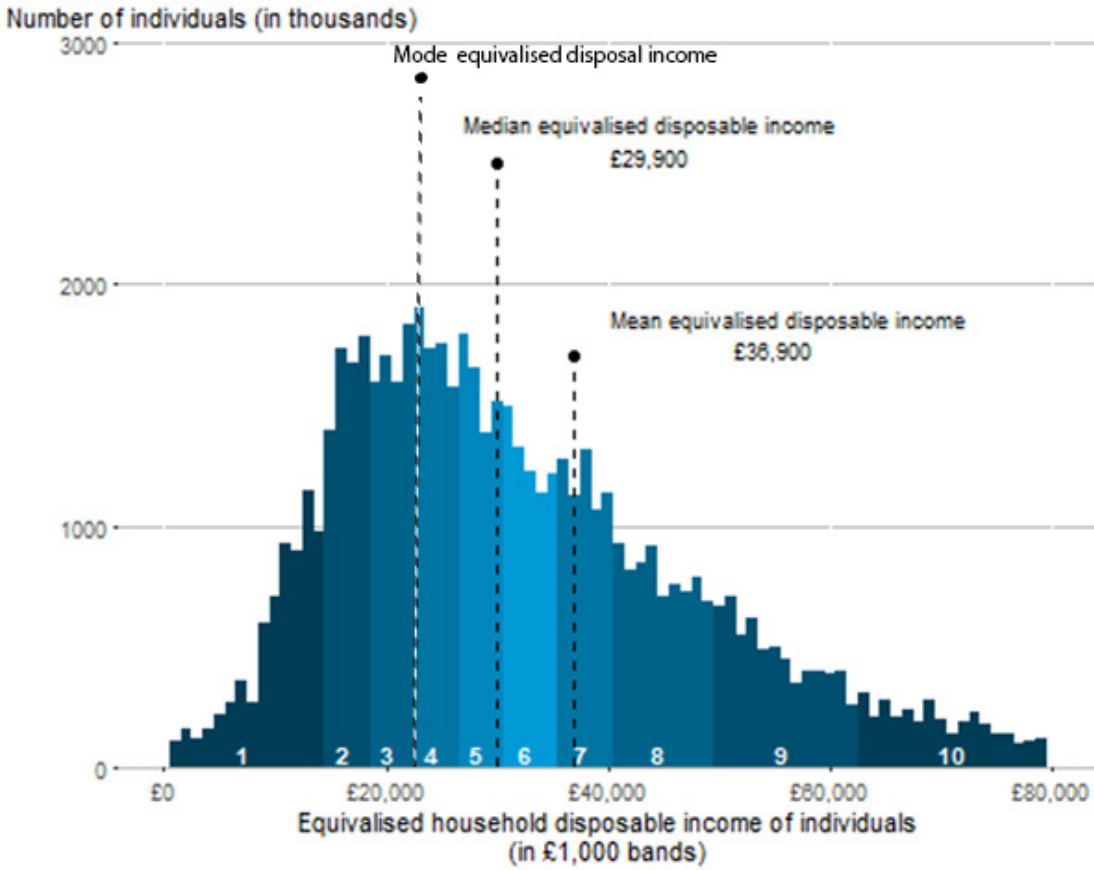

Figure 1: Disposable household incomes in bands of £1,000. (Ibid.)

Secondly, this is equivalised household income; that is all income coming into a household from adults living there whether a single person, a couple or a family with three or more earners formulaically allocated to individuals. A final qualification for this dataset is that it is UK wide. Many sets of figures apply to just England or Scotland or Wales and students need to be reminded to check geographic coverage.

Other issues

There are other statistical difficulties that students struggle with such as dealing with chance, causation v correlation, base rate fallacy, sample sizes and statistical significance that I do not have the space to discuss here, but that are equally important. Students should be advised to read the books recommended below (Blastland and Dilnot and Chivers and Chivers) and listen to the Radio 4 More or Less podcasts that discuss topical items in the news based on figures.

Data sources

Sources of data are important as are sources for anything. Politicians are an important and authoritative source but often play fast and loose with data as do many campaigners. Figures of all sorts are regularly bandied about yet turn out to be nonsense.

FullFact the fact checking web site revealed that during a debate on a private member's Bill about damages for adverse vaccine effects, Sir Christopher Chope, the Conservative MP for Christchurch, made a number

of misleading statements to the House. Introducing the second reading of the Bill Mr Chope talked about the Yellow Card scheme that allows doctors and the public to notify the authorities of potential adverse vaccine effects: "Essentially, what the Yellow Card scheme shows… is that there have been 435 reports of major blood clots and low platelet counts, including 74 deaths. It shows that there have been 767 cases of inflammation of the heart, a condition that is almost unheard of in medicine on a normal day-to-day basis. It shows that there have been some 35,000 reports of menstrual disorder, and there are all sorts of other effects set out in the comprehensive report. Very worryingly, it says that there are 1,632 reports of deaths having taken place shortly after vaccination." Mr Chope went on to say that there were "Hapless families – 10,000 of them or maybe more" who had suffered "real, serious damage".

FullFact said they did not know how Mr Chope reached the figure of 10,000 and he did not respond when contacted. The organisation also pointed out that the Yellow Card scheme was for reporting suspected adverse effects and that the data submitted cannot be used "to reliably say whether or not a condition was directly caused by the vaccine… With the early vaccination rollout concentrated on the most elderly and vulnerable in society, it's sadly unsurprising that there were '1,632 reports of death' (at the time of Mr Chope's claim) shortly after Covid-19 vaccinations—a point the MHRA has also emphasised."

Fullfact are a good source for checking stories and it is worth following them on social media.

Other good sources of data include the Office of National Statistics that carries figures on all aspects of UK life. Statisa.org is another good source of data, although it only allows a few free uses before seeking a subscription, and indexmundi.com carries stats from around the world. Similar organisations exist in a number of countries to give national figures. There are a number of other excellent sources available online.

Crosschecking data

Checking sources can be time consuming and, with looming deadlines, it can often seem easier to take information on trust.

A Guardian story reporting the conjunction of Saturn and Jupiter in 2020, claimed that Jupiter was at its closest to Earth at only 886,000km away. In fact the figure should have been 886million km. It is another problem of large numbers. There were sufficient zeros in the given number for the sub not to notice that it should have been one thousand times larger. Had those editing the column in the Guardian (it was removed very shortly after publication) remembered a basic pub quiz statistic they may have considered that our moon is already about half that distance from earth.

By learning a few basic numerical facts, a journalist can often check whether something is plausible or how big or small it is very quickly. Knowing that the Moon is typically around 240,000 miles or 430,000km from the earth, it would be easy to spot that someone had got the distance to Jupiter wrong. If one needs to know the precise distance at any time, then there is always Google. Knowing such basic data also has the added advantage of always being a good pub quiz winner.

So, the population of the UK is around 70million. Births in 2020 were as few as 615,557 and people start to die in significant numbers from their mid sixties, meaning a cohort of 85 year olds is as small as 265,000 but there are almost 940,000 55-year-olds and that means we can say that there are approximately 800,000 people in each year cohort from birth until 60. How many schoolchildren are there? There are 10 cohorts from 5 to 15, so 10x800,000=8m. In fact,according to ONS, there are 8.26m in that age group. Close enough to allow the use the 800k cohort as a rule of thumb guide.

The population of the USA is just under five times as many (329m), India is four times the size of the USA at 1.35bn and China a little larger at 1.44bn. Knowing basic, rounded, figures such as these can often prevent making silly mathematical errors.

Other easily remembered numbers around economics, population and geography are always useful.

Confidence building

Students' confidence can be built by developing exercises or assessments that use statistics. Developing such exercises is not hard but does require reasonable control of data sources. It is tempting to use current stories for such exercises and the ready availability of data from the pandemic makes that a serious candi-

date. Figures of cases, deaths and hospitalisations are published daily in newspapers and more detailed stats are available on the Government's Covid Dashboard (https://coronavirus.data.gov.uk). Asking students to find a news story from that data is good practice and can help them build confidence in reading and understanding statistics. The dashboard gives daily figures and access to past figures for cases, vaccinations, deaths and healthcare. This can make a useful early start to statistical stories. The data gathering and presentation is already done for you.

The dashboard can be used as a both a teaching aid and resource. The charts (as below) allow for discussion about interpretation and questions can be asked. For instance, why does the data dip every few days only to rise again three days later? A quick look soon confirms that cases fall on Fridays and drop further over the weekend, only to rise again, often dramatically, on Monday. We can perhaps suggest that people feel it more important to confirm infection when they need to attend work and risk infecting others rather than when they are relaxing at home with their family. Or maybe some of the testing laboratories only report their daily data on a Monday. I think we can discount the idea that the virus takes the weekend off.

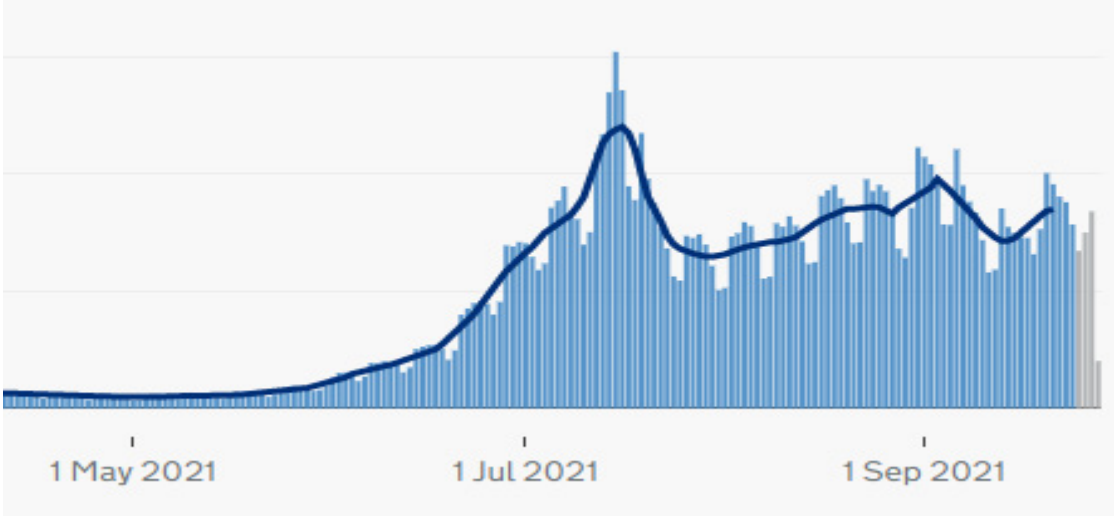

Figure 2: Tracking positive Covid tests (Coronavirus 2021)

The site allows for geographic investigation, great for local papers but also for pointing out that some areas are far more affected than others. Wales, the Midlands, Cumbria and the Scottish Borders were particularly badly hit compared to the South of England at the time of writing.

Vaccination figures can be closely examined. Are the anti-vaxxers having a significant effect or are they just a small group given way too much publicity as social outliers often are? Hospitals and healthcare are an additional area that can be examined. How is vaccination affecting hospitalisation, for instance? NHS England publishes data for daily hospital admissions by age band (NHS England 2021) This shows figures for admissions to hospitals in England up to Sept 2, 2021 at time of writing. For the last ten days average admission was:

Age	Avg number	Per year cohort (avg number divided by cohort size)
0-5	8	1.6
6-17	7	0.6
18-64	178	3.8
65-84	112	5.8
85+	37	Cohort size reduction too dramatic to be meaningful

Table 1: Hospital admissions in England during 2021

Interestingly the admission numbers for 6-17 year olds, which had started to climb through August, started to fall noticeably during September. This might be attributable to the start of the vaccination programme for those aged 16+ and the later programme for 12-16 year olds. Certainly the data shows that the large disparity seen between the older and younger age groups in admissions the previous autumn had been largely banished by the vaccine programme. Admissions for those under 18 remained fairly steady throughout 2021 but admissions for over 18s fell steadily as the vaccine programme progressed.

Media regulation data

Although the Covid data provides an excellent set of exercises in statistics stories, these rely heavily on data gathered and provided by official sources. The sources are reliable, if often some days old, but nonethe-less have been collected by someone else. For an exercise that takes data in a much rawer form and also allows students to examine issues of journalism I would suggest asking them to research complaints made about newspapers or broadcasters. This means looking at the websites of the different regulators such as Ofcom, the Independent Press Standards Organisation or Impress. Ofcom puts out regular bulletins about complaints and these are useful in themselves for guiding students in the use of the Ofcom broadcast code. But it is probably easier to manage the data that comes from IPSO in that there is less of it.

The rulings section of IPSO's website (see https:// www.ipso.co.uk/rulings-and-resolution-statements) allows access to all the adjudicated complaints listed in an index that gives details of the publication, the complainant, the code provisions concerned and the outcome. This index then allows access to the full adjudication.

The index allows filters to be used to limit the time, search for a particular complainant or publication or to limit by code provision. This allows some ability to search for particular categories of data. However, the search filters are particularly clumsy and the date filters reset to the most up to date time period between each filter setting change. The index probably works well for the occasional searcher of a specific complaint, but as a regular user of this index in my research, it is incredibly frustrating. Whatever the reason though, for its structure, it is possible for students to gather information on IPSO's performance at the same time as learning about why the public complains and what concerns them most.

It is also possible to download IPSO's annual reports (IPSO 2021). These contain data about the number of complaints and how they were handled. IPSO no longer publishes full details about which code provisions are most complained about.

This can make a useful exercise. IPSO gives the data to allow examination of such questions as "which publication is most complained about?" and "which code provision attracts the most complaints?". These questions can be limited to a particular year so that you can make the workload easier. In 2020, IPSO dealt with 30,126 complaints and enquiries and investigated 496 of them, adjudicating 292, so there is a substantial workload in just one year. Earlier years have fewer complaints; 2015 for instance had only 12,278 inquiries, investigated 307 and adjudicated 243.

The two lots of data, that in the annual reports and that in the rulings pages can of course be combined to evaluate IPSOs work and performance.

Data manipulation

Having gathered the data they require, students often seek advice on how to store and manipulate the data.

The simplest method is to store it on a checksheet, but this is not the most flexible of systems and few students would consider not using a computer. A checklist on a computer in Word is probably the easiest method with a table set up that either allows figures to be updated as information is added or that automatically totals data as it is added.

Better still is a spreadsheet, something with which most students are familiar that allows data to be inserted and totalled and comparisons to be made as required.

Best of all is a database, such as Microsoft's Access that allows the data to be entered into a table, item by item and then allows reports to be produced, extracting the appropriate data as required. However, this requires some experience and expertise with Access that few students (and possible few tutors) will have. Whilst students with an aptitude and a future requirement to gather data might find it useful to become proficient with Access, it is probably not worth the time required to expect other students to use it so I would recommend sticking to Excel as the usual source of data collection.

Once a data method is set up then it is simply a matter of entering data. Because of the clumsiness of the IPSO website, this can be a frustrating process and I would recommend keeping exercises short. For instance, seeking out which publication has the most complaints upheld over the lifetime of IPSO would require filtering by the full range of dates of IPSO's existence for each publication in turn with all the upheld adjudication options turned on. It is a laborious process so to help out, here is the league table of the top 10 to July 2021. Complaints upheld drop dramatically for subsequent publications.

IPSO Complaint totals

Newspaper	total complaints	Uph off	Uph corr/adj	resolved	U/h as %	Reject
Mail online	163	12	13	74	15	61
The Sun	115	17	12	28	25	58
The Times	106	15	10	19	24	62
Daily Mail	96	12	4	33	17	48
The Daily Telegraph	76	8	10	18	24	39
Express.co.uk	75	21	9	23	40	23
mirror.co.uk	69	9	5	16	20	38
The Sunday Times	53	11	5	12	30	24
Daily Express	53	12	7	7	36	27
Daily Record	52	7	5	12	23	27

Figure 3: IPSO complaint totals

Asking students to discover which code clauses are most often breached is another relatively easy use of the IPSO website. Again for your guidance, here are the figures to July 2021.

Total IPSO adjudications by type

02-Oct-21

Total individual complaints against a newspaper: 2023

Total of all code categories of complaint made: 3481

Total upheld:	449	22.2 %
Uph as offered by member:	256	12.7 %
Uph with correction/adjudication:	193	9.5 %
Resolved:	505	25.0 %

	Total complaints	Total as percentage of all complaints	Percentage of total upheld	percentage of category upheld
Accuracy:	1729	85.5	413.0	23.9
Privacy:	681	33.7	111.0	16.3
Harassment:	260	12.9	49.0	18.8
Intrusion:	202	10.0	29.0	14.4
Children:	122	6.0	17.0	13.9
Children sex cases:	13	0.6	5.0	38.5
Hospitals:	12	0.6	0.0	0.0
Payments to criminal:	6	0.3	0.0	0.0
Listening devices:	88	4.3	10.0	11.4
Victim of sexual assault:	34	1.7	13.0	0.0
Financial journalism:	5	0.2	1.0	20.0
Confidential sources:	32	1.6	9.0	28.1
Discrimination:	152	7.5	30.0	19.7
Payments to witnesses:	4	0.2	0.0	0.0
Suicide:	29	1.4	3.0	10.3
Reporting of crime:	6	5.5	20.0	17.9

Note: many individual complaints cover several clauses of the code hence the percentage of complaints may not total 100.

Figure 4: Complaints made against each code provision.

Conclusion

Anyone interested in ethical news reporting understands the importance of being able to question politicians, scientists and other professionals about statistics they use in their work. Reporting stories based on statistics or using statistics extensively and accurately requires journalists to be able to query, manage, manipulate and report statistics and other data accurately and fairly to properly inform readers.

But many students are nervous of dealing with figures, they found maths confusing at school and have not had a comfortable management of figures inculcated into their working methods. What is being laid out here is not a maths course, but an attempt to help develop that feeling for accuracy and the clear identification of what the stats are really telling us that sets the standard for the good journalist.

Reference

Blastland, Michael and Dilnott, Andrew (2007) The Tiger That Isn't: Seeing Through a World of Numbers Glasgow: Bell and Bain Ltd

Chivers, Tom and Chivers, David (2021) How To Read Numbers: A Guide to Stats in the News (and knowing When to Trust Them) London: Weidenfield and Nicholson

Coronavirus (2021) https://coronavirus.data.gov.uk/details/cases

Europa (2021) https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Equivalised_disposable_income

Frost, Chris (2021) 'Pandemic and the Media: Counting the Pandemic's Human Tragedy' in Mair, John and Keeble, Richard (Eds) The Pandemic: A Year of Mistakes? London: Bite-sized Books

Frost Chris (2014) 'Journalism Ethics Through Accuracy in Numbers' in Keeble, Richard and Mair, John (eds) Data Journalism: Mapping the Future London: Abramis

IPSO (2021) www.ipso.co.uk/monitoring/annual-reports/

NHS England (2021) https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-hospital-activity.

ONS (2021a) www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/articles/provisionalbirthsinenglandandwales/2020andquarter1jantomar2021#:~:text=Based%20on%20birth%20notification%20data,annual%20decrease%20in%20a%20row

ONS (2021b) www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyear2020

UK Government (2021) www.gov.uk/government/publications/how-public-spending-was-calculated-in-your-tax-summary/how-public-spending-was-calculated-in-your-tax-summary

Youtube (2021) https://www.youtube.com/watch?v=_3VFmmqFCoM