

If the line fits: inequality, statistics and The Spirit Level

Richard Harris, School of Geographical Sciences, University of Bristol

Introduction

“Gap between Britain’s rich and poor now wider than ever” (Headline: The i, March 14, 2016)¹

An important area of social and economic research is economic inequality between people and places, and the sense it is growing. Two reports from Oxfam have made the headlines. The first said that the combined wealth of the worldwide richest 1 per cent is equal to the total for the remaining 99 per cent.² A second said that the richest 1 per cent of Britons has received more than one-quarter of the £4 trillion increase in national wealth since the year 2000.³

What the reports claim is that there is a growing gap between the wealthiest individuals and the rest. This is not to say that the differences between countries are growing also. In fact, the UN Millennium Development Goals Report (2014) says, “in 1990, almost half of the population in developing regions lived on less than \$1.25 a day. This rate dropped to 22 per cent by 2010, reducing the number of people living in extreme poverty by 700 million.”⁴ There is a long way to go before extreme poverty is eliminated and all people can claim their right to food, safe drinking water, shelter, protection from violence, and the opportunity for employment. Yet, significant progress has been made.

Nevertheless, whereas inequalities *between* countries seem to be decreasing, those *within* countries are said to be expanding, with some commentators voicing concern that the gap between the wealthiest and the poorest has become too great. A report by the think-tank, The Equality Trust, suggested that the effects of inequality can be measured through its impact on health, wellbeing and crime rates, and that it is costing the British economy more than £39bn a year.⁵ In no society is it likely that everyone will have an exactly equal share of the national income or have the same wealth. Nevertheless, the talk is about the growth of the ‘super rich’.

A growing number of academic books have been published about inequality, its causes and consequences, and what might be done to reduce it. Some of these are by the Oxford geography professor and social commentator, Danny Dorling⁶ (RGS-IBG School Members can view a lecture by Danny Dorling about inequality via www.rgs.org/SchMemArea). The purpose of this report is not to take a political position on the national and international policies that, depending upon your point-of-view (and who you read), fuel or restrain social and economic inequalities; nor about

¹ You can view the headlines for that day at <http://news.sky.com/story/1659220/mondays-national-newspaper-front-pages>

² Oxfam (2016) An Economy For the 1%. <https://www.oxfam.org/en/research/economy-1>

³ Oxfam (2016) Ending the Era of Tax Havens: Why the UK government must lead the way. <http://policy-practice.oxfam.org.uk/publications/ending-the-era-of-tax-havens-why-the-uk-government-must-lead-the-way-601121>

⁴ United Nations (2014) The Millennium Development Goals Report. <http://www.un.org/millenniumgoals/2014%20MDG%20report/MDG%202014%20English%20web.pdf>

⁵ <http://www.theguardian.com/society/2014/mar/16/inequality-costs-uk-billions>

⁶ For example, Dorling D (2014) Inequality and the 1%. London: Verso Books.

whether economic inequality is an unfortunate but better-than-the-alternative by-product of trade and economic growth. This is something you could discuss in class, perhaps in the context of Sir Anthony Atkinson's fifteen proposals to tackle inequality (which does, of course, presuppose that something should be done)⁷.

Instead, this report takes inequality, and data about it, as a context for discussing a number of statistical ideas and concepts, understandings of which are important not just for geography but across the sciences, social sciences and humanities. The case for why knowledge of quantitative methods is important for geographers is made in another report, available [here](#). Whereas that report provided a general introduction to quantitative geography, this report looks at some specific methods, taking as good practice the opportunity to teach them not in the abstract but embedded in a topic of substantive interest to geography within the GCSE and A Level specifications. In doing so, it will be possible to bring greater understanding to key areas of geographical enquiry, be it inequality in this case study or other areas of geographical study for which data provide knowledge and inform discussion.

How is income distributed?

To talk of an inequality is to say that two or more 'things' are not of equal value. In the case of economic inequality we are saying that income or wealth are not the same for different groups within society. For example, there has been a disproportionate rise in the number of 'ethnic minority' groups in England and Wales that live in areas of high unemployment and, where work is available, it is characterised by part-time jobs. (The increase in the number who find themselves in these neighbourhoods has risen at a rate greater than the growth of these population groups).⁸ A 2007 report by the Joseph Rowntree Foundation showed that the around two-fifths of people from ethnic minorities lived in income poverty, twice the rate for white people.⁹

Income is not the same as wealth. Income most easily can be understood as the amount declared on a tax return (assuming all earnings are declared) – it is the amount received through wages and other earnings such as shareholdings, bonds, capital gain through the sale of second homes, and so forth. Factors affecting peoples' 'take home' income include their salary, whether they work full or part time, tax rates and any bonuses they receive. Some nations also have a legal minimum for salary levels¹⁰, though there has been concern about the use of 'zero hour contracts' to reduce wages and other in-job benefits.¹¹

Wealth is wider and includes other assets such as the value of a home, of cars, of art and jewellery, and of things that are owned or invested in and that have a monetary value (including savings, investments, pensions and shares). Income, because it forms the basis of a tax return, is the easier to measure. It may reasonably be assumed that wealth inequalities exceed income ones.

Because not all incomes are the same, they have a **distribution** – some incomes are above the **average**, and some are below it. Someone will have the lowest income, which is the **minimum** for the data, and someone will have the highest, which is the **maximum**. From the minimum to the maximum is the **range** of the data.

⁷ see <http://www.tony-atkinson.com/the-15-proposals-from-tony-atkinsons-inequality-what-can-be-done>

⁸ Harris R, Johnston R, Manley D (2015) The changing interaction of ethnic and socio-economic segregation in England and Wales, 1991–2011. Ethnicities, in press.

⁹ <http://www.poverty.org.uk/reports/ethnicity.pdf>

¹⁰ In Britain the national minimum wage in 2015 was £6.50 for 21 year olds and over, and £5.13 for 18–20 year olds. www.gov.uk/national-minimum-wage-rates

¹¹ <http://www.independent.co.uk/news/uk/politics/one-million-british-workers-will-soon-be-relying-on-zero-hours-contracts-for-their-main-job-a6921721.html>

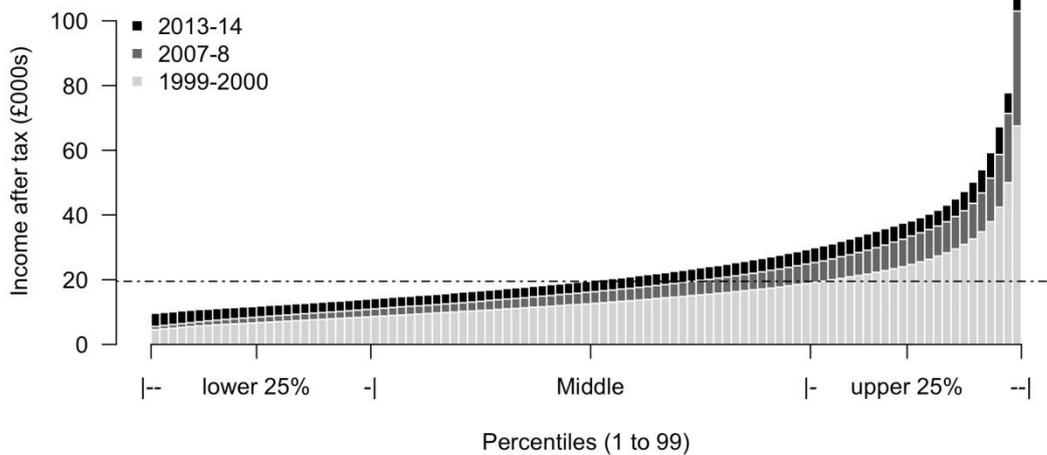


Figure 1. The income distribution, after tax, of UK taxpayers in the tax years 2013–14, 2007–8 and 1999–2000. The median income in 2013–14 is shown by the broken line.

Figure 2 plots the data in a different way, using **histograms**. The skew in the data is now more evident – more frequently taxpayers have incomes to the left and lower end of the charts but there are a few with much higher incomes that appear to the right. Over the time period, incomes have increased: observe how the bars for 2013–14 are higher than for other years and that the distributions shown by the histograms shifts rightwards. This does not mean that people are necessarily wealthier. To know the real worth of the ‘pound in their pocket’ we have to adjust for inflation (see below). Nevertheless, we can still see that the gap between the richest and the rest does seem to have grown from 1999–2000 but perhaps not from 2007–8, after the financial crisis. The **income ratio** between the highest earners (at the 99th percentile) and the middle group was 5.5 in 2013-14, telling us that the highest earners earned 5.5 times the median average. That ratio was greater than in 1999-2000 (a ratio of 5.4:1) but less than in 2007-8 (a ratio of 6.4:1).

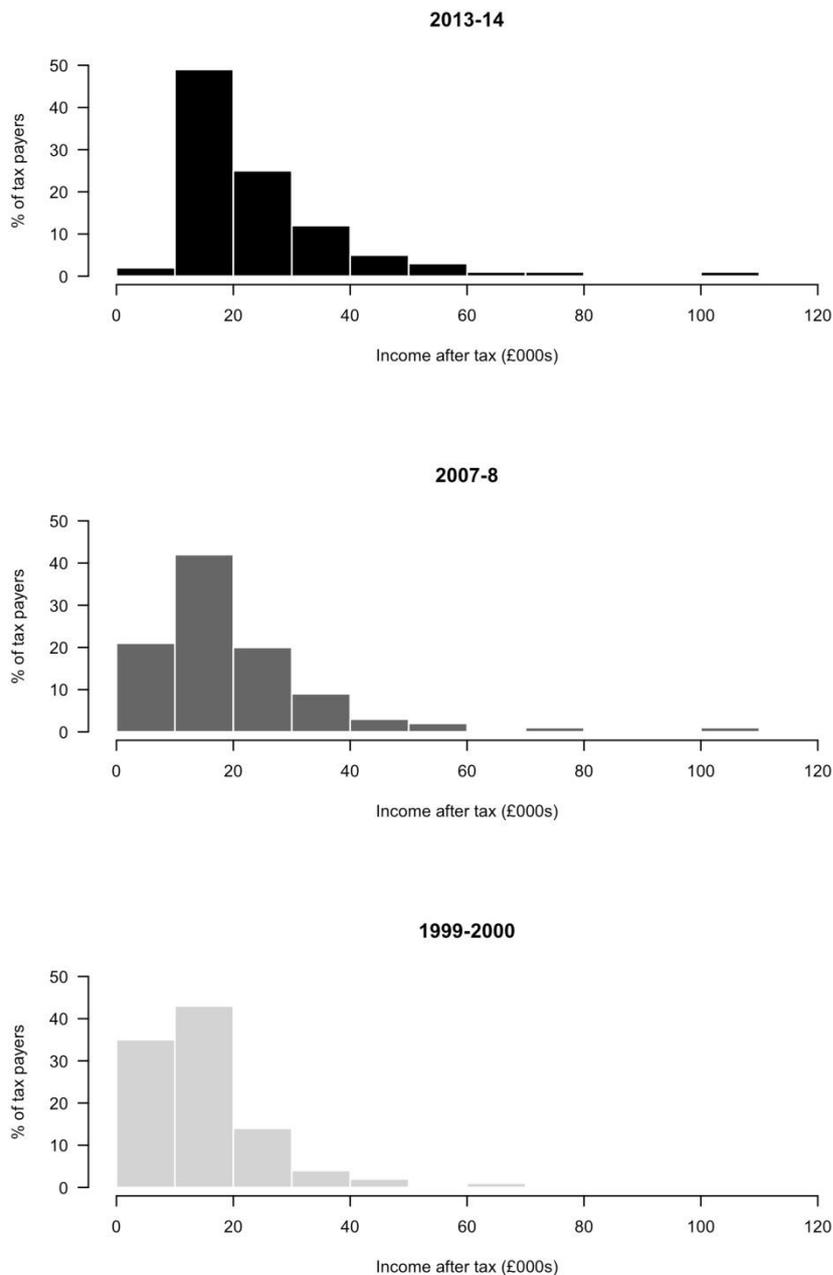


Figure 2. The income distributions in each of the three tax years. The distribution is skewed with a ‘tail’ of highest earners.

One way to adjust for inflation is to look at the annual change in the consumer price index.¹⁵ If a person receives a salary of £20,000 for two consecutive years but the price of consumer goods rises by 1.5 per cent over the period, then in real terms that person’s salary (their buying power) has shrunk because of the price increases. Figure 3 shows, as a series of **boxplots**, the incomes adjusted (somewhat crudely) for inflation for all years in the data set.

¹⁵ Inflation is a measure of whether the price of goods and services in an economy are increasing (rising inflation), or decreasing (negative inflation)

Boxplots are a good way of illustrating the median, 'mid-spread', range and any skew of the data, and also to identify potential outliers. The median is the vertical line near the centre of the box. The width of the box is showing how spread out the middle 50 per cent of the data are, from the first quarter (first quartile, also 25th percentile) to the third quarter (third quartile, 75th percentile). The distance from the first to third quarters of the data is known as **the interquartile range**. We might regard this as the typical range for the data. However, the actual range is greater and that is what 'the whiskers' extending out from the box are indicating.¹⁶ In some circumstances they will extend to the minimum and/or maximum values but not if the data are skewed with what are then highlighted as outliers (the square blocks).

For the incomes data we can see that the whiskers extend out to the right of the median further than they do to the left but even then there are outliers beyond them. It is not surprising that the skew is at the higher end: the lowest taxable income cannot be less than zero (actually, some value greater than zero because of the minimum wage and also various forms of social security benefit), whereas the maximum income is, in principle, unlimited. Of relevance is the amount of skew, which is indicative of the level of income inequality. We can see that the highest incomes were steadily pulling away from the rest until the time of the financial crisis in 2008. (The data for 2008–9 are missing). They then fell back but may be rising again. Looking closely at the median income, we can see evidence of a decline in real wages after 2009–10, when the median and the interquartile range fall.

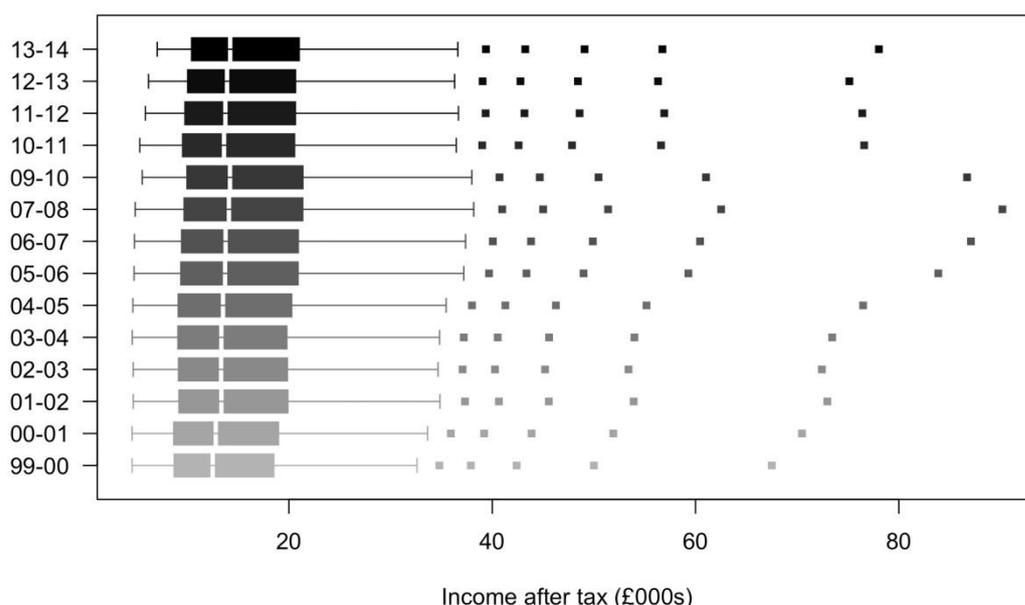


Figure 3. The income distributions shown as box plots for the tax years 1999-2000 to 2013-14 (the year 2008-9 is missing).

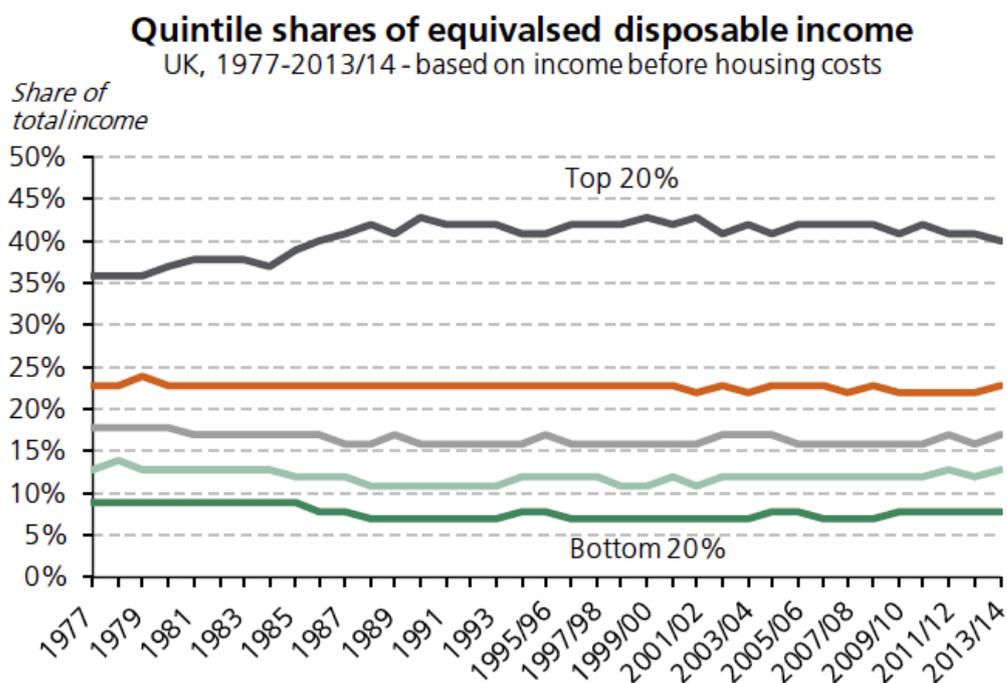
Figures 4 and 5 offer a longer-term perspective; both appear in a House of Commons briefing paper (their original sources are credited below them).¹⁷ Figure 4 shows the share of the total national incomes going to each of five groups: the most wealthy 20 per cent; the next 20 per cent; the next; and so forth down to the bottom 20 per cent. In a financially equal society each group would receive a share of their income in proportion to their size, and since the size of each group is equal (each contains 20 per cent of taxpayers), that expected share would be 20 per cent. In fact, the 'top' 20 per cent regularly got double that, and the 'bottom' 20 per cent, less than half.

¹⁶ A box plot is sometimes known as a stem and whisker plot

¹⁷ McGuinness F (2016) Income inequality in the UK.

<http://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7484>

The top and bottom grew apart from the 1970s to the early 1990s but after that the gap appears to stabilise, perhaps even narrow.



Source: ONS, *The Effects of Taxes and Benefits on Household Income, financial year ending 2014*, using data from the Living Costs and Food Survey and Family Expenditure Survey

Figure 4. Graphic published in a House of Commons Briefing Paper about Income Inequality in the UK.

Figure 4 might imply that concern about growing inequality isn't warranted in the UK, with the distribution of wealth between the five groups being relatively static for the last 15 years (with a slight uplift for the middle three groups in 2013–14). However, what happens when we look within the top 20 per cent and concentrate only on the most wealthy amongst them? Figure 5 provides the answer and it is graphs like this that raise concern about the super-rich growing apart from the rest. As a trend, it seems clear that from 1977 the share of disposable income going to the top 1% has risen in all but four years.

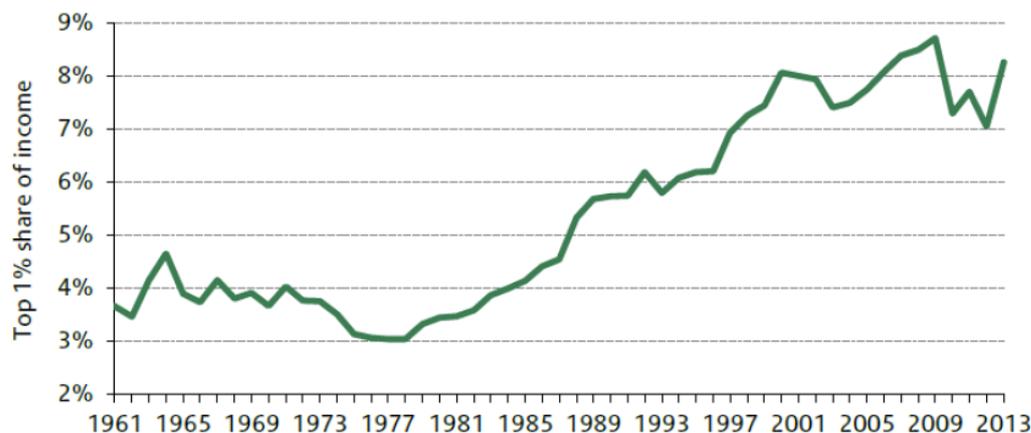
However, we should add a few caveats. First, even within the top 1 per cent there is growing inequality – for example, the difference in earnings of the top 0.1 per cent and the next 0.9.¹⁸ An infographic by Oxfam shows a falling number of billionaires who together own the same wealth as half the world's population: 388 in 2010 to 62 in 2016.¹⁹ Even amongst the 'super rich' the richest seem to be growing apart. Second, the trend does not mean that the same people are getting wealthier year-on-year. It is possible that people are dropping in and out of the wealthiest groups.

¹⁸ See <http://www.theatlantic.com/business/archive/2014/03/how-you-i-and-everyone-got-the-top-1-percent-all-wrong/359862/> for a commentary about this in the American case.

¹⁹ <http://www.oxfam.org.uk/media-centre/press-releases/2016/01/62-people-own-same-as-half-world-says-oxfam-inequality-report-davos-world-economic-forum>

Share of equivalised disposable household income going to top 1% of individuals by income

Great Britain, before housing costs



Source: chart taken from Institute for Fiscal Studies, *Living Standards, Poverty and Inequality in the UK, 2015*

Figures have been calculated by Institute for Fiscal Studies based on data from the Family Resources Survey and Family Expenditure Survey

Figure 5. Graphic published in a House of Commons Briefing Paper about Income Inequality in the UK.

The Spirit Level

The Spirit Level, by Richard Wilkinson and Kate Pickett has been an influential book about the potential social consequences of economic inequality within countries. Its subtitle, 'Why More Equal Societies Almost Always Do Better' makes no secret of their point of view. Whether you agree with that or not, there can be no denying that their publication provides a master class in using relatively simple statistics – ones that can be easily visualised and comprehended – to send out a clear and challenging message, and to promote debate.

The basic tool in their book is a **scatter plot**, like the one shown to the left of Figure 6. It uses some of the authors' data (available for a donation from The Equality Trust).²⁰ What it suggests is a **negative relationship** between life expectancy and income inequality, where income inequality is measured by comparing how much richer the top 20 per cent of people are when compared to the bottom 20 per cent. The word negative is not a social comment on the nature of the relationship (although that might be valid); it means only that there is evidence of a **negative correlation**: above-the-mean values of income inequality are associated with below-the-mean values of life expectancy, whereas below-the-mean values of income inequality are associated with above-the-mean values of life expectancy. In short, countries with higher income inequality tend to have lower life expectancies according to the data here. Whether one **causes** the other is a matter for debate. Here we are just highlighting the association, not the causality. An effective way to highlight it – and a second statistical tool in Wilkinson's and Pickett's book – is to add a **line of best fit**, as in the right of Figure 6.

²⁰ <https://www.equalitytrust.org.uk/civicrm/contribute/transact?reset=1&id=5>

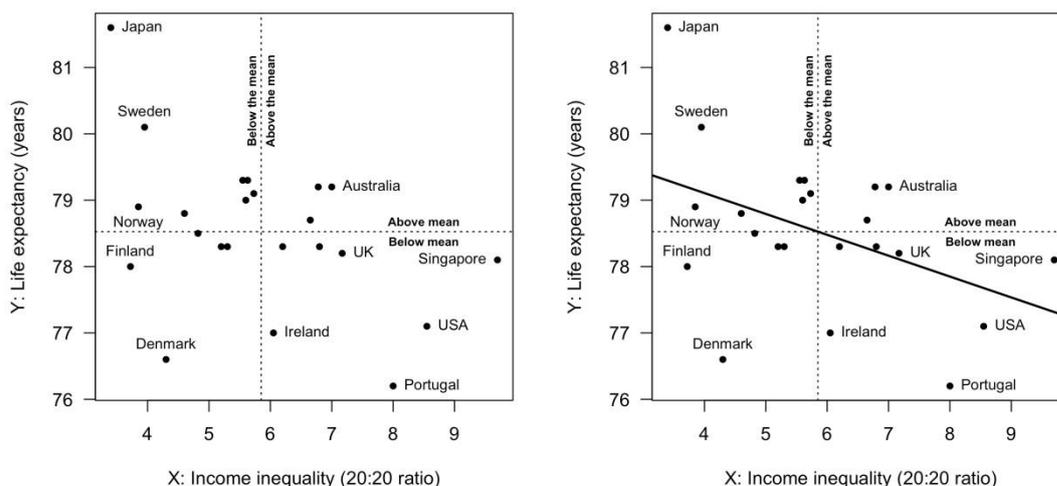


Figure 6. Using data from The Spirit Level book with a scatter plot and a regression line (right) to imply a negative relationship between income inequality and life expectancy in various countries.

The line of best fit is a **regression line**. It is a line of best fit, not of perfect fit, which is evidenced by the variation around it (some of the points are above the line, some are below it). The idea of regression is to place the line in the position that minimises that variation – to get it as close to as many of the points as possible. There are different ways of plotting a regression line from placing it 'by eye' to mathematically calculating its position. How this is done is not something we need to go into here. We can simply accept it as something that provides a summary of the relationship between income inequality and life expectancy for the data shown.

Of the two variables in the scatter plot, it is customary to describe the one along the horizontal axis as the **X variable** (also the predictor and independent variable), and the one along the vertical axis at the **Y variable** (the dependent variable and the response). When the regression line is fitted, the values of the X variable are used to explain (and to predict) the values in the Y variable. That implies that X leads to Y not the other way around. The Spirit Level book considers a number of relationships, showing, for example, how increased inequality is associated with decreased social trust but increased levels of obesity.²¹ Whether one is causative of the other is a matter of some debate. The regression line in Figure 6 is downwards sloping – the gradient, and therefore the relationship between X and Y, is negative. If the gradient was in the other direction we would describe the relationship between X and Y as positive.

A feature of the regression line in Figure 6 is that it is **statistically significant** at an agreed upon level. The concept of statistical significance takes a little explaining and getting used to. Many statistical methods, including chi-square, the Mann Whitney U test and Spearman's rank correlation (each of which have a long history in geography curricula) generate a **test statistic** – the number that arises as a result of the calculations that are followed in order to complete the test. In very loose terms, the interest is then in whether the test statistic has arisen due to chance or suggests something more concrete. For the regression line, the test statistic is related to the gradient of the line, and we ask whether the gradient might actually be zero, given what we know about the variation around the line, as well as the number of **observations** in the data set (there are 23 in the Spirit Level data, which is the number of countries shown in Figure 6).

²¹ The relationships they look at can be viewed and discussed by downloading the slides at <https://www.equalitytrust.org.uk/spirit-level>

One means that the two variables are interchangeable; one can be used as a substitute for the other because essentially they are measuring the same thing.

With Japan removed, the value of -0.319 is not strong but still indicates a moderate degree of correlation. Some critics of The Spirit Level have questioned whether some of the apparent social effects of income inequality are evidence of a general relationship or driven by unusual cases (such as Japan). These arguments and the authors' defence are interesting examples of how quantitative approaches don't necessarily settle debates but they do raise and allow us to participate in important areas of discussion.²⁵

See for yourself

The relationship between income inequality and life expectancy can be seen in the accompanying [Excel spreadsheet](#). The data both include and omit Japan.

- By looking at the columns of data, can you identify Japan in the first scatter plot (the one that includes Japan)?
- In what way does Japan seem to be unusual in comparison to the other countries?
- Can you see how it is an outlier in the sense that it seems a long way out from the regression line of best fit and also from where the other points (the other countries) cluster on the scatter plot?
- Compare the equation of the regression line with and without Japan in the data. What effect does removing Japan from the data have on the gradient of the line?
- As the line becomes flatter, what happens to the p-value and to the statistical significance of the relationship?
- Is there a stronger correlation between income inequality and life expectancy with or without Japan in the data? (A stronger correlation is one that moves closer to -1 in the case of a negative correlation or to $+1$ in the case of a positive correlation)
- A correlation between -0.3 and -0.5 (or between 0.3 and 0.5) is sometimes described as moderate. Is this true of the relationship between income inequality and life expectancy even omitting Japan from the data?

Closing comments

This report has used debates about rising income inequality in the UK and elsewhere as a way into describing some important statistical methods and principles, how they might be applied in geographical research, what they reveal, how they help with the interpretation and analysis of data, and their potential limitations. These methods include visual and numeric methods to describe the centre (average) and spread of data, to consider their distribution, to identify skew, to look at the relationship between variables, and to introduce the idea of statistical significance (as well as some of its shortcomings). An important underlying principle is that statistical methods are far better taught within a geographical context – in relation to a topic of substantive geographic interest – than in isolation from the sorts of themes and debates with which geographers engage.

The quantitative ideas covered were: averages (mean and median), distributions, the range (minimums and maximums), outliers, percentiles, skew, bar plots, histograms, (income) ratios, boxplots, interquartile range, scatter plots, correlation, regression lines of best fit, X variables, Y variables, statistical significance, test statistics and p-values.

²⁵ See, for example, <http://www.policyexchange.org.uk/publications/category/item/beware-false-prophets-equality-the-good-society-and-the-spirit-level> and <https://www.equalitytrust.org.uk/authors-respond-questions-about-spirit-levels-analysis>

.....

Follow-up activities

A collection of slides in regard to The Spirit Level book and exploring some of the possible consequences of inequality can be downloaded from <https://www.equalitytrust.org.uk/resources/the-spirit-level>. Take a look at them and see if the class is persuaded by the regression relationships they show. Can the class offer explanations as to how inequality can cause the outcomes that are suggested (assuming that it does)?

The Spirit Level has inspired a documentary film: The Divide. You could watch and discuss the trailer at <http://thedividedocumentary.com>

The documentary Don't Panic – How to End Poverty in 15 Years argues, with data and visualisation that “that recent global progress is “the greatest story of our time – possibly the greatest story in all of human history.” It can be viewed at <http://www.gapminder.org/videos/dont-panic-end-poverty/>

About the author

Richard Harris is a Professor of Quantitative Social Geography at the School of Geographical Sciences, University of Bristol, where he also is director of the University of Bristol Q-Step Centre, part of a multimillion pound national initiative intended to provide a step change in the quality of quantitative training provided to social science students. He also is author of the books Quantitative Geography: the basics (published by Sage, 2016), and co-author of Statistics for Geography and Environmental Science (Routledge, 2011). He was the recipient, in 2014, of the Royal Geographical Society's (with IBG) Taylor & Francis Award for excellence in the promotion and practice of teaching quantitative methods.

This project was funded by the Nuffield Foundation, but the views expressed are those of the author and not necessarily those of the Foundation.