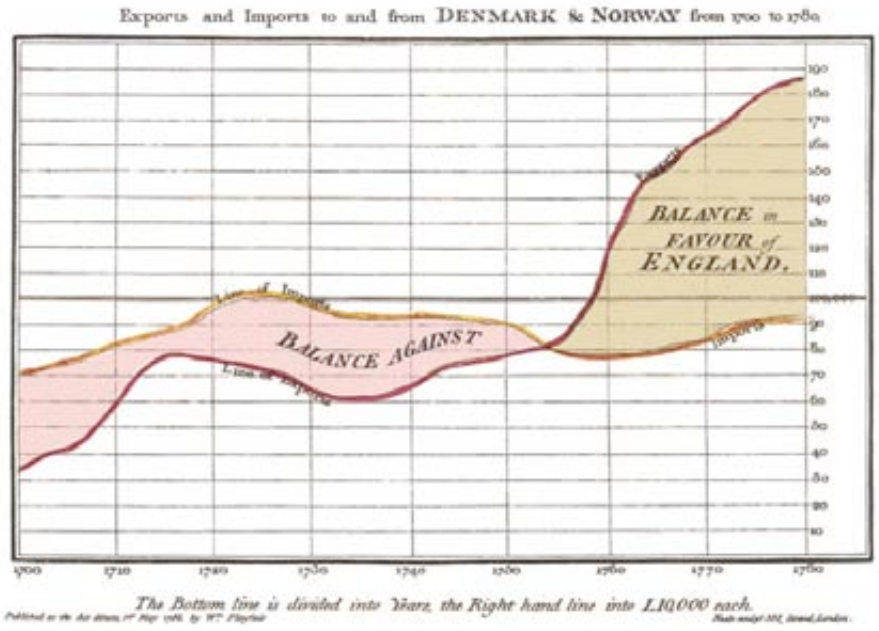
Integrated Content for User Engagement and Statistical Literacy

Smith, Alan
Data Visualisation Centre, Office for National Statistics (UK)
Segensworth Road
Titchfield PO15 5RR, United Kingdom
Alan.smith@ons.gsi.gov.uk

Introduction

It is over two hundred years since Playfair (1801) wrote "*As knowledge increases, and transactions multiply, it becomes more and more desirable to abbreviate and facilitate the modes of conveying information from one person to another and from one individual to many*". Despite its age, such a statement could have been written for the Age of the Internet equally as well as the Age of Enlightenment. Playfair used those words to justify his graphical innovations – line charts (Figure 1), bar charts and pie charts – arguing, "*Statistics is the driest of all subjects, unless the mind and imagination are set to work*". Playfair's imagination has undeniably had a lasting impact on the way statistics are presented.

Figure 1 – William Playfair line chart from 1786



While Shaughnessy (2010) argues that statistical literacy is "*much more than the ability to read graphs*", there is no doubt that visual methods of presenting statistical data allow and encourage a wider audience to engage with official statistics. Bolton (2010) argues that although statistical literacy is linked to reading literacy, "*it is also directly connected to how data are displayed in a graphical format as this is one of the main ways in which statistics are used.*" Nevertheless, traditional static images have their limitations. While the use of geometric visual variables (position, size and angle) to represent quantitative data may allow readers to readily recognise differences in magnitude and proportion, they do not automatically facilitate statistical reasoning – the informational implications of the variation being displayed. Playfair

himself was acutely aware of this and skilfully introduced limited annotations using text directly on his charts to aid understanding (e.g. in Figure 1 above, "Balance in Favor of England"). But there are real restrictions on how much supplementary text can be added directly to such images – especially with complex, large or rich datasets – without reducing the clarity or the impact of the quantitative visuals.

"Modes of conveying information"

Using Playfair's terminology, it is the "mode" of conveying information, not our ability to "abbreviate and facilitate", which has changed the most in the two centuries since his graphical innovations: Playfair himself would undoubtedly recognise most statistical graphics in a contemporary newspaper as direct descendents of his fine charts – but would probably find it more difficult to adjust to the concept of an Internet-enabled web browser as a replacement for copperplate engraving.

Web literacy is now seen as essential skill for progression in the modern labour market: ONS statistics show that the rate of Internet use in the UK decreases in line with income. In 2010, 98% of people with an income over £41,600 had used the Internet, but just 69% of adults with an income of less than £10,399 had done so. The same report found that just 45% of adults without any formal qualifications had used the Internet, compared with 97% of those with a degree. These figures suggest that the demands of the labour market are at least playing a part in increasing web literacy, reflected in the transformation of the UK communication infrastructure.

In 1998, the ONS Family Expenditure Survey estimated that 9% of all UK households had an Internet connection. By 2010, that figure had risen to 73% (in fact, a quarter of all households had a broadband internet connection installed between 2006 and 2009 alone). In 2010, 30.1 million adults in the UK (60%) accessed the Internet every day or almost every day. Although there remain valid concerns about those who are not accessing the Internet, the inherent risk of a so-called 'digital divide', there has been a rapid progression in the overall 'web literacy' of UK society. How does that compare with the skills required to deal with statistical information?

Statistical Literacy

Statistical literacy is widely recognised as a key skill for individuals to participate fully in a modern society. The ability to interpret, reason, evaluate and communicate statistical information has arguably never been more a more critical part of participative citizenship. But statistical literacy is difficult to measure directly.

Numeracy skills are often cited as a pre-requisite to statistical thinking, Bolton (ibid) reasoning that "*if someone is innumerate they are much more likely to be statistically illiterate*". This is a cause for concern in the UK because, for example, the Wales Assembly Government estimated in 2006 that 46% of the adult population of Wales lacked Level 1 numeracy skills. Based on these figures, UK society would now appear to be happier using a computer than dealing with numbers.

It is not hard to find examples where poor numeracy skills have manifested themselves in public. In 2007, the UK's national lottery operator Camelot introduced 'Cool Cash', a new lottery scratchcard. Players were required to identify if their allocated number was lower than a given number. As both numbers were often negative (part of the card's wintry design theme), comparing them often proved difficult for many and the card was subsequently withdrawn from sale. According to the Manchester Evening News, one of those confused customers noted that "*the card doesn't say to look for a colder or warmer temperature, it says to look for a higher or lower number*". In this case, poor numeracy (and lack of appropriate

contextualisation) prevented the player from reaching a goal (identifying a warmer or colder temperature) which they were perfectly capable of achieving.

In the UK, the Royal Statistical Society (RSS) has recently launched its GetStats campaign, a 10-year (2010 to 2020) campaign which aims to "*close the gap between the knowledge, skills and perceptions of statistics*". GetStats is a 'networked campaign', aligning with other initiatives. Identifying the need for a 10-year campaign reflects the difficulties in achieving significant improvements in statistical literacy, yet acknowledges the societal benefits of achieving those improvements.

Acutely aware of the gap between data and meaningful information, the UK Statistics Authority (the independent body operating at arms length from government, responsible for the ONS and independent scrutiny of all official statistics produced in the UK) produced a statement in 2010 on The Value of Statistical Commentary, outlining the responsibilities of producers of official statistics under the Code of Practice for Official Statistics, introduced in January 2009. Under the Code, producers are obliged to "*prepare full and frank commentary and analysis that aid interpretation including…factual information about the policy or operational context*…[and] *formats for the presentation of statistics in graphs, tables and maps that enhance clarity, interpretability and consistency*". The statement concludes "*the public has a right to know the statistician's understanding of the messages from the statistics, just as they have a right to the data itself. It is common, and proper, for statisticians in government departments to brief policy colleagues on the substance of the statistics. That knowledge should, as a matter of principle, be shared more widely*". The challenge connected with this last statement is *how* to share that knowledge more widely, given the issues of numeracy and statistical literacy One possible approach is to revisit the form and content of official statistical outputs and check to see if they are working hard enough to engage the wider public and promote statistical literacy.

Towards richer content

Many official statistical outputs have transitioned from paper to print in name only – articles still consist broadly of static text, static tables, static charts and static maps – they are, in effect, no more than functional abbreviations of their older print counterparts. At ONS, such content is often called, appropriately, a 'nugget' (Figure 2). Links on the side of the web page are restricted to taking the user to other documents, not to explore the content on the page itself. More interactive content – interactive charts and maps – tend to be separate to any narrative, meaning that users – especially users with lesser statistical skills – might not be able to use them fully as standalone applications because there is not enough contextual explanation.

Figure 2 – A typical ONS 'nugget'

As a relatively juvenile medium, Internet content is still rapidly evolving. But even now, it is easy to recognise the fundamental differences between fully 'native' web content and traditional print: It is non-linear, characterised by interactive hyperlinking, the ability to jump directly to another document, or simply to another view in the same document. Maturing web browsers and faster Internet connections now mean it is an increasingly visual, interactive medium too.

Hyperlinking can be used to link to different views within a document. In this respect, it is similar to the 'data brushing' exploratory data analysis technique first developed independently by Newton (1978) and McDonald (1982). Data brushing offers users alternative views of the same data, allowing each different display method (e.g. scatter-plot, histogram, line chart) to highlight different characteristics of a dataset. ONS has recently published a graphic designed to explore the ability of web content to display interactive, linked views of information in much the same way. But rather than rely on purely quantitative linked display as in brushing, this graphic explores a range of presentational techniques aimed at delivering an engaging statistical narrative to a wide audience.

'Snapshots From The Census Years'

The product (Figure 3) was conceived as a promotional graphic to encourage interest in the 2011 Census in England and Wales. The plan was to promote the upcoming census as the latest in a rich legacy of snapshots, each chronicling how major events have left an indelible mark on the structure of the UK population. Data for the graphic consisted of all the age/sex national counts from 1911 (the earliest year for which individual age band data are available) though to 2001. This data was graphed using ONS' previously developed animated Flash population pyramid template, and appropriate demographic stories for the period were identified and summarised. Consideration was then given to how this narrative could be more effectively integrated with the visual representation of the data.

Gal (2002) observes "statistical information may be represented in three ways – through text (written or oral), numbers and symbols, and graphical or tabular displays, often in some combination". The Snapshots From The Census Years product contains all of these methods, bound into a fully hyperlinked display. The descriptive text is both textual and oral (as an MP3 audio stream), synchronised to the time slider and next/previous buttons. The text contains hyperlinks to the graph – so clicking on a story (for example, 'Post World War II Baby Boom'), selects the appropriate cohort in the population histogram. Whenever a selection is made (either by clicking on a story or the user interacting with the graph directly), a tabular display of the age/sex data, together with contextual information (the year(s) of birth for the cohort, the proportion of males and females) is displayed. Additionally, appropriate contextual photographs representing the era in question are displayed (sourced from Flickr and The National Archives).

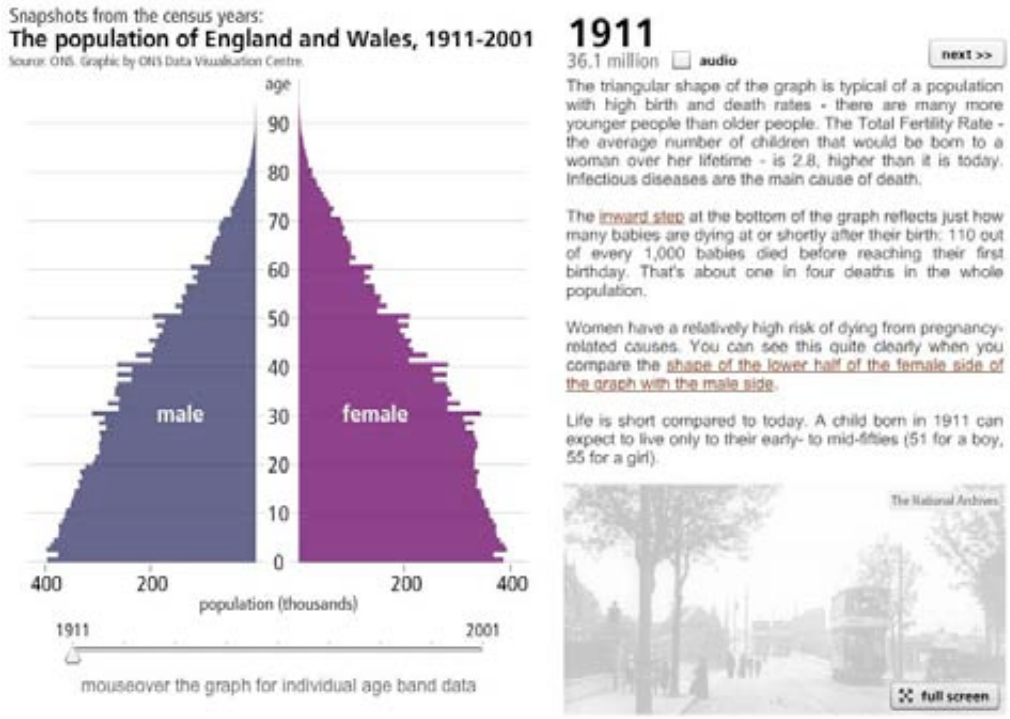
Web Accessibility considerations for integrated content

Web accessibility requirements are a very real issue for developers of rich interactive applications, and particularly for National Statistics Institutes who must adhere to World Wide Web Consortium accessibility standards. It is important to recognise that not all accessibility requirements 'pull' in the same direction. One group of accessibility requirements may appear to contradict the requirements of another. To minimise the potential negative effects of interactive graphics on some groups and maximise the benefits for others, the following accessibility measures were implemented on the Snapshots From The Census Years project:

- A full transcript of the commentary and the underlying data are available separately on the Census website, so that the interactive graphic is not the sole source of the information.

- Full Screen functionality is available, for visually impaired users and for presentation environments
- Audio switched off by default – to avoid interference with screen readers
- Keyboard Accessibility – it is not necessary to use the mouse to achieve basic interaction with the application
- Fonts were embedded into the graphic to ensure that no scrolling, vertical or horizontal, is required on a standard 1024 x 768 resolution display
- A Welsh language version was also produced, complete with embedded Welsh glyphs and Welsh language audio.

Figure 3 – Snapshots From The Census Years – an integrated graphical narrative



Reception and future plans

The graphic received over 7,500 page views in the 2 months between the end of January and the end of March 2011. With very little discrete promotion, these figures represent a promising start, nearly 10 times the mean page views for an ONS nugget over a similar period. The product will be updated to include a 2011 'snapshot' and published on the main ONS site as an early part of 2011 outputs. Given the number of page views and the mean page view duration (a healthy 1 minute and 30 seconds), there would seem to be real potential for this type of integrated content to be used for educational purposes, both in formal educational settings and as part of public engagement. ONS' Data Visualisation Centre intends to explore this approach further in future products, one possible extension being to allow users to add and share hyperlinked stories of their own to the graphics, reflecting the bi-directional flow of content between producers and users typified by the maturing Internet.

REFERENCES

Bolton, Paul (2010) Statistical Literacy: A Review of Initiatives, Royal Statistical Society (October 2010)
Gal, I (2002) Adults' Statistical Literacy: Meanings, Components and Responsibilities, International Statistical Review 70,1,1-51

Leeming, C. (2007) 'Cool Cash' card confusion, Manchester Evening News [online] November 3. Available at http://menmedia.co.uk/manchestereveningnews/news/s/1022757_cool_cash_card_confusion [accessed 11th May 2011]

McDonald, John A. (1982). Interactive Graphics for Data Analysis.

Newton, Carol M. (1978). Graphics: From Alpha to Omega in Data Analysis. In Graphical Representation of Multivariate Data. Academic Press, New York, NY.

Office for National Statistics (2010) News Release - '9.2million UK adults have never used the Internet' [online] 27 August. Available at http://www.statistics.gov.uk/pdfdir/iahinr0810.pdf [accessed April 13th 2011]

Playfair, W. (1801) The Commercial and Political Atlas, T. Burton (London)

Royal Statistical Society GetStats campaign (2010) [online] Available at http://www.rss.org.uk/site/cms/contentChapterView.asp?chapter=25 [accessed May 12th 2011]

Shaughnessy, J. M. (2010) Statistics for all – the flip side of quantitative reasoning. National Council of Teachers of Mathematics President's message, August

United Kingdom Statistics Authority (2010) The Value of Statistical Commentary [online] 23 July. Available at http://www.statisticsauthority.gov.uk/news/the-value-of-statistical-commentary.html [accessed May 14th 2011]