Emerging Trends in Data Visualisation: Implications for Producers of Official Statistics

Alan Smith OBE
Principal Methodologist, Data Visualisation Centre,
Office for National Statistics (ONS), Titchfield, PO15 5RR United Kingdom
alan.smith@ons.gsi.gov.uk

Abstract

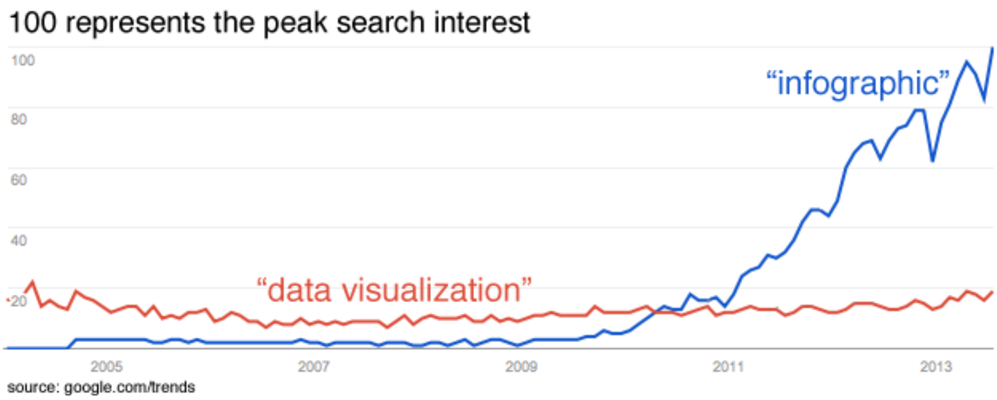
The recent rapid growth of data visualisation on the World Wide Web has been typified by a proliferation of online content types ranging from static infographics through to sophisticated visual analytics applications. Many of these innovations have been driven, with varying degrees of success, by the IT and media sectors rather than the statistical community. Meanwhile, producers of Official Statistics continue to wrestle with the problem of migrating traditional product portfolios from print to web, while increasingly being challenged to reach out to wider user bases. This presentation will showcase some of the opportunities for exploiting the expertise and insight of official statisticians in visual interfaces that promote clarity and methodological best practice while embracing the rich, tactile nature of modern web interfaces. Informational implications, based on user reaction to dynamic visuals, are also discussed using recent examples from ONS.

Key Words: Internet, Graphics, User Engagement, Statistical Literacy

1. Visual Content on the Web

Content on the World Wide Web is increasingly characterised by its emphasis on the *visual*, dominated by multimedia content types ranging from photographs and videos, through to diagrams, charts and graphics, many of which include representations of data in both static and interactive form. There are clear trends in such data-driven graphics, neatly summarised by contrasting recent Google search terms for 'data visualization' and 'infographic' (Figure 1). Infographics clearly represent a boom sector in the visual representation of information on the web, even allowing for the fact that their definition remains ambiguous and includes an enormous variation in the type and quality of content that might fall under that banner. But, regardless of the merits or otherwise of infographics, they are at least clear evidence of the appetite for visual representations of data and other types of information on the Web.

Figure 1 – Trends in Google searches for 'data visualization' and 'infographic'



For producers of Official Statistics (National Statistics Institutes, or NSIs), this shift towards visual content represents both an opportunity and a challenge. For example, graphical methods form a well-documented and popular approach for statistical communication, for example Tufte (1982), Cleveland (1985), Wilkinson (1999) and Few (2004). But producing high quality graphics for the Web – in infographic or interactive form - requires skills which might not be present in great numbers in statistical organisations that are still struggling to adapt traditional print-based product portfolios to the Web. And yet, it seems strange that the huge appetite for statistical data and information in visual form is not directly met by the organisations who produce the data, as graphics would seem to be the ideal vehicle for conveying expertise and insight acquired during the production process.

2. The case for NSI involvement

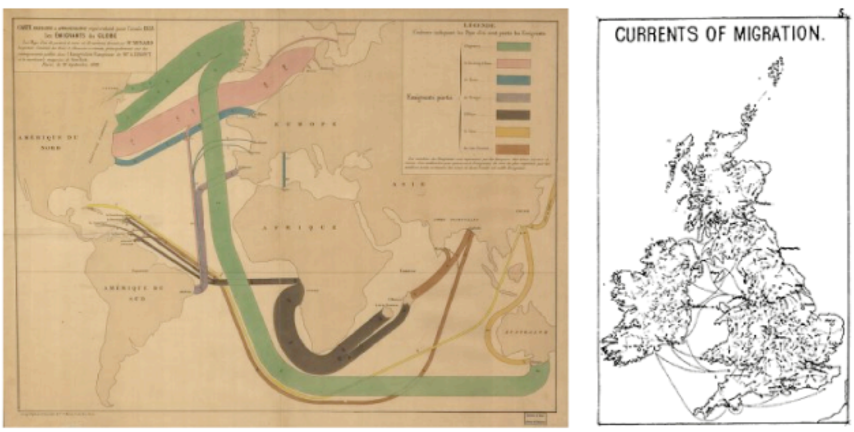
Kirk (2012) sees data visualisation as an increasingly inter-disciplinary activity. Recognising this, the key challenge for NSIs is not so much to appoint a single data visualisation 'expert', but to assemble a team with the correct balance in terms of skills and experience. A crucial factor in getting the balance right is to recognise the unique strengths of an NSI. Many data graphics are produced by media organisations that are traditionally strong in journalistic, design and IT skills – but weak in the area of statistical insight and analysis. By contrast, NSI efforts in the graphical space should emphasise the statistical expertise of the organisation as producers of the data: Producing intelligent, engaging data graphics represents a unique opportunity to differentiate NSI content from other reporting of official data on the Web, adding unique value. The following section presents a brief case study in the development of an interactive visualisation tool that aimed to do just that.

3. Case Study – Visualising Internal Migration in England & Wales

Produced by the UK Office for National Statistics (ONS), the Square Matrix of Local Authority moves is a 348 x 348 grid, containing over 120,000 cells of data with estimates on internal migration moves in England & Wales, based on data from a number of different sources. This data is too voluminous to sit inside one Excel worksheet, and so there is a big task for users to manipulate the data even before considering the information within the dataset. In addition to its complexity, migration data is a high profile output, being used for a number of policy and planning purposes in the local context, making it an ideal candidate for visualisation.

Figure 2 – 19th Century flow maps of migrants

a) Minard, Global Emigrants (1862) a) Ravenstein, Currents of Migration (1885)



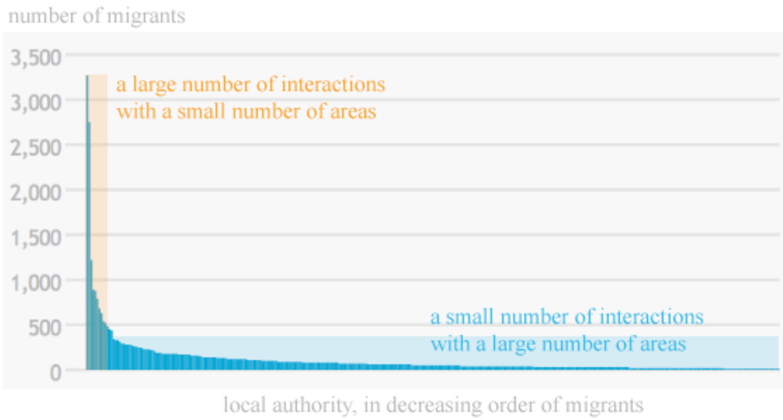
The inter-disciplinary Data Visualisation Centre at ONS assessed a variety of approaches to visualising this dataset that could make an improvement to how users interpreted it. A map seems to be a natural approach, but traditional choropleth statistical maps are not designed to convey movement – each value in the dataset is visually *contained* by a boundary and does not effectively symbolise flows *across* regions. In the literature, static images that present effective images of migration flows include the ground-breaking works of Minard (1862) and Ravenstein (1885), using flow lines to denote movements from one place to another (Figure 2). For visualising movement, this approach offers a more favourable symbology, offering information to the user at a direct perceptual level.

4. Visualising a Matrix

Regardless of symbology, many visualisations of migration data – particularly static ones - inadvertently dilute the information content of the data by failing to recognise the matricular structure of the source data: Very often they depict the relationship between a single place and many other places, whereas the full migration matrix contains information on the relationship of many places with many other places. This is because, in a static image, exposing the full matrix in a single view would be overwhelming to the user, creating a 'spaghetti-like' appearance that would be virtually impossible to interpret. The proposed solution to this problem was to implement interactivity as filter, allowing the user to determine the data to be displayed, a single area at a time: As the user moused over the map, pseudo-animation would allow variations in patterns of flows between multiple areas to be seen.

Even allowing for this, given the size of the matrix, there was still the issue that the sheer number of flows (300+ in some areas) might prevent the user from seeing what was important for any given area. The internal migration dataset under consideration is characterised by a 'long tail' of migrations (Figure 3), a high volume of interactions with only a few areas, but many interactions at a much lower volume with other areas. This is a characteristic completely in keeping with many origin/destination matrices, linked to the 'first law of geography' proposed by Tobler (1970) that "everything is related to everything else, but near things are more related than distant things". What was needed was a way of highlighting for the user significant flows, near or otherwise.

Figure 3 – Birmingham, inward immigration, by total number of migrants



Haggett and Holmes (1977) proposed a method for identifying significant flows in matrices that was assessed for suitability. It was found to be suitable for the characteristics of the internal migration data, while its relative simplicity made it easy to convert into a javascript routine which could potentially allow the user to perform the test in real time when interacting with the map. With the approach to symbology defined, the next step was to find a method of implementing the interactive flow mapping in a web-based environment.

5. Implementing the visualisation in a web browser

There is a clear trend in web technology towards data visualisations in native HTML format, marking a departure from a period in the previous decade that was characterised by an abundant use of the proprietary Adobe Flash technology. The widely-cited reasons for this [e.g. see Jobs (2010)] could be summarised by the shift in the way that web content is consumed, away from traditional desktop computers, towards mobile tablets and smartphones, with different interface paradigms (largely based on touch/multi-touch) and lower energy use.

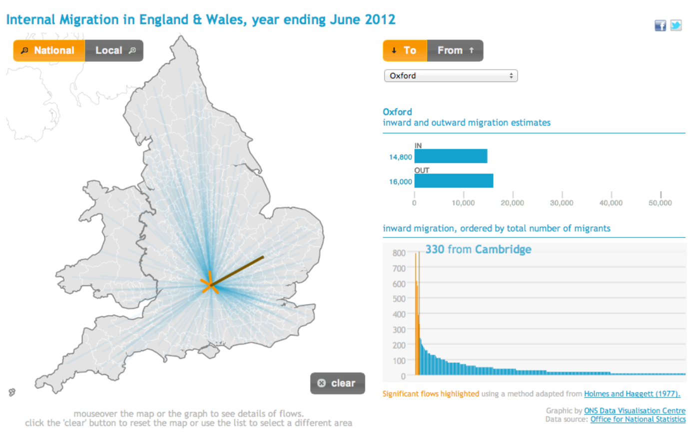
The migration visualisation tool was created using the HTML5 family of technologies (Table 1 provides a summary). The resulting application enabled the entire Square Matrix to be dynamically visualised in flow map and graph form and interrogated by a user on any modern web browser on a PC, tablet or mobile phone. Mousing over/touching an area on the map, or selecting it from a list, displays the flows to/from an area, with 'significant flows' as identified by the Holmes & Haggett algorithm clearly highlighted. Once selected, an area can be interrogated in more detail to find discrete values from the dataset.

Table 1 – the role of the HTML5 family in the migration visualisation

Technology	Description	Role
HTML	HyperText markup Language.	Provides basic document structure and most of the text for the application
SVG	Scalable Vector Graphics.	Provides a drawing API for vector graphics, used for all mapping and interactive graphic elements
CSS	Cascading Style Sheets	Provides the presentation layer – the styling rules for laying out the application
Javascript, jQuery jQueryUI	Scripting language executed by the browser	Provides the interaction layer, dynamically altering the appearance of the graphics based on user interaction.
Canvas	Programmable bitmap graphics	Provides a drawing API for simple graphics where vectors are not needed – used for the total migrants chart.
XML	eXtensible Markup Language, a platform-independent language for describing structured data	Was used for converting the Excel-based data into a structured form suitable for a web browser.

The completed visualisation (Figure 4) features data 'brushing' – that is, linked displays of the same data. In this implementation, it means the flow map and a ranked bar chart of flow values are linked together interactively: Mousing over the map highlights the appropriate flow on the chart and vice versa. Also, the facility for users to switch the direction of flow was implemented, allowing for rapid comparison of inward and outward flows.

Figure 4 – Interactive Flow Map of Internal Migration in England & Wales



As a final touch, 'parameterised URLs' were also introduced, allowing users to share stories with deep links into specific views within the visualisation. – Table 2 provides some contextualised examples:

Table 2 – parameterised URLs (map selection information encoded into URL)

URL	Description
/index.html	The native state of the visualisation with nothing selected (a blank map)
/index.html#00ML,nat,to	Flows to Brighton (area code 00ML), national map view
/index.html#00GA,loc,from	Flows from Herefordshire (area code 00GA), localised map view

6. Release and Reaction

Pre-release versions of the map were shown to media agencies, such as the BBC and The Daily Telegraph. Initial comments were extremely favourable and quite revealing. One journalist commented that they would never have devised the idea of showing 'significant' flows, but would be much more likely to show a 'top 10' or similar simple ranking, which would have reduced the ability of the map to focus on what was meaningful: In some areas, such as Cornwall, there is only one significant flow in and out [Plymouth], thus a simple 'top 10' would give a false impression that positions 2 to 10 were somehow menaningful in the context of the flows.

The map itself was formally published with new migration data in September 2012 and received widespread positive reviews from both journalists and the general public. The Daily Telegraph syndicated the map and used it as a central feature in their reporting of the migration data release for the general public. More specialist users also appeared satisfied. For example, Warwickshire Observatory published a blog article about the tool, calling it "*an excellent tool for visualising and examining migration patterns*". The blog went on to use the tool to identify significant flows within Warwickshire (using the Holmes/Haggett functionality). There was similar enthusiasm on social media, with several news editors providing links and retweets to the map.

This collective reaction endorses the notion that syndication (both through formal press channels and via social media) is a potentially enormous tool for maximising the outreach of both official data – and the insights that official statisticians can help generate with that data.

7. Future development

Looking at the reaction to the map as it was being used via social media proved helpful in understanding possible future directions for the visualisation. For example, in addition to showing in/out flows, many users posted a logical call for significant *net* flows. Similarly, being able to swap over origin/destination matrices within the visualisation (for example, to explore demographic trends within migration) seems a natural extension of the current framework. Another intriguing possibility is to re-use the visualisation with other datasets. A prototype map of Significant European trade flows highlighted a potentially useful direction in this respect.

8. Conclusions

The project described makes it clear to NSIs why visualisation can be a powerful tool for dissemination and communication. And it is a trend that looks set to continue as devices connected to the web become more powerful. The fact that a smartphone such as an iPhone or Android device, can load 120,000+ cells of data and dynamically execute Holmes & Haggett's significant flow algorithm on a subset of flows reflects just how far web browser performance has advanced in recent years. This has profound implications for the future ambitions of interactive graphics – the user can use 'guided interfaces', such as that offered by the migration map, to perform robust statistical tests in real-time.

More generally, if the underlying official data is of value to society, then visualisation represents a great way of engaging with the wider public, without necessarily 'dumbing down'. Our thoughts are that the success of the migration map is in the *way* it helps the user, not by stripping information away, but by presenting the full complexity of the data in a way that users feel confident exploring and interacting with.

References

Cleveland, W.S. (1985) *The Elements of Graphing Data*. Hobart Press, Summit, New Jersey, USA

Few, S. (2004) *Show Me the Numbers* Analytics Press, Oakland, USA

Haggett, P. and Holmes, J. (1977) "*Graph Theory Interpretation of Flow Matrices: A Note on Maximization Procedures for Identifying Significant Links"* Geographic Analysis, Vol.9, 4, 388-399

Jobs, S. (2010) *Thoughts on Flash* retrieved from http://www.apple.com/hotnews/thoughts-on-flash/ 10th April 2013

Kirk, A. (2012) *Data Visualization: a successful design process*. Packt publishing.

Minard, C. (1862) "*Carte figurative et approximative représentant pour l'année 1858 les émigrants du globe, les pays dóu ils partent et ceux oú ils arrivent*" Mr. A. Legoyt et le merchant's magazine de New York ; autog. Regnier et Dourdet.

Ravenstein, E.G. "*The Laws of Migration*." Journal of the Statistical Society of London, Vol. 48, No. 2. (June, 1885), pp. 167–235.

Tobler W. (1970) "*A computer movie simulating urban growth in the Detroit region*". Economic Geography, 46(2): 234-240.

Tufte, E. R. (1982) *The Visual Display of Quantitative Information*, Graphics Press, Connecticut, USA

Wilkinson, L. (1999) *The Grammar of Graphics*, Springer, New York

"Interactive graphic: 230,000 quit London for the regions" retrieved from http://www.telegraph.co.uk/earth/greenpolitics/population/9565594/Interactive-graphic-230000-quit-London-for-the-regions.html 12th April 2013

"New tool for visualising migration patterns" retrieved from http://warksobservatory.wordpress.com/2012/06/29/new-tool-for-visualising-migration-patterns/ 13th April 2013

The interactive flow map of Internal Migration in England & Wales retrieved from http://www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc25/index.html 15th May 2013