# DESIGNING INTERACTIVE DISPLAYS TO PROMOTE EFFECTIVE USE OF EVIDENCE

Sinclair Sutherland, Spencer Hedger, Mark Ireland and Jim Ridgway
Durham University, UK
sinclair.sutherland@durham.ac.uk

*Interactive displays are increasing being used to convey information, and are a significant factor in promoting statistical literacy (and illiteracy). Durham University and the House of Commons Library are collaborating to create data visualisations (DV) which will be accessible to politicians, researchers and journalists. The focus of this paper is a DV designed to be useful in the run-up to the 2015 general election. The aim was to assemble a rich resource from multiple sources, and to make it easy for target groups to manipulate data and draw conclusions. We identify important changes to the DV as it evolved over 13 iterations, and draw conclusions about appropriate design processes and validation.*

INTRODUCTION

Statistics education increasingly makes use of technology, for both distance and face to face instruction. Print media and TV, and social networking sites present statistical data routinely; these are important sources of statistical literacy (and illiteracy). There is an urgent need to foster the community which promotes good design in the presentation of statistical information via interactive displays, and supports a community of designers. Since Playfair's (1786) seminal work on data visualization, there is much that can be built upon (e.g. Tufte, 2001; Rosling, 2010).

A number of sites and commercial packages offer 'off-the-shelf' data visualisations (e.g. *Tableau*, *Many Eyes, Visual.ly*); these are not always successful (the demise of *Swivel* provides an example). A danger in using off-the-shelf products is that, of necessity, they cannot be tailored to provide an exact fit to user needs, and there may be compromises in the exactness of the fit between the visualization the structures in the data to be presented. The alternative approach of creating custom DV for particular data sets and audiences offers the prospect of functionality geared to specific data sets, and designed to fit user characteristics and needs, but this approach incurs all the costs and risks associated with the creation of new software products.

The context of this paper is a collaborative venture between the House of Commons Library (the Library) and Durham University which sets out to create interactive displays that will facilitate communication between the Library and a range of users. The Library is a major source of data to politicians (and to journalists, as a by-product of their publications). One ambition for the collaboration is to support a climate where more and more people can access 'open data' (and choose to do so). A major problem is that 'open data' are often available in a form that cannot be accessed by people with little experience of spreadsheets, or interrogating databases, and so an immediate goal was to access data on behalf of users, and to create displays that convey large amounts of data in ways that are accessible to these key users – here, politicians, their researchers, and journalists. We also want to create a climate where users expect and demand good quality data displays, and so encouraging use was an important design requirement.

Here, we explore the ways in which ideas and designs emerge by tracing the evolution of the Constituency Explorer; we aim to draw some general conclusions about the design process. The Constituency Explorer is a DV created for an audience that is significant in a number of ways: first in terms of their importance for promoting the effective use of evidence, and second in terms of the high standards of data quality and presentation that are demanded.

Work began ahead of the 2015 General Election. We anticipated that there would be considerable demand for data displays relevant to the election. In the UK, the House of Commons is made up from people elected from different areas called constituencies. There is a good deal of data about each constituency, so it was decided to make the presentation of constituency data the focus of the first DV. Data are available on more than 150 variables (on topics such as demography, health, household composition and the economy).

The development team is dispersed across offices in Durham, in London and in Belfast. We used *Trello* as a means of communicating. *Trello* is a platform where team members can post

cards in columns on any aspect an on-going project. In this project, there were columns devoted to: data files; key issues; general and specific 'to do' lists; and to each DV family we were creating. For each DV, there is a card which gives details about the address of the version (all versions are archived), the changes and fixes that were incorporated, and (usually) the reasons for the fixes and changes. This paper is based on the records of the first 13 versions of the Constituency Explorer.

Throughout the design process, there have been a number of technical developments. These include: creating new DVs by loading data via an API, rather than uploading individual csv files; ensuring smooth functioning on iPads; ensuring compatibility with a range of browsers; adding Analytics features; fixing bugs; positioning data panels to make iPad use easier; varying the size of clickable areas; addressing accessibility issues. Although essential to effective design, these developments are of marginal interest to statistics educators, and so are beyond the scope of this paper.

The first version of the Constituency Explorer is shown in Figure 1. It is based on a DV created by the Data Visualisation Centre at the Office for National Statistics, with whom we collaborate (see Ridgway and Smith, 2013).
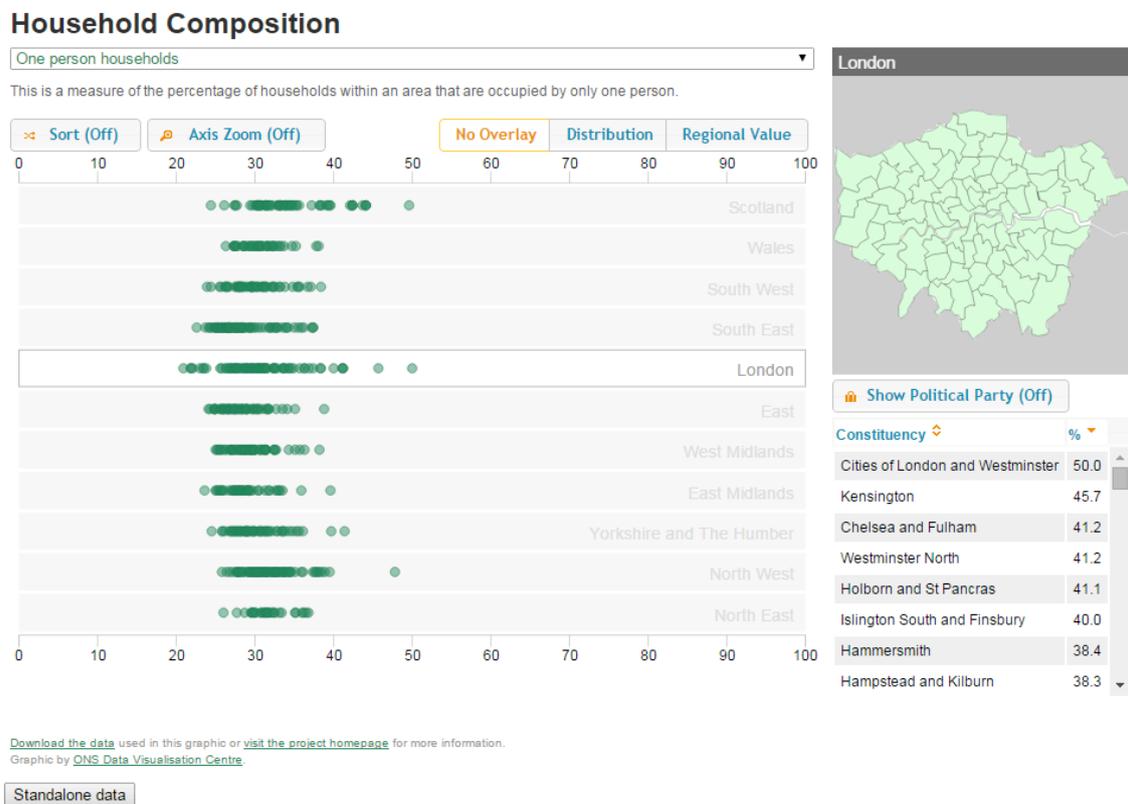


Figure 1: The first version of the Constituency Explorer

Each circle represents a constituency; constituencies are grouped by geographical region. The DV shows the location of each constituency on the chosen variable. Variables can be selected via a pull-down menu. They can be disaggregated into (colour coded) political parties by clicking on a region name, or on the map. It is easy to overlay the mean value on a particular variable for each region, and to superimpose a box plot onto all the constituencies in a particular region. Context is important, and many data visualisations have the default of an expanded scale, meaning that the user has to actively look at the scale to identify the arbitrary range being used in the display. We took the design decision that percentage measures would always open showing the full 0 – 100 range (as in Figure 1) so that the user immediately and automatically forms an impression of the approximate centre of the distribution, and of the spread of values across constituencies. There is a facility to rescale the axis in order to increase the dispersion of the constituencies, and a

facility to reorder the regions from highest to lowest, on the basis of the Regional Value. Constituencies can be sorted by name (in ascending or descending order) and by their score on the variable of interest (again, in ascending or descending order). A click on a circle identifies the constituency via an outline on the map, and by highlighting the name in the constituency list. It also brings up a panel (set below the display) for that constituency with the score on every variable available in this DV, together with the score for the relevant region, and a calculation of the deviation from the Regional Value. Users can download the data as a spreadsheet.

There are many sensitivities associated with releasing a DV which is badged with the House of Commons logo, because of the imperative to be seen to be politically neutral. For example, in what order should regions appear? (answer: in number order from an existing European classification). When a region is disaggregated by constituency, in what order should the parties appear? (answer: put the party with the largest number of seats at the top). What colour dots should you use, bearing in mind issues of accessibility to all users? (answer: ignore accessibility – you must use parties' own-brand colours), and so on. There are political nuances to many phrases, and we defer to our Library colleagues in all such matters – notably creating abbreviated variable names.

ARCHAEOLOGY OF A VISUALISATION
Figure 2 shows Version 13 of the Constituency Explorer.
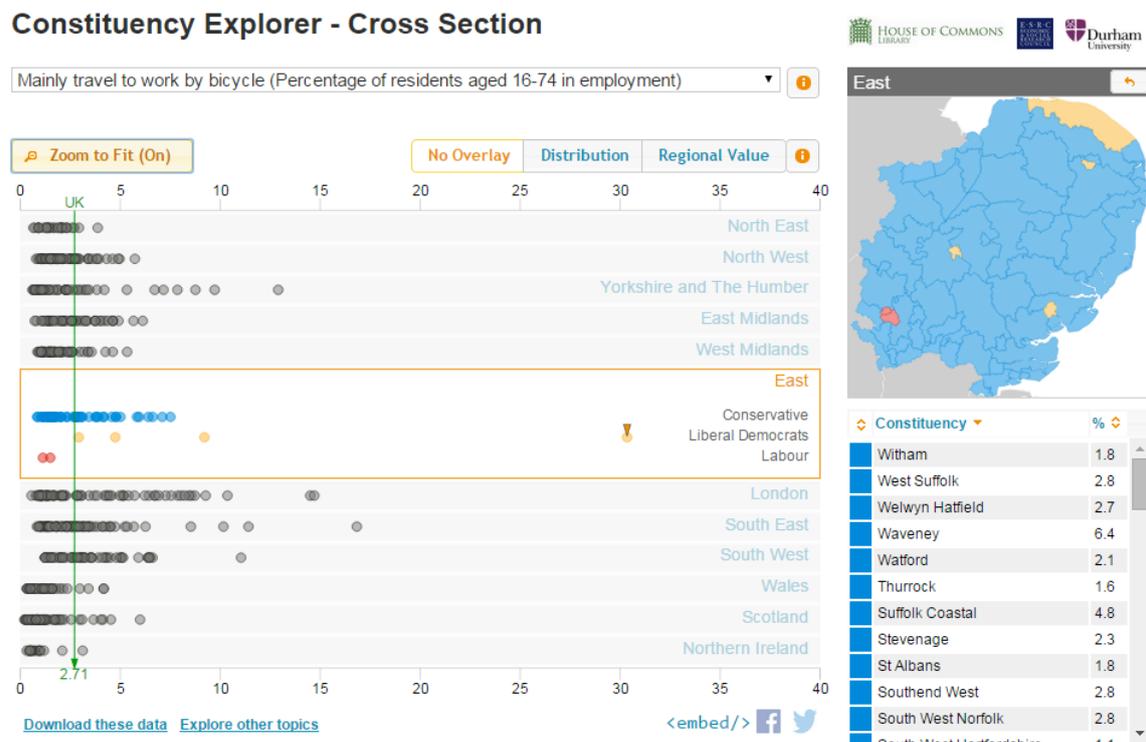


Figure 2: Version 13 of the Constituency Explorer

We describe changes in the design of the Constituency Explorer that were triggered by different groups of people.

*Significant changes from trials with data and computer literate users*
It is obvious that users need to be able to understand what the display is showing, and how information is to be interpreted. A very low hurdle to test comprehensibility is to ensure that sophisticated users can work with the new display, and draw some sensible conclusions. We asked co-workers who had not been involved with the development of this DV to explore it, while we watched.

Initially, users did not realize that the circles were constituencies, and that our belief that driving the display would be intuitively obvious was quite wrong. As a result, we have incorporated a brief on-screen description of the display, and a 'getting started' tutorial.

There were issues that high or low values on variables would lead to circles overwriting the region names (this was often true when the Zoom facility was used). Both Zoom and the location of labels is now done 'intelligently'.

*Significant changes from trials with a key user group*

It is essential to involve members of the target user group throughout the development process. Experiences with the display highlight ergonomic factors and conceptual issues, as well as helping shape the design specification itself.

We ran four focus groups in the House of Commons over a 3 month development period, with a variety of audiences – in particular, with MPs' researchers. As well as identifying some ergonomic issues about the placement of pop-up displays, there was a request for some functional enhancements. A key request was for a facility to 'pin' one or more constituencies. This would allow the user to change the variable of interest, but not have to locate the constituency they were interested in (usually their own!) again, because it could be found easily via its pin. Pinning would also facilitate comparisons between constituencies.

There was a request for UK wide data, as well as regional data, on the pop-up display. Prompted by this, we also added a vertical line to the display, showing the UK average value, as well as the regional values, when users click the 'Regional Value' button.

Users also wanted more and different information about the results of the last election.

We want users to tweet comments on specific displays. Discussions in focus groups emphasized the need to relate a tweet to a specific display (a particular variable within a named DV) rather than to a generic one. The DV now creates tweets with relevant text (DV name followed by variable name) and links to the appropriate page within the named DV.

When conducting user trials, we ask users to bring their own computers. This might seem like folly! However, because of the nature of the target audience, we need to test software on a range of platforms and web browsers. While there were no problems with the Explorer itself, the Quiz (designed to run on smartphones and mobile devices) was associated with problems on some android devices, where dragging a slider actually moved the whole display.

*Significant changes from design team discussions*

Creation of successful DVs is not a passive process characterized by *design specification* to *coding* to *user testing*. The creation of every new version of the DV provokes ideas and discussions, and can lead to changes to the design specification itself (here illustrated by the fact that we are analysing the evolution from Version 1 to Version 13 of the same DV). The design team was a major source of changes.

Initially, we had envisaged a web page where users would choose a topic of interest (say health or ethnic composition), and would exit one DV before entering another. We decided instead to have a facility to move from one DV to another, while retaining the user's key settings (notably the constituencies they pinned in an earlier DV).

We judged the 'sort by Regional Value' to be a distracting feature, and one that disconnects the spatial arrangement of the display from the map; the sort function was disabled.

We took a stronger line on metadata – they are now intimately linked to their data. Metadata can now be accessed directly on screen, and users who want to download data are obliged to download the metadata, too.

Originally, we had planned to offer downloaded data in a spreadsheet format. We have decided to use csv as the format of choice, to avoid any apparent endorsement of a particular commercial product.

In the pop-up panel for each constituency, information is provided about every variable. We have added a facility whereby users can click on any variable name in the table, and the display will change so as to show data for that variable.

An early design decision was to make it easy for users to embed our DV in their own websites. We moved the instructions from the credits to the front page.

The Zoom facility has been modified so that circles do not override Region names; the scale now accommodates count data and currency, as well as percentages; the Sort facility now allows users to sort by political party, and secondary sorting (e.g. rank within party) is supported.

Some data are only available for England and Wales, or for England, Wales and Scotland etc. because of some country autonomy in the census questions. Titles of displays are now adjusted intelligently, depending on the data provided.

At the level of product design, we saw an opportunity (quite late in the design process) to link resources in a way that would seem intuitively obvious to users. A parallel software development was the creation of a Quiz about each constituency. We hoped that the Quiz would be attractive to a broad target audience. We now provide a link from the Quiz to the Explorer, to encourage more user engagement with data.

We also provide links to existing pdfs on the Library website, so that users (notably researchers) can access a single resource with extensive information about any constituency.

GUIDING DESIGN

This analysis of the design process reveals a number of important principles. Most obvious is the need for agile development – rapid prototyping, fast feedback from relevant groups, reflection and revision. Engagement with users from the outset is essential. Feedback provides reality checks (for example, expert computer users who could not make sense of our 'intuitive' displays – provoking the need for more on-screen information, and a help system) and ideas about ways to add functionality (for example, the use of pinning to locate and compare constituencies, and the improvements to the facility to engage with *twitter*). Engagement with target user groups also led to the incorporation of new data, notably on UK averages as well as regional ones, and more detailed information about the 2010 election results. The circumstances of trialling change as the prototype gets to be increasing close to the finished product – so intial trials on a non-representative group proved to be very informative, but trials with a representative group, using their own computers and authentic data was an essential activity, later in the trialling process.

The design team introduced important design changes during the design process. Some of these came about from reflections from a user perspective about the functionality of the whole system (integrating the Quiz and the provision of pdfs with the Explorer DV). There were changes associated with statistical ideas, such as changes to the ways that box-plots are explained, and an increased emphasis on providing metadata – both on-screen and as an integral part of any data download. Data are downloaded as a csv not as a spreadsheet, to avoid any perception of us favouring any particular commercial software provider.

An important design decision was to remove an attractive 'infographics' feature, where the display reordered the regions on the basis of their mean value, whenever a new variable was selected. This was judged to be visually engaging, but a distraction that actually uncoupled the spatial arrangement in the display from that in the map. Design is about deciding what to leave out, as much as what to build in.

CONCLUSION

Despite the plethora of off-the-shelf DV, there is a place for DV that are designed with a specific class of data sets, and specific user groups in mind. Developing such resources is not straightforward. Development is not a simple process of moving from specification to prototype to product. Criteria for good design come from a variety of user perspectives, and failure to address an important perspective can result in overall failure. For example, insensitivity to possible political bias expressed via an inappropriate choice of colours, or a rigid ordering of political parties in a display could lead to reputational damage; autoscaling response variables, or failing to provide metadata could lead to statistical misconceptions and misinterpretations; ignoring ergonomic aspects of the display could lead to poor user experiences; failure to engage users early in the design process could lead to missed opportunities for enhanced functionality.

Appropriate development can only be undertaken by groups with appropriate knowledge of: the contexts of data access and use; detailed knowledge about, and access to, target user groups; software skills; statistical skills; resources; and humility.

Here, the messiness of the design process is exposed by excavating 13 layers of a single DV, created over a three month period. Agile design proved to be essential. User focused design was critical, along with regular user engagement throughout the development process.

Statistics educators have a valuable role to play in the design of interactive displays, and have a great deal to gain, if we can create a 'pull' climate where users (politicians, journalists, citizens) demand high density, high quality interactive data displays as part of the process of debate and decision making.

ACKNOWLEDGEMENT

REFERENCES

Playfair, W. (2005). *The Commercial and Political Atlas and Statistical Breviary.* Cambridge University Press. Originally published in 1786.

Ridgway, J & Smith, A. (2013). Open Data, Official Statistics and Statistics Education – Threats, and Opportunities for Collaboration Keynote Talk: *Proceedings of the first Joint International Association for Statistics Education and the International Association for Official Statistics*, Statistics Education for Progress, Macau.

http://iase-web.org/documents/papers/sat2013/IASE_IAOS_2013_Paper_K3_Ridgway_Smith.pdf

Rosling, H. (2010). *The Joy of Stats.* http://www.gapminder.org/videos/the-joy-of-stats/

Tufte, E.R. (2001). *The Visual Display of Quantitative Information* (2nd edition ed.). Graphics Press.