A01.2

Unlocking information through data visualisation - Some case studies in using SVG

Alan Smith and Alistair Calder, Office for National Statistics, Statistical Framework Division

1 Introduction

The advent of the Internet has not entirely passed national statistical institutes by. Most have become highly adept at converting dull publications into dull websites. While many statistical agencies have started to introduce the use of web specific tabulation packages the aim has been little more than to efficiently serve tables of data - reproducing the old approach in a new, more flexible, format. The potential of the new media for presenting information instead of numbers and to allow users to genuinely interact with the data is only now beginning to be properly considered.

This paper describes our early, though we feel highly successful, experiments in presenting statistics using SVG (Scalable Vector Graphics - a new format and standard for vector graphics). We hope it will be of interest to those responsible for providing data as well as to those who want to use graphics to lever more information out of their numbers.

2 Data Visualisation and Interactivity on the Web

To date effective data graphics have been sadly lacking from most statistical presentation on the web. Most graphics that have appeared have been static reproductions or PDF files of published graphs or maps. For a variety of reasons (largely around the issues in serving data) Flash, for example, has not gained the foothold here it has in other parts of the web. The one area in which a degree of interactivity has been possible has been in the use of GIS which is now widely used in statistical agencies. However, although online GIS has allowed a degree of interactivity in statistical or thematic mapping on the web, this has not been without its problems.

The dominance of raster over vector graphics has meant that most attempts at visualising data on the web, and especially maps, suffer from screen resolution issues. Although vector based GIS solutions are available these have failed to gain wide acceptance or application. Porting GIS to the web has normally either resulted in server-side generation of static bitmaps (via off-the-shelf web servers like ESRI's ArcIMS) or custom-programmed Java applets. It is rarely possible to subtly change the appearance of a graphic without having to reload the entire graphic from a server. One of the web's promises is interactivity - but server-side generation of bitmaps brings with it delays and problems of server-load and bandwidth. The user's experience of interactivity is severely hampered and frustratingly 'clunky'.

Our early experiments with SVG, described below, suggest the potential to move us forward here.

3 SVG: A new and open format for vector graphics on the web

Scalable Vector Graphics (SVG) is an emerging standard for vector graphics on the web. It is XML-based and has become an official W3C recommendation (currently at version 1.1, though 1.2 will be with us soon). Importantly, because it is an open standard, it is not tied to proprietary tools.

Currently, SVG content is displayed in browsers by using a plug-in. A variety of plug-ins are freely available - for example, the Adobe SVG Viewer (currently at version 3.0). Vectors are anti-aliased (smoothed) on screen

which gives them a superior display quality over Java applets. Major advantages of the format are its small file size due to the use of vector information in the first place. SVG content can also be compressed using gzip and decompressed in the browser 'on-the-fly', meaning even smaller file sizes.

It allows the use of cascading stylesheets to separate content from appearance and, perhaps most importantly, interactivity and animation can be introduced through the use of JavaScript in the client. SVG has a layer based approach to drawing objects that is similar to that in GIS systems. As a vector format it is suitable for printing at high quality. SVG content is also accessible to search engines.

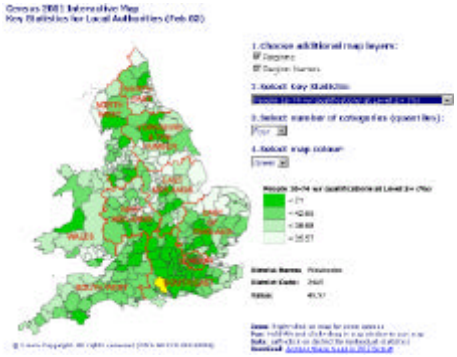
SVG is generating interest within the GIS community not just because it is a well-specified format for rendering high quality maps in a web browser, but because it is not restricted to just mapping and GIS outputs: SVG has the potential to integrate web-mapping with other forms of data visualisation on the web.

Case Study 1: Thematic Mapping of Census Data

For the 2001 Census, the Office for National Statistics (UK) has released thematic maps of Key Statistics for the 376 Local Authorities of England and Wales. Eighteen variables have been mapped, each in a PDF file at approximately 550kb each - a total of around 10Mb of PDF files to download for the complete set.

Using the same data, those maps were then put together in one interactive SVG map that weighed in at under 180kb with the following additional benefits:

- Users can instantly switch the variable being displayed in the map;
- Users can change map colour;
- Users can change the number of classes in the thematic map;
- Exact data for all variables can be displayed in an HTML window by clicking on an area;
- Lossless zooming and panning is very responsive since it is all carried out on the client - there is no need to keep going back to the server to refresh the map.



(all URLs are included in the references below)

Technically the boundaries are displayed by one SVG file and the data is loaded from a different file using JavaScript. This allows the map to be easily updated, for example, by adding new variables. To construct this map, GIS data files were prepared as SVG content using Adobe Illustrator and Avenza's Mapublisher. Scripts were then prepared using a simple text editor, based on the work prototyped by the carto.net project.

Case Study 2: Focus on London 2003

Following on from the initial success of the Census Key Statistics map, a series of SVG maps were produced to accompany the launch of National Statistics' 'Focus on London 2003' publication. Two of these maps have now been launched on the National Statistics website.

The first contains thematic mapping of Oxford University's Index of Multiple Deprivation, allowing users to choose between the various indices (e.g. income, child poverty, education etc) that make up the Multiple Index. The map is at Ward level (760 wards in Greater London) and contains similar functionality to the Census Key Statistics map. Two significant extra features were added:

- The ability to change how the thematic classes were organised was introduced - users can choose between simple quantiles and a skewed set of ranges aimed at identifying the more deprived wards.
- Support for background raster mapping was added (SVG allows raster images to be linked to vector content)

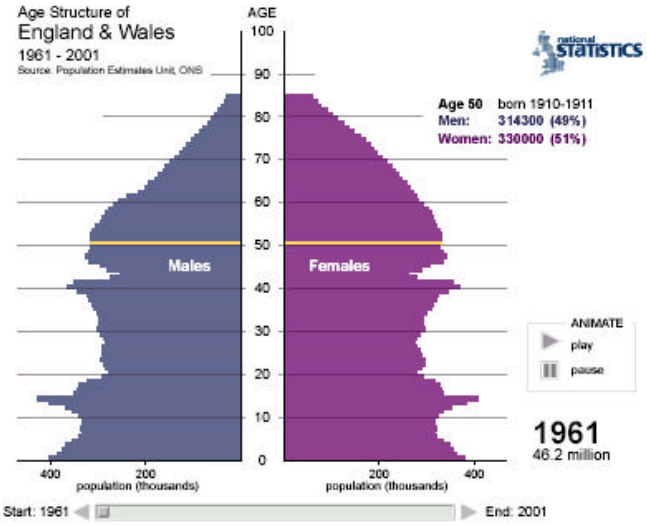
The second map - at the Borough level - was notable mainly for the amount of data it allowed the user to access: 29 rate variables for thematic mapping and 15 count variables fired by clicking on the map. Allowing for background imagery, SVG content was just 35kb.

By the time these maps were produced, SVG implementation had begun to appear in GIS software - Safe Software's Feature Manipulation Engine (FME) was used to generalise ward boundaries and produce SVG content, with text editors being used to script the interactive elements. (Graphic software was not used in the production of this example)

Case Study 3: Data-driven Population Pyramid

Population pyramids have long been popular as a way of visualising demographic patterns in population data. Using SVG, it is possible to produce interactive, data-driven population pyramids, pushing the role of this type of graphic further towards that of a data explorer.

For example, quantifying the figures depicted on a population pyramid has always been difficult. With SVG, a simple 'mouseover' the bars on the pyramid allow instant access to discrete data (stored, in this example as JavaScript arrays within the SVG file itself). It is important to realise that the data are changing the size of each bar on the pyramid: Changing the data changes the appearance of the graphic. Similarly, since SVG content can be animated using JavaScript, it is easy to produce pyramids, which depict change over time. With sufficiently detailed data, the visual effect of this animation can be extremely powerful.



By driving the graphic from the data, it is easy to re-use the pyramid with other datasets: A German version of the pyramid can be viewed on the Federal Statistical Office of Germany's website.

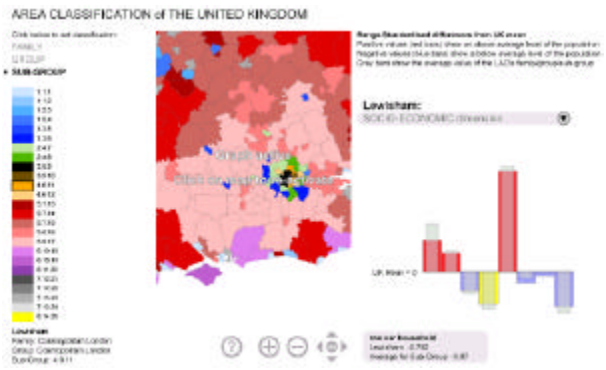
Critically, for ONS, it also proved the concept that SVG could be used for non-map based outputs - with even smaller file-sizes: The Population Pyramid of England & Wales, with 41 years time-series data for 180 age bands (Male/Female, 0-90 years) is 22kb in size. It was constructed using a simple text editor.

Case Study 4: Area Classification of the United Kingdom

Having now produced both maps and charts using SVG, the next logical step was to combine the two - something very difficult to do using a more traditional web toolkit.

The ONS Area Classification is currently being prepared using 2001 Census data to classify UK Local Authorities into a hierarchy of Families, Groups and Sub-Groups. The classification is being constructed from a multivariate analysis of 42 variables across 6 dimensions (demographic, household composition, housing, socio-economic, employment and industry sector).

SVG content has been prepared that allows the users to browse and query an interactive map - and use the map to drill down and explore the multivariates in a graph.



The early indications are that this is a really exciting area for further work - the ease with which different graphics can be constructed and integrated, working live from the same data is rich with potential.

4. SVG - A Summary

- A file format - not tied to proprietary tools;
- Small file sizes (often just a few Kb) - perfect for the Internet;
- Scalable;
- Fully scriptable for interactivity - but at the client end;
- Easy to build templates & design properly to standards;
- Stylesheets provide the same graphic to multiple media;
- An emerging open standard;
- Easy to drive direct from separately held data;
- Equally capable of handling vector and raster data.

5. Moving forward

SVG is still relatively new. This brings with it both the prospect of exciting new outputs and uncertainty about the format's long term future. Because the web browser market is in a state of flux at the moment, it looks increasingly likely that SVG content will continue to rely on the use of a plug-in in the short-term, which is a shame.

However, the benefits of being able to deliver compact, interactive, 'XML-ised' graphics direct from databases are indisputable. In terms of satisfying users, initial feedback to our demonstrators has been extremely favourable and for that reason alone, more work will be carried out looking at extending the use of SVG for data driven graphics in ONS. We are now focussing on options for the automatic creation of SVG files from large databases and considering how it might be applied to serving different types of spatial data.

That said we are fully aware that SVG is not a panacea and is not the only solution possible here. Our real interest is in the concept of imaginative and well designed interactivity as a lever into information. Our focus will be on this and on pressing for this type of approach to be more widely adopted within ONS while remaining responsive to technical changes.

6. Visualisation as an Exploratory Tool

By giving the user the ability to genuinely interact with the data online we are hoping for wider use and interest in our data. See the Appendix for some of the feedback that suggests we are heading in the right direction.

We do not, however, see this type of graphic solely as a presentation tool.

Adding interactivity to the population pyramid moves it from being a simple presentation of the data to a genuine exploratory analysis tool. This fact, combined with the ability of SVG to be directly driven by data, suggests that there is potential to use this approach much more widely for analysis both inside and outside of ONS. We are currently considering the potential of this type of graphic as an interactive exploratory, analysis and QA tool - perhaps providing SVG templates to which ONS business areas can apply their own data.

It is important to note that the normal caveats must apply here. The easy integration with real graphic tools such as Adobe Illustrator and the ability to use stylesheets and templates makes producing professional results and maintaining standards more straightforward. However the technology does nothing to ensure the statistical validity of the results. The same care that needs to be taken with any statistical graphic needs to be applied and in some respects there is even more danger here.

Adding polish and animation does nothing to stop a graphic being statistically misleading or just plain wrong. We need to guard against complacency here as well as the temptation to add complexity to simple stories.

Fresh challenges in this area come from the fact that the user is going to be able to alter the appearance of the graphic. This brings with it extra pressure to think about what is being presented and either limit what is possible of provide clear guidance as appropriate.

7. Conclusions

We have been astonished both by the ease of use, flexibility and potential power of SVG as a format and by the strength of positive user feedback on this work to date.

There are still some outstanding questions and we intend to continue and expand our work in this area but the key messages to date would seem to be:

- The use of interactive graphics - and the effective combination of maps with other types of graphics - can be a extremely powerful in levering information out of complex data;
- Although we do not see SVG as a complete solution we believe it has enormous potential in this area;
- As ever, there is a critical need for standards and real care in the way such graphics are designed and used.

References

http://www.w3.org/Graphics/SVG/Overview.htm8

http://www.adobe.com/svg/viewer/install/main.html

http://www.statistics.gov.uk/census2001/censusmaps/index_new.html

http://www.statistics.gov.uk/london/interactive_maps/default.asp

http://www.statistics.gov.uk/populationestimates/svg_pyramid/default.htm

http://www.destatis.de/basis/e/bevoe/bev_svg2.htm

http://www.carto.net

Comments or questions on any aspect of this paper are strongly welcomed.

We will shortly be preparing a technical paper and set of examples to guide others interested in investigating SVG further.

For further information contact svg@ons.gov.uk

Appendix A - Feedback

Feedback on our published examples (see the URLs above) is presented below. These comments are not provided to show off or to cheer ourselves up - they are intended to demonstrate the strength of interest and demand for this type of material that this work has already revealed.

User feedback from svg@ons.gov.uk :

"....it was really good - very quick and easy to use. More data available this way would be great"

"I think that this is an excellent tool"

"The Census 2001 interactive maps are very easy to use"

".... an enjoyable half-hour browsing the data interactively.. how useful this might be for children in investigating patterns"

".... extremely well designed and effectively implemented"

"What an excellent application"

"You have re-defined the word 'interactive'"

"the population pyramid is most didactic - bon courage"

"....very impressed"

"These maps are great"

"I am in awe"

"Keep up the good work!"

"The pyramid is excellent"

"Just discovered the interactive population pyramid and am most impressed"

"I was very impressed"

"Very fine SVG"

"This age pyramid is very interesting! Using animated SVG to show the

evolution of age's population is really didactic. Please continue to develop charts and graphics using SVG to provide more comprehensives statistics demonstrations. Bon courage.

"Just wonderful!"

In addition publication of the population pyramid has already elicited a strong interest from a range of other international statistical agencies. In addition to the UK and Germany pyramids which have already been published work is currently underway to produce customised versions of the same graphic for Northern Ireland, Canada and the USA. These can be easily created by changing the datasets referenced - a clear sign of the transferability of the SVG approach.

Alan Smith and Alistair Calder, Office for National Statistics, Statistical Framework Division, Segensworth Rd, PO15 5RR, Titchfield, United Kingdom
alan.smith@ons.gov.uk, alistair.calder@ons.gov.uk

with thanks to Michael Neutze
Federal Statistical Office of Germany