

Point data in mashups: moving away from pushpins in maps

Aidan Slingsby, Jason Dykes, Jo Wood
giCentre, City University London

Department of Information Science, City University London,
Northampton Square, London, EC1V 0HB

http://gicentre.org/point_mashups/

Cluttered pushpin maps are often the result of mashups based on large point datasets, from which it is frequently difficult to detect spatial patterns. We demonstrate how server-generated density estimation surfaces, chi surfaces (expectation surfaces) and ‘tag maps’ may be more appropriate means to explore point datasets. We use a gazetteer of British placenames (~9000) – classified by generic form and linguistic origin¹ – and Google Earth to illustrate this. The surfaces and tag maps are generated on a web server, on-demand and at a resolution appropriate for Google Earth’s current zoom level. This allows patterns and relationships to be explored at different spatial scales.

Reasons why such spatial summarisation techniques are not more widely used include a lack of knowledge as to how to handle spatial data, lack of software tools to summarise the data on-demand and the high computation time that is often associated with this. Our mashup uses two prototype Java applications that run on our web server; one that retrieves and returns point data from a database and one that uses this output, generates the surfaces² and returns a KMZ file. These applications are accessed through standard HTTP requests (in which parameters including Google Earth’s BBOX parameter are encoded) that return KMZ files. The worst-case computation time (all 9000 placenames, all surfaces and all tagmaps) is about 15 seconds, but this time is reduced significantly when fewer outputs are requested (e.g. only density surfaces) and when the query is geographically constrained. Code optimisation, precomputation and caching are strategies that will improve the computation time.

Figures 1 and 2 show the strong spatial patterns in placenames’ linguistic origins. In figure 1, density surfaces of absolute counts are used with the same colour scaling applied to all six surfaces. It is clear that Old English placenames dominate in absolute numbers. In figure 2, we use a simple statistical surface (based on the chi statistic) showing the deviation from a null hypothesis that the placename distribution is spatially invariant across the UK for each language group. This reveals that the Celtic languages have much higher local prominence than the other language groups, a pattern that is not otherwise detectable.

¹ *Note:* this dataset is based on data from <http://geonames.org/> and is classified by hand using http://en.wikipedia.org/wiki/List_of_generic_forms_in_British_place_names. Only those places names which could be classified with a generic form are present.

² The density estimation surface computation is based on the (free) LandSerf API (<http://www.landserf.org/>).

See more details at http://gicentre.org/point_mashups/

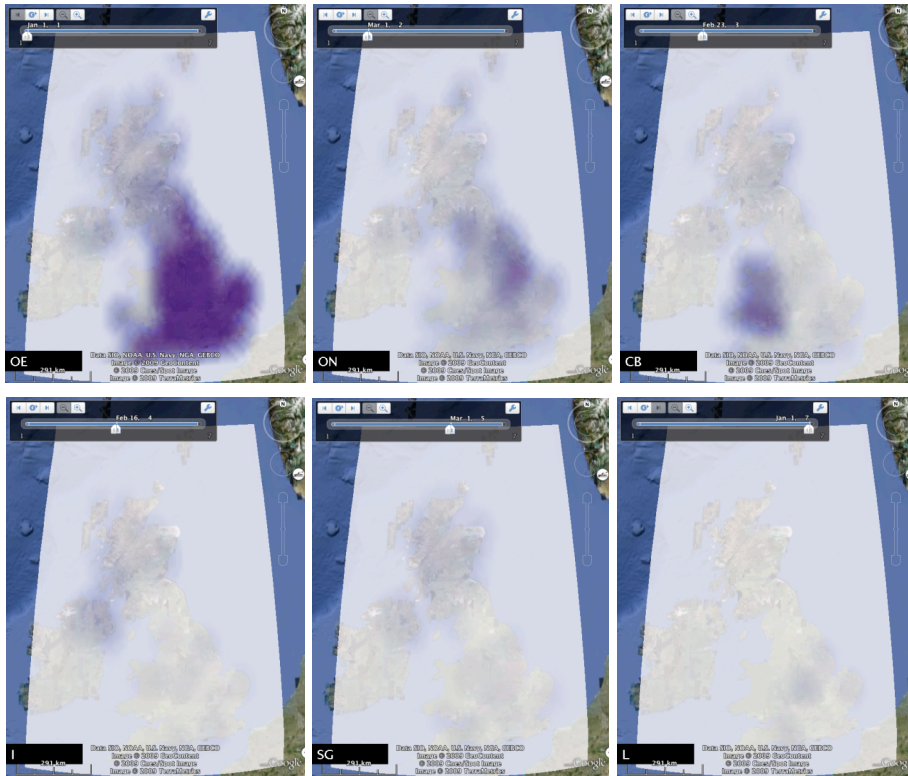


Figure 1: Density surfaces of settlement names with Old English, Old Norse, Brythonic, Irish, Scottish Gaelic and Latin origin, from left to right in order of decreasing maximum local density, with identical colour scaling³.

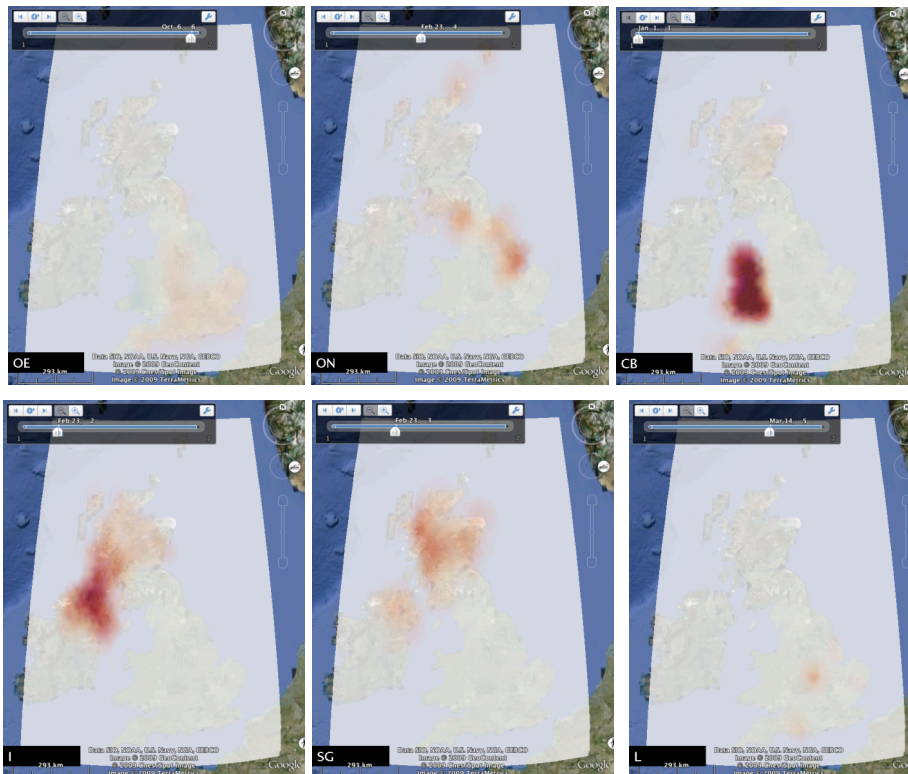


Figure 2: Chi expectation surfaces where red indicates a higher than 'expected' local concentration of settlements with the specific linguistic origin. Note that a symmetrical diverging colour scale is used and that all negative values are too low to be strongly coloured.

³ Colour schemes from <http://www.colorbrewer.org/>

See more details at http://gicentre.org/point_mashups/

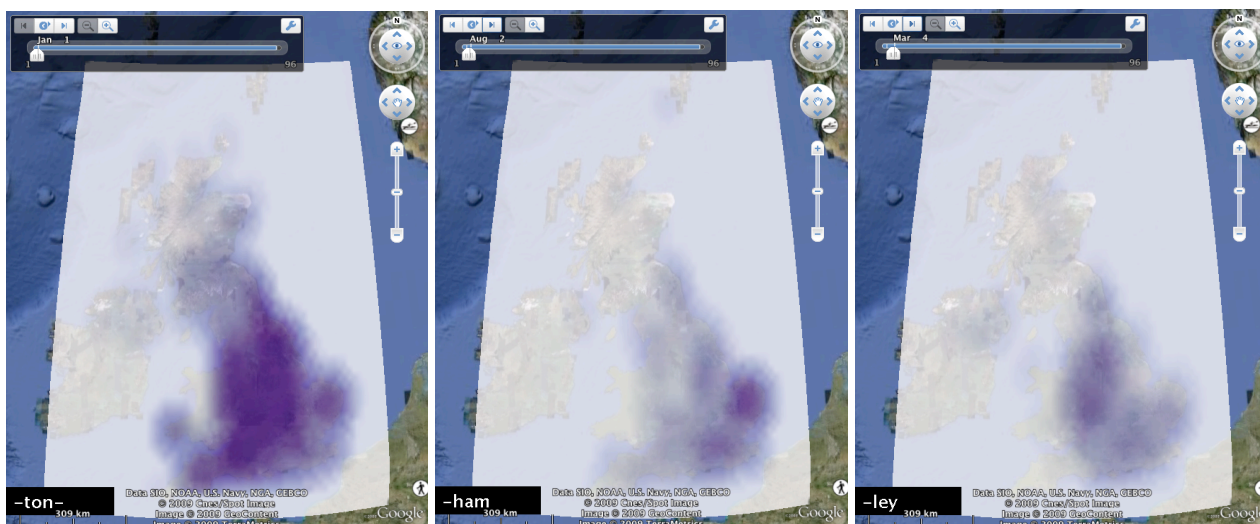


Figure 3: Density surfaces of the top three placename forms when a common colour scheme is applied to the three surfaces.

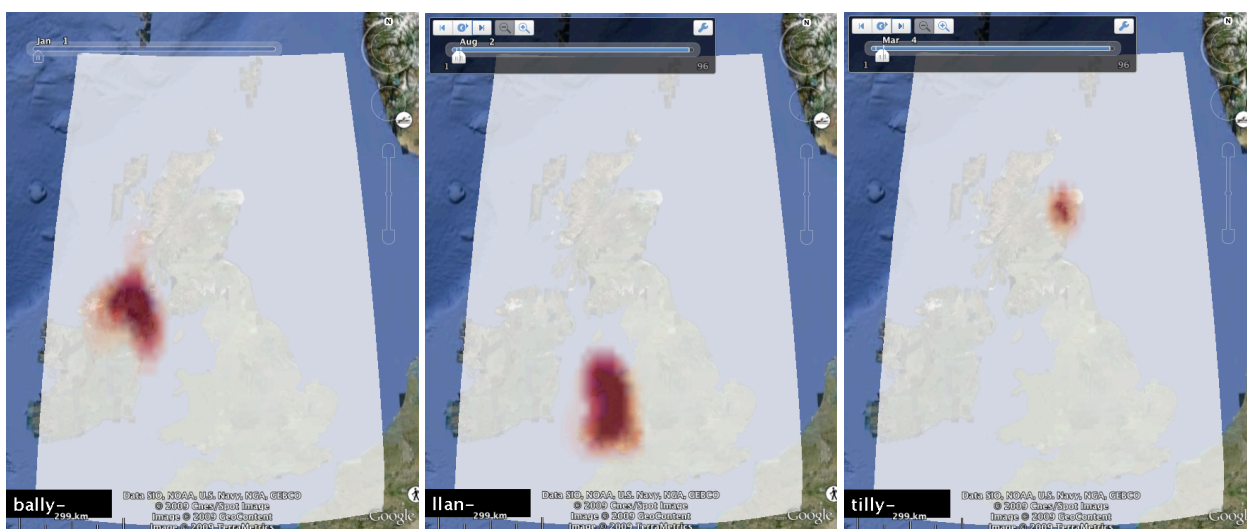


Figure 4: Chi expectation surfaces of the three placename forms with the highest maximum local deviation from expectation, using a common diverging colour scheme for all three surfaces.

Figures 3 and 4 show density and chi surfaces ordered by maximum value (the top three are shown – note these top three are different). These are ordered using the Google Earth timeline slider, which along with the Places panel, gives access to any individual surface. In all cases, the surfaces and tag maps are based on 100x100 rasters for the current field of view, with a 10x10 kernel with an inverse-distanced weighted function (Cressman).

An obvious drawback of these surfaces is that only one variable can be shown at once. In figure 5, we show multiple placename forms concurrently using a tag map – a spatial version of a tag cloud – where words are sized by their local prominence. Prominence can be measured using absolute counts or the chi value – note that these give very different spatial patterns. Words are placed at local maxima in the density and chi surfaces, and sized by the value. In figure 6, the tag maps have been constrained to the Google Earth viewing area based on density surfaces at a larger scale, again based on 100x100 rasters for the area shown.

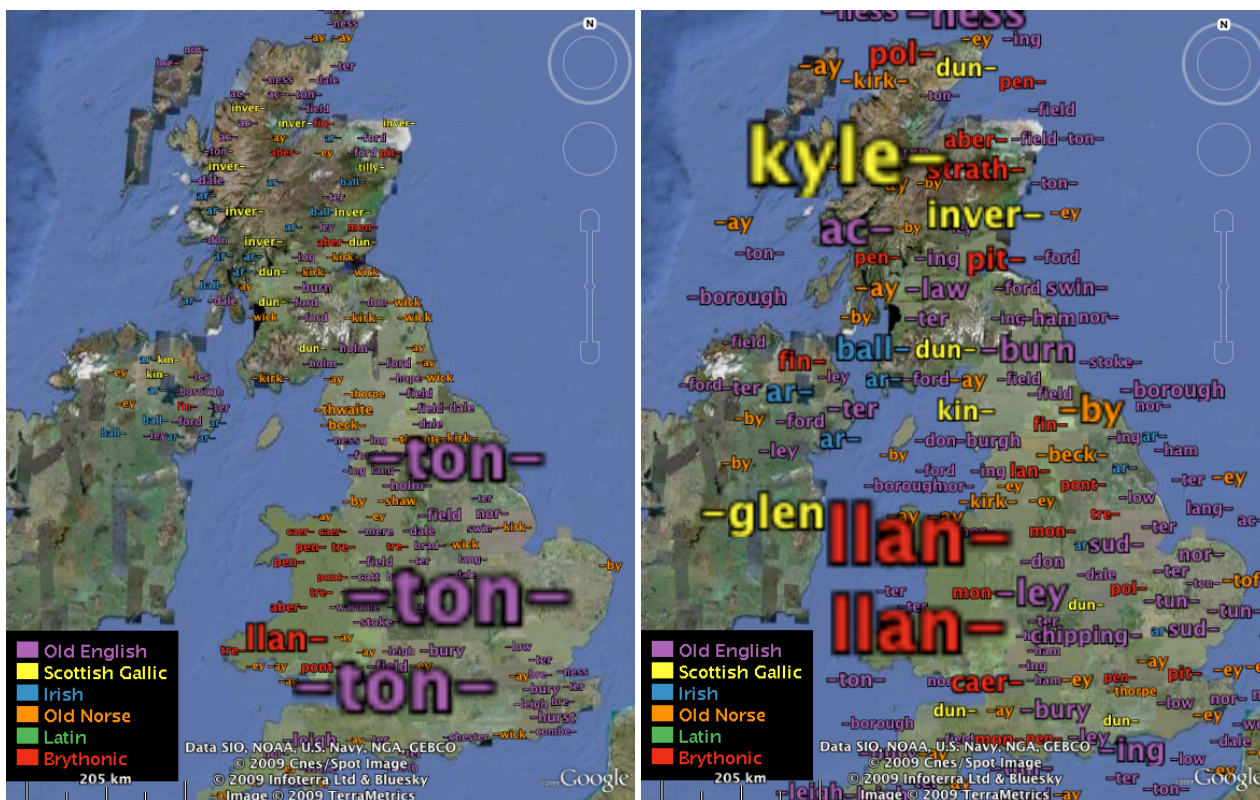


Figure 5: Tag maps of placement forms. Left: words are size by their local (absolute) density. Right: words are sized by their local deviation from expected.

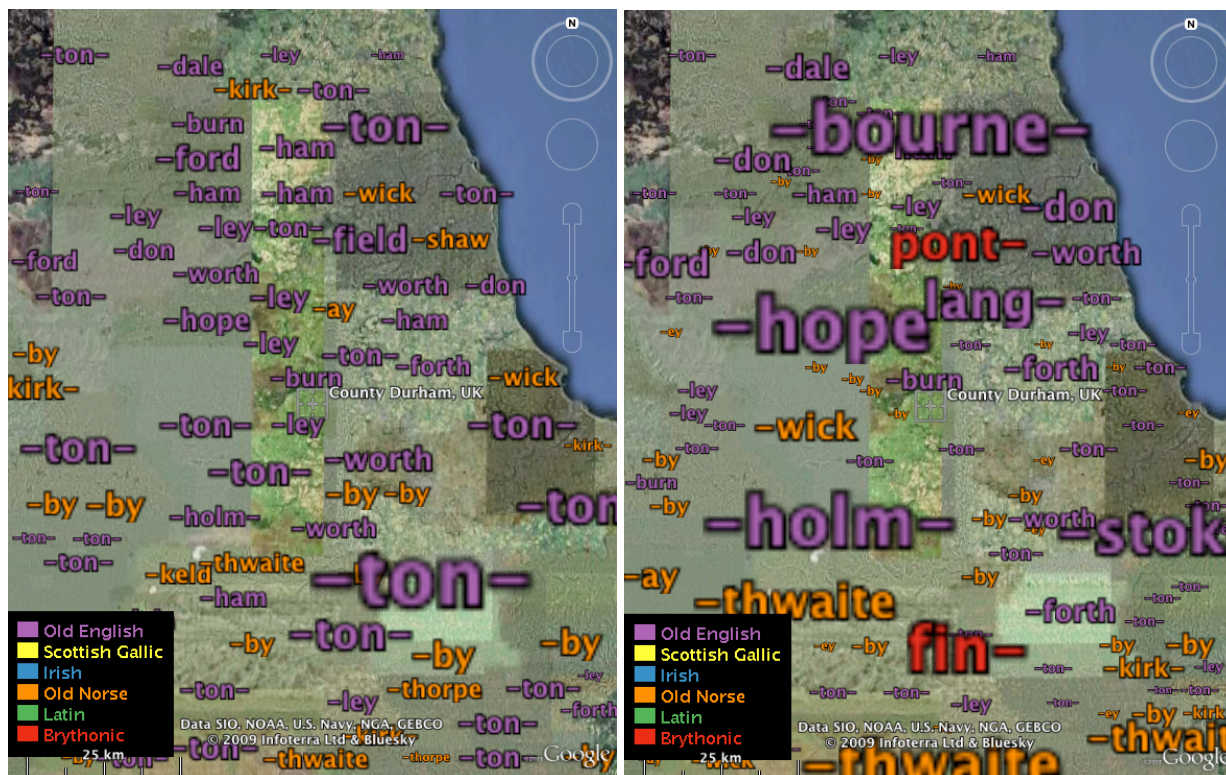


Figure 6: Tag maps as above, but computed for the North East, centred on Durham.

Here, we illustrate how mashups based on point data may benefit from using appropriate scale-dependent summarisation techniques. Our prototype shows one way that this might be implemented, perhaps as a web-service for use in other mashups.

See more details at http://gicentre.org/point_mashups/

References:

Slingsby, A. Dykes, J., Wood, J. and Clarke, K. 2007. Interactive Tag Maps and Tag Clouds for the Multiscale Exploration of Large Spatio-temporal Datasets. *In Information Visualization, 2007. IV '07. 11th International Conference.* pp. 497-504.

Slingsby, A. Dykes, J., Wood, J. and Clarke, K. 2007. Mashup cartography: cartographic issues of using Google Earth for tag maps. *In Commission on Maps and the Internet. Warsaw, Poland.*

Wood, J., Dykes, J., Slingsby, A. and Clarke, K. 2007. Interactive Visual Exploration of a Large Spatio-temporal Dataset: Reflections on a Geovisualization Mashup. *IEEE Transactions on Visualization and Computer Graphics, 13(6), 1176-1183.*