

# DESIGNING AN IRISH SURNAMES MAP USING GIS

Dr Kenneth Field and Dr Linda Beale

## Abstract

*This technical paper describes the design and production considerations for a map illustrating the distribution of Irish surnames. The large-format A0 print document uses classic Irish symbology and themes to create a visually striking representation of historic and numeric data relating to Irish surnames from the 1890 Irish census. The map was designed and produced entirely using Esri® ArcGIS®, as an attempt to illustrate the flexibility and quality afforded by modern geographical information systems (GIS) for high quality cartographic design and production without recourse to alternative software.*

## Keywords

*Genealogy, Ireland, thematic, typography*

## 1 Background

The purpose of the map is to provide a visual representation of the spatial and quantitative distribution of birth data from the 1890 Irish census of population. The birth data identify the surname of every child born, by county. The surnames on the map have either historic or numeric importance to the counties of Ireland and as such the map provides an illustration of the associated geographies. Indigenous Irish names beginning with O', Mac, Mc, De, Le, and others indicate historic significance and plotting these on a map begins to reveal geographical patterns of the distribution of people with similar names. The 1890 census contained over 2,500 surnames throughout the counties of Ireland and in addition to showing the spatial distribution of names, the data allow the numeric significance of the births by county to be mapped.

A map illustrating the spatial pattern of the distribution and quantity of surnames could be used as a mechanism to support research which is not otherwise visible through tabular exploration. By presenting the data in this way, patterns of dominance of certain surnames in parts of Ireland become visible. The almost universal spread of some surnames across all counties can also be explored. As a means of investigating this aspect of genealogical history, such a map provides a fascinating insight into an otherwise simple dataset and reveals far more than the list of surnames alone could reveal. A second purpose of making the map was to demonstrate the capabilities of Esri ArcGIS as a cartographic design and production environment. Such a map is capable of being produced using a range of desktop illustration software, but the approach here was to demonstrate database-driven cartography and to use various cartographic tools solely within ArcGIS to create the desired map. It was created as a means to explore the range of tools now available in

ArcGIS for cartographic design and production of high quality products. Whilst many of the techniques applied are not beyond any skilled GIS professional or practitioner, many users are still reluctant to move beyond the familiarity of desktop illustration software.

## 2 Design Considerations

The design of the map was guided by two principles. Firstly, it had to present the spatial and quantitative information relating to the surname data. Secondly, because of the large number of surnames to place, the printed map would be large format. The overall design philosophy was to create a visually-engaging and eye-catching product suitable for marketing purposes (e.g. for tourism by capitalising on Irish heritage) as well as to illustrate the census data meaningfully as a research tool.

The database of surnames was derived from the 1890 census of Ireland at county level (the finest available geographical descriptor). The database itself consists of three columns: surname, county, and birth count. Consequently, for the c.2,500 rows of data there were three pieces of information that needed cartographic representation. Census data are normally illustrated using traditional thematic mapping techniques such as choropleth or proportional symbol maps. A choropleth is used to indicate the relative magnitude of a value of data, by area, to allow comparison of differences across the map (e.g. percentage unemployment). A proportional symbol map places (often geometric) symbols to indicate the location of phenomena with their quantity being represented by changes to the size of the symbols. The pattern of differences in quantity across the map is seen through the range of differently-sized symbols (e.g. total number of people). Both of these techniques are normally applied to single variables though it is possible to create bi- and multi-variate choropleth maps and also to assign more

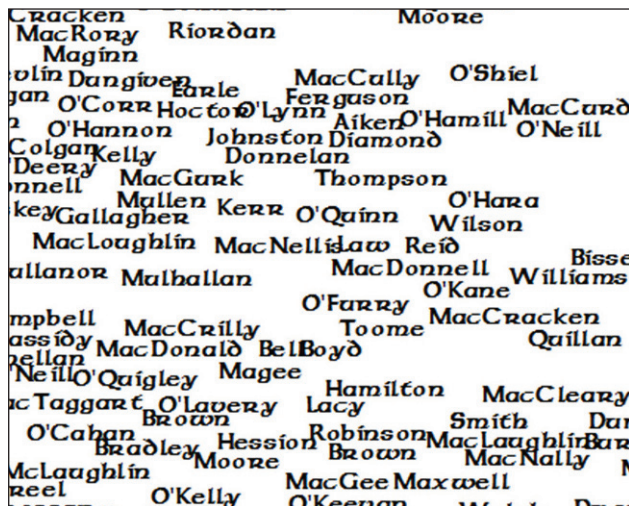


Figure 1a Labels symbolized in the same font size

than one variable to a proportional symbol by, for instance, changing colour, shade, or by using a pie chart. However, neither of these standard methods for representing census data is suitable for the surnames dataset without substantial loss of detail and the resulting map would be somewhat limited and visually uninteresting. Crucially, we required a way of illustrating c.2,500 individual pieces of information (i.e. individual symbols) as well as some way of indicating the magnitude of birth counts. Whilst there are only two variables being mapped, this equates to over 5,000 pieces of information to be symbolized effectively to ensure legibility and visibility.

The design of the map was to be highly artistic as a poster. For the map components to hang together visually, a Celtic theme and connotations with the ‘Emerald Isle’ would add to the impact of the map as a product that would inspire people to view, interact with, and study; to evoke feelings of longing, recollection, and familiarity. These would serve to give the map a strong identity and one which people would ‘warm to’. To achieve this, a design principle was developed around a typical ‘Irish’ theme that played on familiar Irish imagery and styles that would be easily identifiable and which would create an immediate bond of attraction for the user. Ultimately, if a map were to be successful, it should attract a map reader and encourage them to spend time exploring what it has to offer, and the design was chosen to achieve this whilst properly reflecting the map theme.

### 2.1 Surname Label Symbolology

The solution adopted to represent the surname data is to modify the principles for proportional symbol mapping. Instead of using a geometric shape to locate an instance of a surname (which would then require labelling), the surname itself is used as a literal locative symbol. By removing a geometric symbol, the name is itself locative and the removal of any traditional geometric symbol or anchor for the label reduces map clutter substantially. The birth count data are symbolized by varying the size of the



Figure 1b Labels re-sized by birth count

surname label in the same way as symbols on a proportional symbol map. In keeping with the overall design philosophy, a suitably Celtic styled typeface is employed to develop the main map content in a graphically-engaging way.

Clearly, there are some drawbacks with such an approach and the main one here is that not all surnames are equal in size when symbolized at the same font size (Figure 1a). Consequently, if all names are labelled equally, then by definition, a longer surname would be more prominent in the visual hierarchy than a much smaller one. However, this is a relatively small consequence given the amount of information display that the approach otherwise allows. It is also hardly noticeable as a problem of visual recognition when the labels are re-sized according to birth counts. This process clearly illustrates those surnames with numeric importance as being much larger and creates a stable visual hierarchy based on the quantitative data being mapped. In Figure 1b, ‘Thompson’ is clearly a more important surname numerically and it is largely irrelevant that ‘Lacy’ and ‘MacDonnell’ differ in relative size simply by virtue of the length of the surname.

### 2.2 Surname Label Scaling and Placement

The key to an effective proportional symbol map display is to ensure a good range of symbol sizes so that areas with low data density do not appear too sparse, areas of high data density are not overcrowded with multiple overlapping symbols, and that the range of symbol sizes gives good differentiation between high and low magnitudes. This problem is often compounded by the fact that large data values occur most frequently in small enumeration areas (e.g. urban areas) and vice versa, so achieving a good balance between commonly conflicting visual issues is a challenge.

The ‘county of birth’ attribute of each surname is used to join with the centroid of the county boundary spatial data and the x and y coordinate geometry used to

create a point feature class of surnames. An annotation feature class is created for the surname labels that is then converted to cartographic representation symbology to provide a richer set of design tools. The 'birth count' data attribute value is then scaled and subsequently used to vary the font size employed to represent the label. At this stage, labels are still over-lapping because they are still located at the county centroid.

The disperse markers algorithm in ArcGIS is used to randomly position labels to spread overlapping points away from their coincident *x, y* coordinates. The algorithm finds representation markers that overlap or are too close to one another and spreads them apart based on a minimum spacing and dispersal pattern, in this case randomly. A minimum rectangular envelope around each label is defined and used to identify if any two marker envelopes are within the minimum spacing of each other or overlap. The outcome of this process gives an initial indication of the relative sizes of symbols and dispersal. An iterative process of re-scaling the birth count attribute data (to refine the font scaling between large and small data values) and changing the minimum distance variable in the disperse markers tool results in a map of surnames that meets the general need to ensure good coverage in sparse data areas, limited overlaps in dense data areas, and a good variation in size between the smallest and largest sized font. As with most proportional symbol maps, the outcome is optimal rather than perfect, since the area in Dublin County is both the smallest and contains the largest proportion of high birth counts for the largest number of different surnames. This area requires some further refinement since making the symbols small enough so they all fitted within the county boundary renders the rest of the map visually sparse.

The disperse markers algorithm does not allow for constraining dispersion to shapes or other containing areas yet, clearly, the requirement was to locate all surname labels from one county within the county boundary. An additional constraint was that we did not want names to be placed across water bodies since the connotation of being able to identify a name with a 'place' would be impaired. Consequently, the water bodies layer is used as an erase feature with the county boundary input layer to create a county boundary which can be used as a constraining layer for further refinement of the label placement. Surname labels are then rendered using the ArcGIS Maplex labelling engine and placement is constrained to the county polygons (minus water bodies) layer with conflict resolution settings applied to ensure good distribution, no overlaps, and to accommodate the pre-set label sizes. The process results in an excellent distribution of labels for all areas except Dublin County because of the small area/large data density and value issue. The usual strategy of allowing proportional symbols to overlap by modifying the appearance of underlying symbols to accommodate others (e.g. using transparency, no fill or cut-out symbols) is inappropriate here because of the need to preserve the label intact. As adapting the label using traditional approaches



**Figure 2** Label symbol scaling, proportional to 1890 census birth counts

would interfere with its literal meaning, overlaps have to be manually removed in an Editor session to permit labels to bleed across the coastline. This is an acceptable compromise to ensure the overall balance in labels across the whole map page.

The final manual movement of labels using Editor provides cartographic finishing to the placement of labels to optimize the appearance of the distribution. These manual location changes are stored as geometry overrides in the representation. Overall, approximately 90 per cent of the process of the final placement is achieved using the automated tools in ArcGIS which are now sufficiently robust and powerful to be used for complex label placement and conflict resolution.

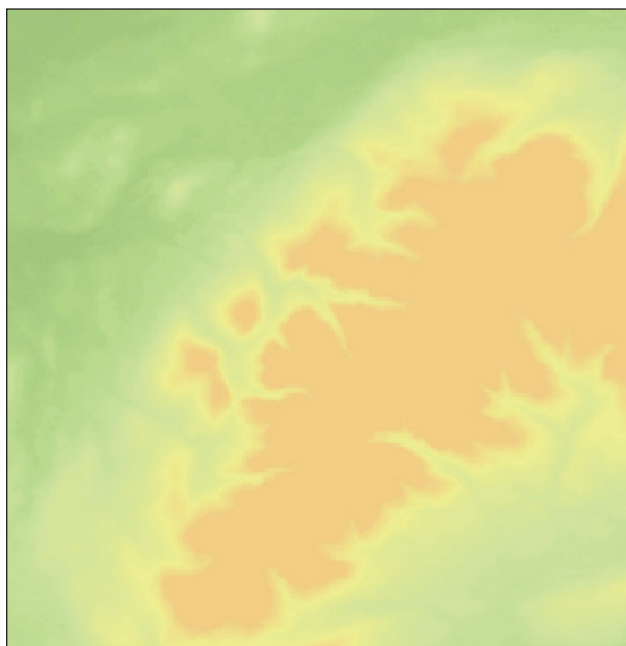
The resultant sizing and placement of labels shows surnames as a geo-located proportional symbol to provide information on the historical geography of Irish surnames as well as quantitative birth counts from the 1890 census (Figure 2).

### *2.3 Symbolizing the Base Map*

Often, the base map of a thematic map is little more than a place-holder for the thematic information and is usually a single, pale colour or white. This is generally a good approach so that the background detail does not clutter the image and distract the reader from the thematic detail. Such an approach is safe and reduces the potential for graphic conflict. For this map though, the design called for something different based on the fact that the design philosophy was to be deliberately eye-catching and, also, that the label symbols in both quantity and representation were bold figures in perception so they could handle a more detailed backdrop without it interfering with information retrieval.

Ireland has relatively low-lying coastal mountains, the highest of which is Carrauntoohil in County Kerry, which rises to 1,038 m (3,406 ft) above sea level, surrounding a central plain, and is characterized by lush green vegetation that gives rise to the name 'Emerald Isle' and much of the association with the colour of emerald green. The base map comprises a high-resolution DEM (digital elevation model), clipped to the Irish coastline and rendered using hypsometric tinting designed to reflect the lush green countryside. The colour ramp was specifically designed for European elevation representation where maximum altitude will be characterised by peat bog and open moorland, effectively a brown landscape (Figure 3a). To provide an improved representation of relief, an Imhof-inspired Swiss hill-shading technique was used to create a subtle yellow-blue tint for the hillshade (Figure 3b). When

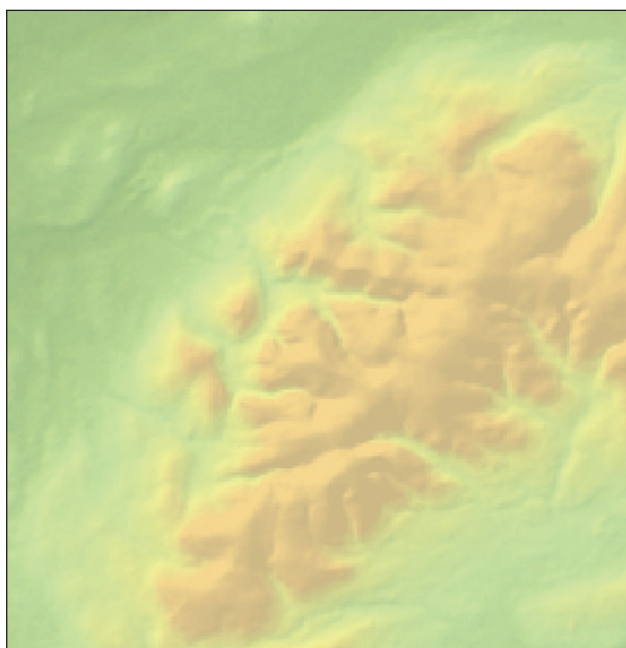




**Figure 3a Green-brown elevation colour ramp**



**Figure 3b Hill-shading (80 per cent transparency)**



**Figure 3c Colour and hill-shading combined**

rendered with 80 per cent transparency over the elevation hypsometry, two effects are achieved. Firstly, the colours of the elevation hypsometry help to accentuate the tonal variation in the hillshade. Secondly, the use of yellows and blues (as opposed to a grayscale) for the hillshade, modifies the colours to give a richer, more realistic tone (Figure 3c).

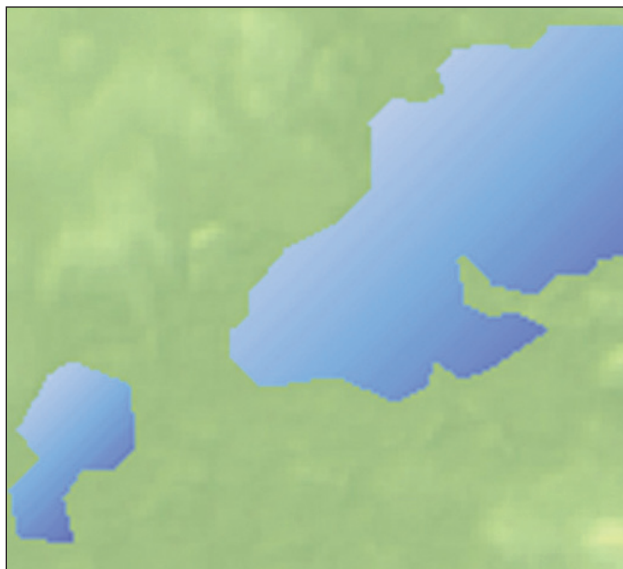
#### *2.4 Symbolizing Water Features*

The base map topography is supplemented by a hydrology layer that contains the major lakes in Ireland. It was decided not to include the river features in order to

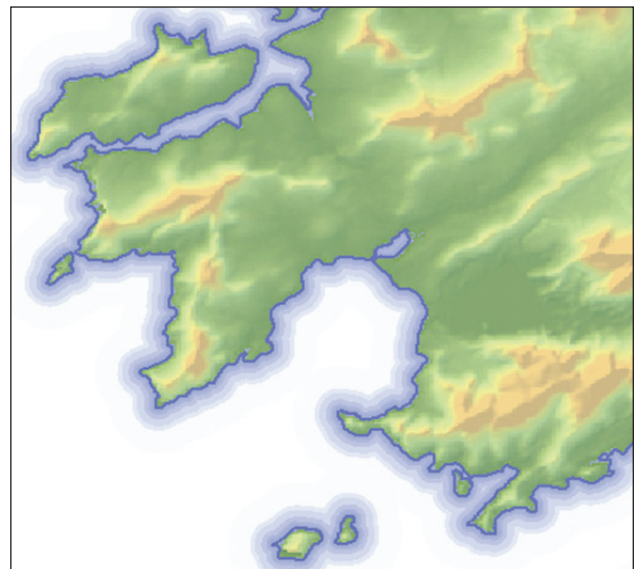
maintain an appropriate balance between the base map as a background and the figural surname labels. Too much detail on the background base map would lessen its impact.

The lakes are symbolized using sun glints to enhance the impression that the blue area fill is a water body. Sun glints are achieved by creating a linear gradient fill from a pale to deeper blue, angled at 45° from the top left to bottom right to enhance the perception of an illuminated source in the top left of the map image. This is the same position that is used for the creation of the hillshade effect and works well in combination with that result (Figure 4a). No polygon outline is used for the lake features so that the lakes appear to sit within the base map as opposed to being made prominent with the addition of a bounding line.

The coastline is represented with a thin blue line as is typical for such features. The usual method of representing the transition from land to sea for a topographic map is to create a vignette that applies a gradient from white (coastline) to pale blue (sea) so that the coastline is highlighted and which then tapers into the pale blue fill used for the water. This gives the connotation that shallower coastal waters (represented in lighter colours) transition into deeper waters without the use of defined edges to represent bathymetry. The approach taken here is reversed, with the intention being to have a different background fill. The map is not meant to appear as if a standard topographic base map had simply been over-printed with the surnames. To achieve this, a pale blue (coastline) to white (sea) vignette is created that allowed the vignette to fade into a white map background. Whilst this is counter to our intuitive view of water depth, here the vignette is simply used as a graphical tool to provide a soft edge to the map and allow the map to appear as being slightly lifted off the page as a background.



**Figure 4a** Use of sun glints to represent water bodies



**Figure 4b** Coastal vignette

Creation of the coastal vignettes is undertaken by applying a series of cartographic representation buffer geometric effects around the coastline polygon at distances of 100, 200, 300, 500, 800, 1000 and 2000 metres to create a non-linear tapered effect. There are a number of alternative ways of creating this effect (such as creating new buffer feature classes with the buffer tool or a single buffered feature class with the multiple ring buffer tool), but being able to control the vignette effect using symbology in the coastline feature class itself was a more efficient approach.

### *2.5 Symbolizing the County Boundary*

The county boundaries are an important administrative boundary dataset to include since they represent the geographical framework within which the census data were collected. An alternative to using a topographic base map is to use a politically shaded map with each county rendered in a suitable pale hue. If shaded using default categories in ArcGIS, each county would be shaded with a unique value which is inappropriate. Instead, a map such as this should be rendered according to the principles of the four colour map theorem which states that given any separation of a plane into contiguous regions, the regions can be coloured using no more than four colours and that no two adjacent regions have the same colour. Despite the theorem often being referred to as the ‘four colour map theorem’ and taking its inspiration from colouring of political maps for countries, the application of the theorem is rarely seen; in part due to the difficulties in determining the representation of individual regions so that the theorem is correctly applied.

In ArcGIS, the four colour map theorem can be implemented by using a four colour map theorem script developed by Gross (2007). The algorithm assigns a number from 1–4 as an attribute in the polygon feature class ensuring that no adjacent counties share the same

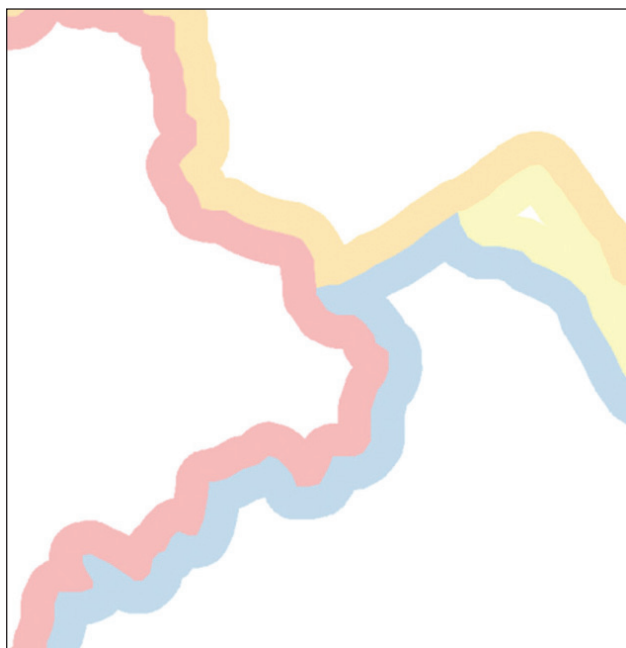
attribute. The county boundary polygons are processed in this way to create a field subsequently used for categorical symbolization; each of the four numbers being symbolized by a pale colour.

A transparent polygon fill was rejected as a way of rendering the county boundary polygons in favour of simply using the polygon boundaries. Applying a colour on top of the base map would have modified the elevation and water body symbology too greatly. To create a thickened boundary between county boundaries, firstly two new feature classes are created using a negative value for the creation of internal donut buffer zones, one at 300 m and the other at 1000 m smaller than the original polygon. These are each used as an ‘erase polygon’ with the original county boundaries feature class, which creates two new feature classes: one of the area within 300 m of each boundary and one of the area within 1000 m of each boundary. These are then symbolized using pale colours (no outline) with 70 per cent transparency applied to the 1000 m polygons and 30 per cent to the 300 m polygons, and then overlaid on top of one another on the map display. The result is a tinted boundary effect between adjacent polygons rendered using the four colour map theorem. Whilst the features are actually polygons in the database, the appearance is as per linework (Figure 5).

### *2.6 Map Marginalia*

The design and placement of map marginalia is an important component of the whole map display. However well crafted the map, a poor overall map page can often ruin an overall display and weaken both user engagement and map communication.

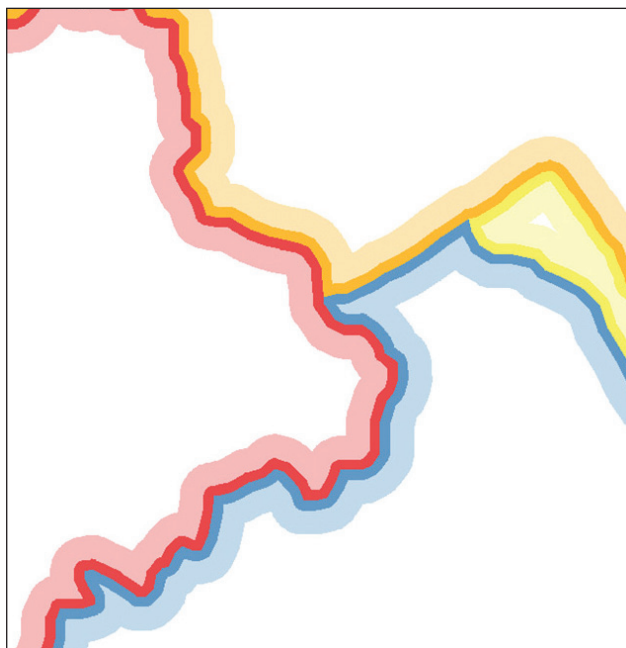
The map page background was to be a highly saturated green to depart from convention and to create an eye-catching display. This was the main reason why the coastal vignette had to be inverted to fade to a white background; so that it could then fade into the green used



**Figure 5a 1000 m internal county boundary polygons**



**Figure 5b 300 m internal county boundary polygons**



**Figure 5c Final county boundary representation**

for the map page more generally. The green was designed to frame the map and was in keeping with the Irish theme. It is generated by first creating a buffered polygon feature class of the coastline (at 25,000 m) which is then used to erase a 'hole' in a further rectangular polygon feature class the size of the page (Figure 6a). A 'donut buffer' geometric effect is then defined to taper the green to white (Figure 6b). Ordinarily, the feature class would be placed underneath all other map components but a limitation in the way in which ArcGIS rasterizes images for print causes the background to the Digital Elevation Model feature class to be rendered in white instead of the 'no fill'

specified, which leads to the problem seen in Figure 6c. This can be overcome by placing the tapered green feature class on top of other map components and fading the green to a 'no fill' to create a genuine hole through which other map detail can be seen (Figure 6d).

Since the map itself uses text as a literal symbol, any further text for county names, physical regions or to label water bodies would create a conflict in interpretation. Thematic maps rarely have locative information beyond the geographical framework used to represent the data so a small location map was used to identify the county names in order to remove the need to place county names on the main map.

A gazetteer was added to the map as a means of locating certain names. With c.2,500 individual labels, the task of identifying the label of interest is akin to searching for a needle in a haystack, so the gazetteer is indispensable for narrowing the search. It locates the position of a name label to a user generated grid. The gazetteer is generated by first creating a polygon feature class grid and coding each grid cell (polygon) with its appropriate designation (e.g. A1, A2, etc.). A series of overlay analyses then attaches the grid designation and county information to each name label. Sorting the 'names annotation' feature class alphabetically then generates a table of names with the counties that can be written to the map layout and arranged around the map shape. The grid was overprinted on the map display to enable the gazetteer to be used effectively.

Usual map inclusions such as a title, in a mix of Celtic and decorative Celtic fonts, short descriptive text, legend and graticule to provide locative and scale information were included. Finally, a Celtic knot border was added, again using a decorative font. The border itself



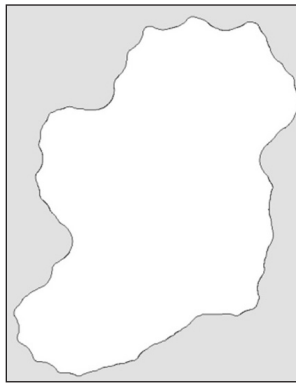


Figure 6a Erased polygon for creation map background

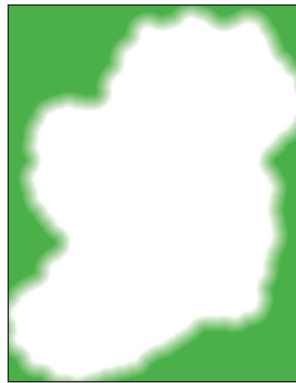


Figure 6b Green to white internal vignette

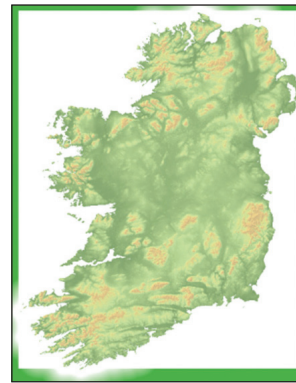


Figure 6c Printing problem created when background polygon is overprinted by rasterized DEM

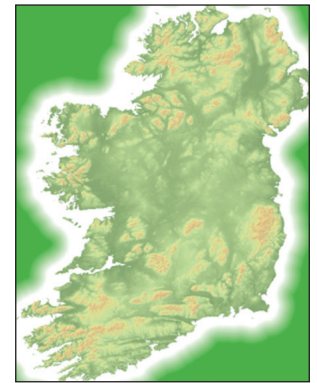


Figure 6d Final solution places the background of polygon with hollow fill over other map detail

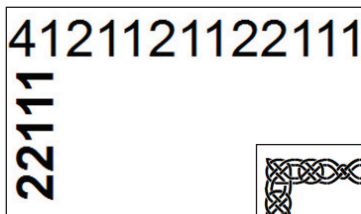
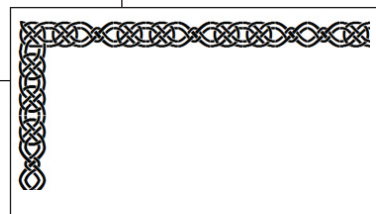


Figure 7a (left) Construction of decorative border

Figure 7b Application of Celtic knot border font using standard font



is simply an array of numbers that are rendered as a decorative border when the font is applied (Figure 7) and individual numbers are aligned.

### 3 Conclusions and Reactions to the Map

The map illustrates the historic and numerical importance of surnames in Ireland according to the 1890 census of population. It is designed in an engaging style based on the use of stereotypical Irish imagery and with a number of innovative and decorative features (Figure 8). As a demonstration of a range of techniques and tools within ArcGIS it illustrates the scope for high quality database-driven cartographic design and production.

The map was designed to be overtly eye-catching and when displayed at the 2009 Esri International User conference in San Diego, California in 2009, it gained considerable interest and provoked many positive responses. It was a focal point of interest during the week as ArcGIS users explored the techniques used in the map's production as well as attempting to identify personal ancestry. It was fascinating watching people view and interact with the map. In many ways it would have provided a superb study in map use with people being drawn to it, and, more often than not, physically touching the map within the first few seconds. Such a human



Figure 8 Geo-Genealogy of Irish Surnames map (original size is A0)

reaction illustrates clearly the way in which map design invokes certain feelings and reactions towards the content matter and its portrayal (Figure 9).

In many ways the map exhibits the premise of what underpins cartography, namely, to design to support effective data communication in a simple yet attractive style. The idea behind the map is a simple one and the



**Figure 9 Interest in the map at the 2009 Esri International User conference**

dataset only consists of two variables, yet with some consideration given to design and production it has been turned into something that adds value to the exploration of the data; revealing new aspects to the patterns within the data, which is visually engaging. It harks back to an age when maps were considered works of art as well as functional documents and only now we have the tools to properly approach the design of a printed map from entirely within a GIS environment.

## Acknowledgements

The map won three awards: ‘Most Unique’, ‘People’s Choice’ and the ‘Cartography Special Interest Group’ prizes at the 2009 Esri International User conference. The authors wish to thank the map gallery judges, professional cartographers and the delegates for these various awards. The map was also awarded the 2010 John C Bartholomew Award for small scale cartography by the British Cartographic Society. Again, the authors extend their thanks to the judges of this award. A high-resolution version of the map at A0 size can be downloaded from the Journal of Maps website at <http://www.journalofmaps.com> (Field and Beale, 2010) from mid-December 2010.

Many of the tools and techniques used in the production of this map are fully described on the Esri Mapping Center website (<http://www.mappingcenter.esri.com>) which contains practical advice on how to design effective and beautiful maps using ArcGIS. The authors are grateful for the documented techniques which helped form the basis on the work herein.

## Software Used

Esri ArcGIS v9.3 (ArcInfo level)

## References

- Field, K.S. and Beale, L. (2010) “Geo-genealogy of Irish Surnames” *Journal of Maps* v2010 DOI: 10.4113/jom.2010.1128
- Gross, T. (2007) *Map Coloring – Four Color a Map Esri Support Center: ArcScripts* available at: <http://arcscripsts.esri.com/details.asp?dbid=14822>

## Biographies

**Dr Kenneth Field** ([kenneth.field@kingston.ac.uk](mailto:kenneth.field@kingston.ac.uk)) is Principal Lecturer in GIS at Kingston University London where he teaches digital mapping and geovisualization courses to BSc and MSc students. With over 20 years experience in UK academia, his work focuses on the use of GIS technology as a design environment for effective cartography and he has produced a number of large format maps that illustrate innovative GIS-based database driven cartography. As well as a Society of Cartographers member he is Fellow of the British Cartographic Society, RGS-IBG and current Editor of The Cartographic Journal.

**Dr Linda Beale** ([l.beale@imperial.ac.uk](mailto:l.beale@imperial.ac.uk)) is Research Fellow in GI Science in the Small Area Health Statistics Unit at Imperial College London where she develops GIS approaches, models and tools for the analysis of large health datasets. She has over 10 years research experience in GIS and is a lead researcher on a number of international projects. Her research interests are primarily in the development of methodologies for spatial analysis, modelling and programming bespoke Geographical Information systems applications. These applications provide tools for the analysis of spatio-temporal phenomena, particularly to explore geographical variation in health outcome and environmental risk for epidemiological study.