MDPI

Article

Multivariate Maps: A Glyph-Placement Algorithm to Support Multivariate Geospatial Visualization

Liam McNabb [1] and Robert S. Laramee [2]

[1] Affiliation 1; 661370@swansea.ac.uk
[2] Affiliation 2; r.s.laramee@swansea.ac.uk
† Department of Computer Science, Swansea University, Bay Campus, Swansea SA1 8EN, UK

Version September 23, 2019 submitted to Information

Abstract: Maps are one of the most conventional types of visualization used when conveying information to both inexperienced users and advanced analysts. However, the multivariate representation of data on maps is still considered an unsolved problem. We present a multivariate map that uses geo-space to guide the position of multivariate glyphs and enable users to interact with the map and glyphs, conveying meaningful data at different levels of detail. We develop an algorithm pipeline for this process and demonstrate how the user can adjust the level-of-detail of the resulting imagery. The algorithm features a unique combination of guided glyph placement, level-of-detail, dynamic zooming, and smooth transitions. We present a selection of user options to facilitate the exploration process and provide case studies to support how the application can be used. We also compare our placement algorithm with previous geo-spatial glyph placement algorithms. The result is a novel glyph placement solution to support multi-variate maps.

Keywords: information visualization, multivariate maps, glyphs, level-of-detail

1. Introduction

Maps are useful for conveying information to both inexperienced and advanced users. There are many types of maps designed to present data but the underlying maps often come with other challenges such as the how the areas are segmented. Fairbairn et al. suggest scale, level of detail, and multivariate data as common challenges for the representation of geo-spatial data [2]. Ward et al. state, *"A problem of choropleth maps is that the most interesting values are often concentrated in densely populated areas with small and barely visible polygons, and less interesting values are spread out over sparsely populated areas with large and visually dominating polygons"* [3]. The challenge of perception (**C1** – size perceivability) is a fundamental one associated with digital maps. Even when trying to rectify this for a univariate map, few solutions enable opportunities to convey multivariate, high-dimensional data. For example, geo-spatial designs (choropleths, cartograms, symbol maps, etc.) only depict uni-variate, or occasionally, bivariate data. This is a challenge for conveying of multi-variate geospatial data (**C2** – multivariate geospatial data). One possibility is glyphs to support multivariate visualization options. However, even if we can present multivariate geospatial data using glyphs, we still run into challenges. If we plot glyphs in their geospatial context, then we risk overlap and over-plotting. In other words, if we place a multivariate glyph at the center of each unit area on a map, the glyphs will either overlap in many cases or be too small to perceive, especially in densely populated areas (see Figure 1) (**C3** – occlusion). Ellis and Dix state *"a glyph representing multiple attributes may need simplifying when reduced in size, resulting in a loss of data"* [4], suggesting that reducing the size of a scalable multivariate glyph can be problematic (**C1** – size perceivability). Another option to address **C3 – occlusion** is to employ structure-driven glyph placement guided by a Cartesian grid. However this common solution de-couples the glyphs from the original geospatial areas they intend to represent. This is the challenge

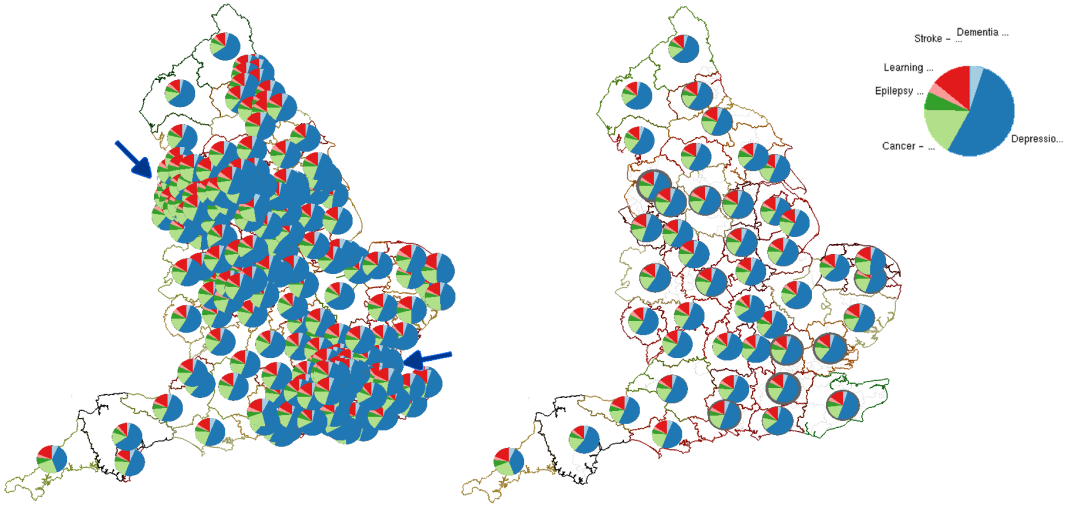


Figure 1. The representation of population health data based on the Clinical Commisioning Groups (CCGs) of England [1]. Refer to Section 5.2 for a case study. (left) Glyphs that are simply placed at the centroid of each region are over-plotted and occluded around London, Manchester, and Liverpool (indicated by blue arrows). (right) Our result using level-of-detail scale-aware maps. Even at a small scale for the figure, we can still clearly differentiate each area's glyph.

of geo-spatial glyph-placement (**C4** – glyph placement). In order to address all four challenges, **C1–C4**, we introduce scale-aware maps, a process of presenting geo-spatial multivariate data based on a desired screen space, that enables dynamic modification to the level of detail shown using both zooming functions and custom scale options. We integrate this with glyphs to present multivariate data in a geo-spatial context to enable interactive exploration, and facilitate easier comprehension with area context using both smooth transitions and uncertainty indicators. We refer to our work as using glyphs as opposed to symbols guided by the definition from Borgo et al. who define a glyph as, *"...an independent visual object that depicts attributes of a data record"* [5]. Our contributions include:

1. A multivariate map with scalable glyph rendering and presentation (in the form of scale-aware maps) (**C1** – size perceivability, **C2** – multivariate geospatial data, **C4** – glyph placement).

2. Dynamic hierarchical glyphs that support zooming, and user-controlled level of detail. (**C2** – multivariate geospatial data, **C3** – occlusion, **C4** – glyph placement)

3. Interactive filters to improve analysis and exploration of multivariate data and comparison of geo-spatial areas. (**C2** – multivariate geospatial data)

In order to do so, we develop solutions that address the four major challenges, **C1–C4**.

2. Related Work

McNabb and Laramee provide an SoS (survey of surveys) for information visualization and visual analytics [6]. The paper includes a section of glyph-focused survey papers, as well as geospatial surveys. Borgo et al. present a survey of glyph design criteria [5]. Fuchs et al. provide a systematic review of experimental studies on data glyphs [7]. Ward presents a taxonomy of different glyph placement strategies (discussed further in the glyph placement section) [8]. We find three survey papers on cartograms [9–11]. We do not consider univariate cartograms within the scope of our work as they distort the boundary geometry of the geo-spatial data, which we avoid in our process. We do not consider magic lenses in our related work [12]. We make this descision considering that magic lenses are manually manipulated, are typically not coupled to geospace, do not necessitate a placement algorithm, and their border transistions are not necessarily smooth [13]. Although we discuss focus+context in the paper, we focus our related work on the topic of maps and glyph placement. We recommend Cockburn et al. for discussion on the topic [14]. We do not consider labels as related work as they do not necessarily apply to multivariate data, and labels do not have to follow

a cohesive hierarchical structure [15]. However the work here could likely be adapted specifically for labeled maps.

Aggregation Techniques: Janicke et al. use a circle packing method to reduce complexity of point-based data at multiple levels of detail [16]. The user can zoom in and out of the map while the point distribution is aggregated to present clear, visible point frequencies. This differs from our work as the data is not coupled to geospatial areas. We also focus on multivariate data which is not featured in their work. Rohrdantz et al. present a multivariate map depicting different new trends across the world using the geospatial map and data proportional glyphs [17]. They use geo-tagging to aggregate their data and do not present any techniques to avoid occlusion. This differs from our work which aims to present glyphs as concicesly as possible, and handles multiple levels of detail with dynamic zooming. Jo et al. present a model for reducing complexity in presenting multiclass data on maps by using aggregative techniques [18]. Their work emphasizes techniques to increase perceivaility without manipulating the underlying geospatial context, whilst ours focuses on increasing perceivability with existing techniques through geospatial unification. Guo creates a technique known as regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP) [19]. Rather than focusing on scale, the algorithm focuses on agglomerating clusters, and Guo discusses six variations of their methodology. Although their technique can calculate varying number of regions, they do not discuss how the number of regions are selected, which differs from our work that dynamically allows restructuring of the hierarchy. The method uses region-based plotting, does not use glyphs, does not consider multivariate data, nor support dynamic zooming.

Related Work with a focus on Cartograms: Dorling visualizes local urban changes across Great Britain [20]. The paper uses multivariate options to review industry distribution, owner-occupied housing, as well as a set of attributes plotted using Chernoff faces as an equal area representation. Slingsby et al. capture the geo-spatial context and transform their results into a grid, which is then represented by a treemap where the hierarchy is based on temporal data [21]. Slingsby et al. present a rectangular cartogram showing the postcodes in Great Britain, where postcode district and unit postcodes form the hierarchy [22]. Cartograms distort geo-space, which we avoid using our procedure. Tong et al. develop Cartographic Treemaps to explore multivariate medical data provided by Public Health England [23]. This is extended to time-varying data [24]. Beecham et al. visualize trends to explain the UK's vote to leave the European Union [25]. They use a juxtaposed view to present equal area cartograms for different variants. Nusrat et al. produce a cartogram that presents bi-variate data using a ring encoding, where the color presents the leading statistic, and the ring thickness presents the value the leading statistic [26]. This differs from our work by emphasizing bivariate design, whilst we provide options for up to nine dimensions to be represented clearly. Our method also supports interactive levels of detail with dynamic zooming.

Related Work with a focus on Multivariate Maps: Multivariate maps have been used in cartography for over 100 years. For example, Minard depicts a multivariate map using pie charts to present cow consumption across France [27]. The pie charts are placed manually. Kahrl et al. present a range of imagery focused on California's water supplies including irrigation applied to crops in the form of dense pixel displays across geo-spatial points [28]. Approaches to add more dimensions to choropleths include bivariate color maps [29,30]. Brewer and Campbell present point symbols for representing quantitative data on maps, including bi-variate options [31]. Although their paper does not focus on glyph placement, their examples place symbols on a region's centroid and exhibit minor occlusion. The work of Andrienko and Andrienko [32] contains a range of examples of multivariate maps using glyphs for thematic maps, including temporal glyphs, and multivariate pie glyphs for forest data. They present glyph placement two ways: region-centroid symbol placement for US states and a Cartesian grid to represent forest data over Europe [32]. They discuss the importance of the link between identifying a symbol and the geo-space it represents (on the map) (**C4** – glyph placement). Slocum et al. provide a chapter on multivariate maps, describing techniques to consider when displaying bivariate, trivariate, and multivariate data [33]. Bertin's Semiology of Graphics is a

Literature		Placement Algorithm	Max No. of Variates	Level-of-Detail	Dynamic Zooming	Smooth Transitions
Aggregation	Janicke et al. [16]	Coordinate-based	1	✓	✓	✗
	Rohrdantz et al. [17]	Coordinate-based	5	✗	✗	✗
	Jo et al. [18]	Region centroid	10	✗	✗	✗
	Guo [19]	No glyphs	1	✓	✗	✗
Multivariate Maps	Minard [27]	Manual	3	✗	✗	✗
	Kahrl et al. [28]	Manual	6	✗	✗	✗
	Olson [29]	No glyphs	2	✗	✗	✗
	Dunn [30]	No glyphs	2	✗	✗	✗
	Brewer [31]	Region centroid	2	✗	✗	✗
	Andrienko and Andrienko [32]	Region centroid/ Grid-based	6	✗	✗	✗
	Slocum et al. [33]	Region centroid/ Grid-based	8	✗	✗	✗
	Bertin [34]	Coordinate-based Grid-based	6	✗	✗	✗
	Elmer [35]	Region centroid	2	✗	✗	✗
	Kresse and Danko [36]	Coordinate-based	2	✗	✗	✗
	Tsorlini et al. [37]	Region centroid	6	✗	✗	✗
GGP	Chung et al. [38]	Scatterplot	9	✗	✗	✗
	McNabb et al. [39]	No glyphs	1	✓	✓	✗
	This	**Dynamic Region centroid**	9	✓	✓	✓

Table 1. A breakdown of the related literature. For each paper, we review the type of placement algorithm used, the number of data variates presented, if multiple levels-of-detail are depicted of the data, whether dynamically moving between levels of detail is discussed, and if so, whether smooth transitions are implemented to increase recognition of dynamic glyphs. The table is sorted into aggregation, multivariate maps, and general glyph placement (GPP) literature.

foundational work on cartography. The work covers many different maps including multivariate maps of up to 6 variates, using grid-based and coordinate-based placement schemes [34]. Elmer reviews symbol consideration for bivariate thematic maps, but does not support more than two variates [35]. Our algorithm supports an arbitrary number of variants depending on the glyph design. Kresse and Danko present geographic techniques from basic principles to applications [36]. They present a table of visual variables to represent data, applied to a given map and symbols. Tsorlini et al. present a taxonomy of thematic cartography symbols, including multivariate options [37]. The symbols are presented as a hierarchy, focusing on the number of attributes, and arrangement. The focus of their work is not on glyph placement, nor dynamic level-of-detail.

Related Work with a focus on Glyph Placement: Ward and Lipchak create a software tool for cyclical, temporal multivariate data. Glyphs are placed in an ordered grid structure to enable easy comparison between similar months, or entire years [3]. They also use a radial structure. Our work differs from this work by focusing the glyph placement coupled to geo-spatial areas. Ward presents a taxonomy of different glyph placement strategies [8]. They introduce glyph designs that can be used, and 15 glyph placement strategies together with a flow chart of how the glyph placement is driven (data-driven or structure-driven). Our placement strategy is considered geo-spatially data-driven. As the modifications are made before the placement process, it falls into original→derived→data-driven. This is expanded by a subsection in a further survey by Ward [40]. Ropinski and Preim present a taxonomy of usage guidelines for glyph-based medical visualization [41]. As opposed to Ward's placement taxonomy, they suggest glyphs should be placed based on physical characteristics or

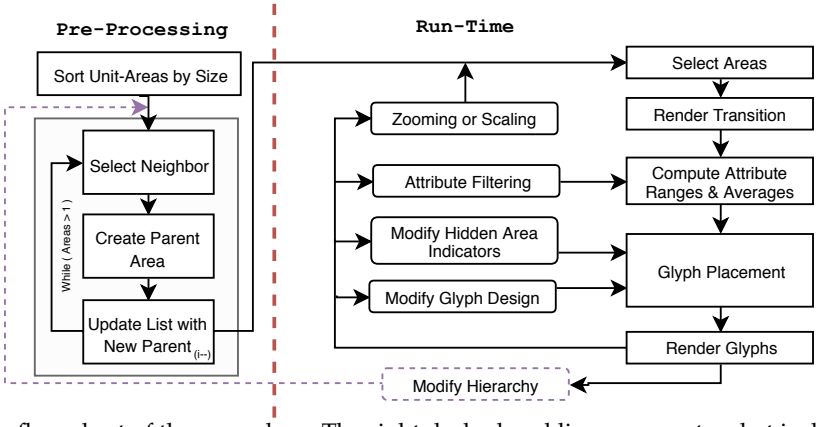


Figure 2. The flow chart of the procedure. The right dashed, red line represents what is discussed in the scope of this paper. The pre-processing steps are discussed in Section 4 and discussed by McNabb et al. in more detail [39].

anatomical features. Borgo et al. provide a section on glyph placement which extends on both of the previous taxonomies by suggesting user-driven placement [5]. Chung et al. discuss glyph sorting strategies and present horizontal axis bins, applying them to sport-event analysis glyphs [38]. Our work differs by guiding our glyph placement strategy based on a 2D geospatial context. As evidenced by Table 1, the algorithm we present offers a novel combination of glyph placement, multivariate data, level of detail, dynamic zooming, and smooth transitions.

3. Design Goals and Tasks

We derive six main tasks to motivate our design process.

T1 – Visualization Overview: Provide a glyph-based overview of multivariate data on a map free from occlusion (**C3** – occlusion).

T2 – Multivariate Map: Offer a selection of informative multivariate glyphs to compare trends between regions (**C2** – multivariate geospatial data).

T3 – Glyph Placement: Clearly couple encoded glyphs to their geo-spatial contexts (**C4** – glyph placement).

T4 – Level-of-detail Leverage scale-aware maps to enable exploration of the data at multiple levels of detail (**C1** – size perceivability).

T5 – Filtering: Support the exploration of multivariate geo-spatial data with user options with varying glyph designs and filters (**C2** – multivariate geospatial data).

T6 – Smooth Interaction: To provide smooth and fluid transitions between the different levels of detail (**C4** – glyph placement).

4. Overview

This section provides the pre-processing steps used to create the scale-aware maps, the run-time process for transitioning between glyph densities, and the options we provide to enhance the exploration of the data. The pre-processing steps are based on previous work by McNabb et al [39]. The purpose of the pre-processing step is to build a map whose areas are always perceivable, unit areas that are too small [42] are unified until they reach a minimum area threshold set by the user. The area-based hierarchy construction is a recursive algorithm broken down into three sub-routines. In these three steps, we select the optimal neighbor for merging, we identify the shared boundary between the given area and its neighbor, and unify them to create a new area which is then inserted back into the list of areas sorted by size. A flow chart of the procedure is found in Figure 2. Once the pre-processing steps are completed, we move to our run-time implementation. The first step is to identify optimally-sized areas, render any transitions between previously rendered and the newly

selected areas, compute the glyph visual mapping of the data, and update the glyph properties, before rendering the glyphs. From here, we provide five options to transform the view. Zooming or scaling to dynamically modify the multi-variate glyphs, attribute filtering to modify the glyph properties and mapping, modification of hidden area indicators to customize glyph design, the revision of glyph attributes to customize the glyph design, and modification of the underlying hierarchy which returns the algorithm back to the pre-processing procedures. We disucss the steps in detail in the sections that follow.

4.1. Pre-processing

We use a recursive procedure to create a hierarchical area-based data structure. An area hierarchy is created for each contiguous region, where each area is merged with its closest neighbor identified using a customizable distance metric [39]. We start with a merge candidate list filled with the sorted unit-areas (for one contiguous region). There are three main sub-routines: (a) neighbor selection, (b) creating the parent area, and (c) updating the merge candidate list. If only a single unit-area remains in the merge candidate list, no further merges can be processed and the procedure terminates. (a) In order to select an appropriate neighbor to join, we use a general and flexible distance metric for amalgamation evaluated between neighboring areas which is used to identify our ideal neighbor, based a 'distance' metric identified.

We use the closest distance considered as the optimal selection for a neighbor, $D = w_a.\frac{a}{a_{max}} + w_d.\frac{d}{d_{max}} + w_\alpha.\frac{\alpha}{\alpha_{max}} + w_{b_s}.(1 - \frac{b_s}{b_{s_{max}}})$. This method is discussed further by McNabb et al. to discuss why these attributes are important for neighbor selection [39]. The measure consists of four constituents: Smallest area (a), euclidean distance between centroids (d), univariate data value variance (α), and shared boundary resolution (b_s). We search and identify each common vertex between neighboring areas to identify the shared boundary. We update the sorted area list by removing the two merged areas, and inserting the newly created parent, which may be used as a new merge candidate. This is repeated until only one area remains in each contiguous region.

Value calculation for unified areas: The Modifiable Areal Unit Problem (MAUP) [43] is an important aspect to consider when discussing the modification of boundaries or values. We address this by providing the user options to customize calculation of aggregated univariate data values as well as the customizable distance metric used to evaluate area merge candidates. The data is linked to the unit areas during the initial loading of the shape files. Before the area tree is built, the user can select the type of value amalgamation. This enables the user to choose options of sums, frequencies, and value averages. When amalgamating values using sums, the value can be calculated using aggregation. Qualitative values are calculated using frequencies. For a detailed description of parent value calculation, see McNabb et al. [39]

4.2. Geospatial Glyph Placement

In order to enable multi-variate maps, a number of technical challenges must be addressed including: 1) A glyph-placement strategy, 2) A hierarchical glyph design, 3) dynamic level-of-detail support, 4) smooth transitions between child and parent glyphs, 5) multi-variate filtering and selection, and 6) customizable interactive user options. Furthermore, the hierarchical glyph design must support the encoding of aggregation error.

We select visible areas and glyphs based on a minimum area scale requirement (a percentage of screen space), m. When the map is rendered, the tree is traversed using a depth-first search (DFS) to identify which areas are rendered. If an area is larger than m we test two criteria: if the area is a leaf node, or if either the left or right child is smaller than m. If either of these true, we render the area. For each area displayed, we create a glyph using the area's centroid to position the glyph. We create a glyph that reflects the given area's multivariate data values (based on the user's selection)We first set the size of the glyph at 2.5% of the screen space as the default, a heuristic we derive from McNabb et al.'s previous user study on perception of scale on choropleth maps [42]. As the zoom level of the map

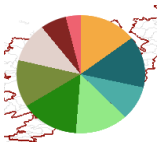
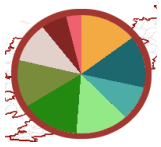
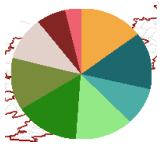
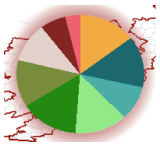
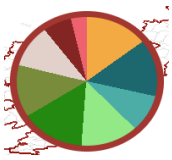
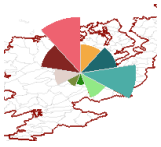
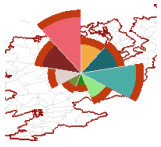
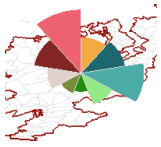
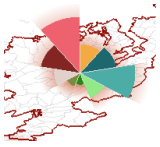
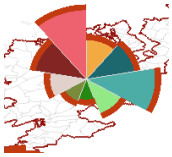
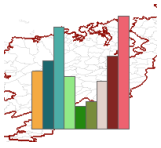
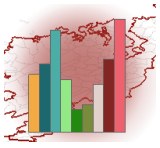
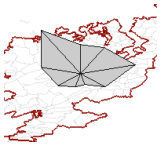
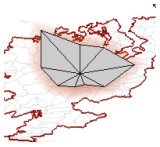
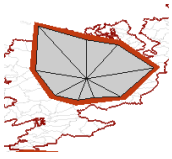
Glyph Design	Hidden Density Indicators				
	No Indicator	Outline	Size	Shadow	Size + Outline
Pie Chart					
Polar Area Chart					
Bar Chart					
Star Chart					

Table 2. Previews of the different glyphs, and the hidden density indicators provided in the application. Each glyph represents the same area, reflecting the same hidden indicator values, and attributes. Refer to our third case studyfor more details on the values.

changes, different areas may meet *m* and therefore be presented, creating a dynamic presentation of glyphs. This addresses **T1 – Overview** and **T3 – Glyph Placement**, by providing a clear overview of the map with no occlusion, and clearly encoded geo-spatial context. As the size of the glyph changes so does the perceived ideal map structure. The user can manually find their own perceived preference using sliders or using naive estimated glyph placement.

4.3. Glyph Selection

We provide the user four common glyph design options to represent the data (see Table 2). We chose these four typical options due to their common occurrence in geo-spatial literature [32]. However, the principles we describe can be applied to any multi-dimensional glyph. The user can switch between each glyph design at any point once the hierarchical data structure has been built. These glyph options are:

Pie Chart: Pie charts are an easily recognizable and practical design, making it a suitable option to present multivariate data. Pie charts are primarily used to present distribution per geospatial area, where the angle of a segment is mapped to each data dimension proportionally (see Table 2).

Polar Area Chart: Originally published by Nightingale [44], a polar area chart is another radial plot but with equal segment angles. The radius or each slice corresponds to the values of each dimension, which facilitates comparison between geo-spatial areas. The polar area chart features different names including the wheel, coxcomb, or wing chart. (See Table 2).

Bar Chart: The bar chart is one of the most visually recognizable visual designs. Values are assigned to bar heights. aligned to the horizontal axis for easy value comparison. (see Table 2).

Star Glyph: Originally presented by Siegel et al.[45], a star glyph presents values using lines originating from the same point, at equal angles. The endpoints connect to form a unique polygon based on the length of each line (see Table 2).

This addresses the requirements for **T2 –Multivariate Maps**. We choose four standard glyph designs

Figure 3. An example of a smooth transition made between two child glyphs that translate toward the new parent node. Both child glyphs decrease in opacity, whilst the new parent glyph increases in opacity.See also the accompanying video for this dynamic behavior.

as a proof-of-concept. Glyph placement, not glyph design is the focus of this paper. The principles we present can be extended to any multivariate glyph.

4.4. Adjusting Level-of-Detail with Glyph Density

Adjusting glyph density can be handled in two different ways. First, we give the user a slider which depicts m, a minimum area requirement. The parameter m represents a percentage of screen space. This is used as the primary variable for the depth-first search (DFS). We also allow the user to interactively zoom in or out of the map. This changes the visible extents of the map, modifying the screen space covered by each area. These options enable the rendering of perceivable glyphs, meeting the requirements for **T4 – Level-of-detail**.

4.5. Smooth Merging and Splitting Transitions

In order to increase the smoothness of user interaction and changes to glyph size when zooming or manipulating m, we apply smooth transitions to child glyph merging and parent glyph splitting. When the user reduces the number of glyphs by either zooming out of the map or increasing the level of detail, glyphs translate towards the origin of their parent in the hierarchy while the opacity is reduced until it is no longer visible. The parent increases in opacity until it is fully opaque, creating a smooth transition. When adding new glyphs (zooming in or reducing minimum scale), the new child glyphs translate away from their parent and increase in opacity to provide a similar effect. Using this technique, we fulfill the requirement for **T6 – Smooth Interaction**. See Figure 3. This dynamic animation is best viewed in the accompanying video.

4.6. Dynamic Average Glyph Legend

We provide a dynamic average glyph legend to present how the multivariate data dimensions of the glyph are encoded. The advantage this provides is that each individual glyph on the map can be compared to the average shown in the legend. Each variate is given a label, which provides context to the user about what is presented. The data used to present the glyph is made meaningful by

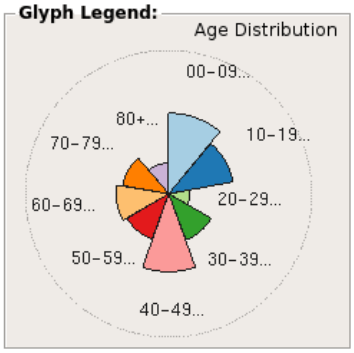


Figure 4. A representation of a glyph legend. The data represents the prevalence of population per age range. The dotted circle represents the full scale of the glyph or the largest value for each dimension. The glyph legend shows the average values over the whole data set.

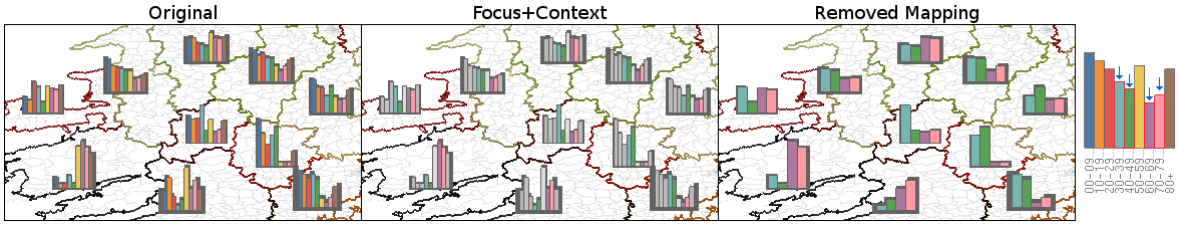


Figure 5. Filtering options. The left shows the original image. The center shows focus+context rendering, which renders the context in greyscale. The right image shows a multivariate glyph mapping filter, which re-renders the data based on the focus dimensions. The legend indicates the focus dimensions using the blue arrows.The Munster area refers to the bottom-left glyph, which is notable for Case Study 3.

visualizing the average value of each dimension. In Figure 4, we can see that there seem to be some extreme values for the 80+ and 20–29 range, causing the average per area to be quite small overall.

4.7. Attribute Filtering

Our first filter option is to re-calculate the glyph design with only the toggled dimensions. Each data dimension can be toggled using a check-box incorporated to represent data variates in the glyph design. This allows the user to focus on or emphasize data dimensions that may reveal trends. We support user filtering using focus+context rendering. We provide a gray-scale option which removes the color from context data dimensions, enabling easier comparison. This supports the requirements we set forth in **T5 – Filtering**. See Figure 5.

4.8. Unit Area Density Indicators

We present unit area density indicators that provide a visual queue indicating how hidden unit areas are distributed, and encourage the user to explore the visualization through multiple levels of detail. When two child glyphs merge to form a parent, the child glyphs are then hidden. Our glyph design maps the number of merges to a range of different visual indicators that generally surround the glyph. See Table 2. We offer four options:

Outline: Outline maps the unit area quantity around each glyph to thickness. The thickness of the outline grows as more areas fall underneath a glyph.

Size: Rather than provide an outline, the glyph's overall size increases as the glyph represents more unit areas. This works especially well with pie charts, that emulate a proportional map.

Size+Outline: Size + Outline uses a combination of the two previous options.

Shadows: Rather than an outline with a constant opacity, we enable for the user to choose a gradient, enabling less occlusion in the representation.

These unit area density indicators are inspired by the work of Chung et al. [38] where the indicator was effective, but used to represent another data dimension (as opposed to the density of a map). We also give the user an option to represent the indicator mapped to color. The color represents the scale the glyph encodes, as opposed to other visible encodings. This enables the user to gain an understanding of how manipulation of glyph density can affect the map if a transition is made. See Figure 2. This addresses our requirements of **T4 – Level-of-detail**.

4.9. Interactive User Options

We provide additional user options to support **T5 – Filtering**. We present a range of user options including value range filters, advanced focus+context rendering options, estimated glyph placement, and context administrative areas.

Data Range Options: We provide data range filtering to enable customized local and global design options for dimension encoding. On a local range, the user can shift the value range to present the data dimensions based on the values found in the leaf nodes (the original dataset), or clamp the ranges

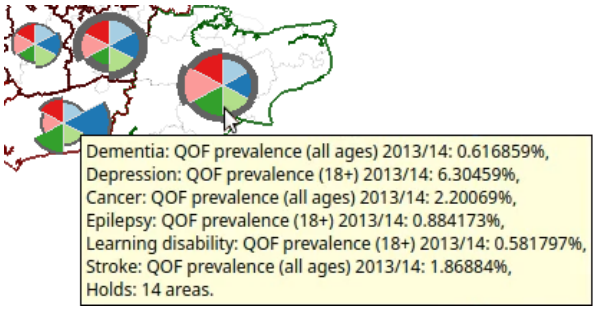


Figure 6. By hovering over a glyph (for this example we use the south-east of England), the user is provided with details on demand of the multi-variate datavalues depicted and the number of areas represented by the glyph.

amongst those that are currently being rendered to enable a more accurate data range to compare data dimensions. We also support global range options by enabling the user to depict each variant based on its own range, or by creating a range based on the highest and lowest value of all mapped dimensions.
Advanced Filters: We include two advanced filters to render focus+context for the user. For numerical values, the user can present focus+context based on values higher or lower than the average value per data dimension.
Color Map: We provide the user with a variety of color maps, selected from published research, including ColorBrewer [46] (Refer to Table 2) and Colorgorical [47] (Refer to Figure 5).
Glyph Scaling: We allow the user to scale the current size of the glyph. This enables the user to explore a ratio between the minimum scale and size of glyphs that meets their own data.
Naive estimated glyph placement: Using the current size of the glyph, we can support the user to make an estimation of the minimum screen space necessary to remove occlusion with the use of a button. This makes it easier to obtain a starting point, in order to decide the design of the map they would like to use. This can also be linked to the glyph scaling to allow for automatic re-placement when scaling the glyph.
Context Administrative Areas: We can provide additional context behind the areas by rendering every leaf area in a context view, which is shown in Figure 5.
Details on demand: We allow the user to obtain precise insight into the multivariate data by providing a textual representation of the values associated with a glyph by hovering over any glyph. We also include the number of areas depicted to give better context to the underlying data. See Figure 6.

5. Evaluation

We evaluate our glyph placement for multivariate maps in two ways. First, we provide three cases for the use of the multivariate maps with varying data sources. We then provide a comparative evaluation of our glyph placement strategy against a standard Cartesian grid-based glyph placement to evaluate its effectiveness and any advantages or potential drawbacks against pre-existing techniques.

5.1. Video and Images

We include larger resolution images as well as video representation of the case studies discussed in the paper. These can be found at the following links: https://vimeo.com/314225790 and in the supplementary video upload.

5.2. Case Studies

In order to evaluate our glyph placement strategy, we incorporate three case studies. In our first case study we examine health indicators coupled with CCGs (Clinical Commisioning Groups) within England. Secondly, we examine the average income of US counties over 10 year periods. Finally, we look at the age distribution across the electoral divisions of the Republic of Ireland.

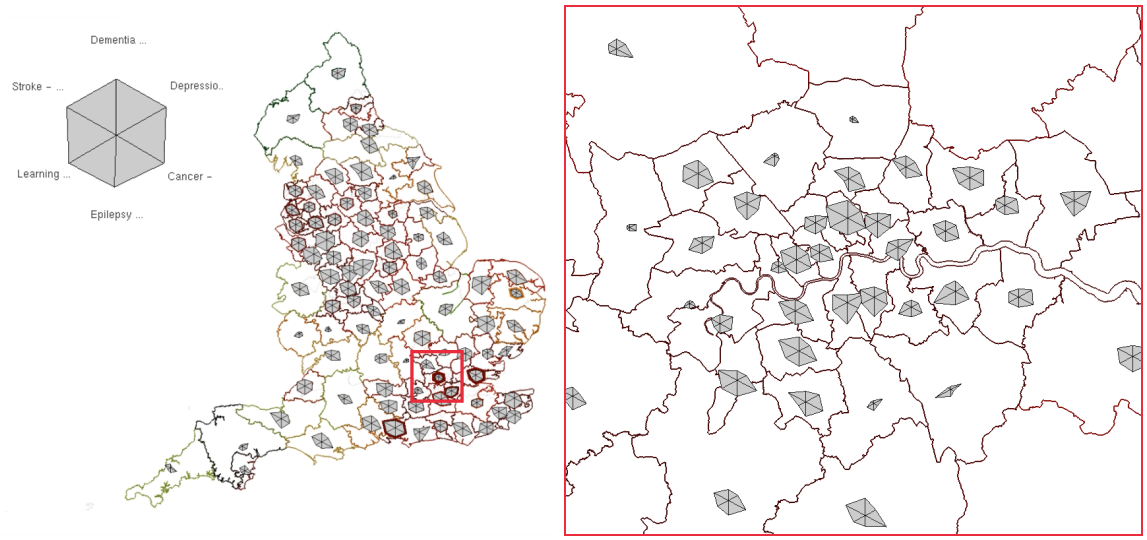


Figure 7. After noticing the southwest of London as exhibiting lower prevalence rates than the rest of London (red box), we zoom in to observe the cause. We can see low prevalence rates are more frequent among the northwest, with some low prevalence rate in the southeast. We can also now identify the particular CCGs. Glyph scale increased for zoomed in view. See Section 5.2.

Case 1: England's Clinical Commissioning Groups (CCGs): Our first case uses a dataset focused on England's Clinical Commissioning Groups, which represent areas of NHS practices. People who reside in the area are generally expected to use the same practices. We explore the prevalence of afflictions per CCG area, including Dementia, Depression, Cancer, Epilepsy, Learning Disabilities, and Stroke. Refer to Figure 1 to show an example of the CCGs represented [1]. There are over 200 CCGs. Figure 1 also compares a naive glyph placement using region centroids with our glyph placement strategy. Placing glyphs in every area centroid is one of the most common placement strategies.

For this example, overlapping glyphs are prevalent around London, Liverpool, and Manchester, if we simply render a glyph at each centroid (Figure 1, left). We start with pie chart glyphs to obtain an overview of the data (**T1 – occlusion**). As pie charts always extend to the maximum radius, combining our level-of-detail glyph placement algorithm combined with the estimated minimum size placement removes most of the occlusion, enabling visual comparison between the points (**T2 – multivariate maps**, **T3 – glyph placement**). The first trend we notice is that depression has a majority prevalence across most CCGs, although we can observe that the Kernow CCG exhibits an uncommon distribution, caused by a larger distribution of cancer as opposed to other pie glyphs (Figure 1). At this point, we can filter out depression prevalence, however we can glean a bit more information by switching glyph design. If we transition to the star glyph, we can see that this is due to both the larger prevalence of cancer and low rate of depression in comparison to other prevalence values for CCGs (Figure 10(a)). As the star glyphs have varying extents, we can reduce m down to 0.8% to increase the level of detail with no occlusion (**T4 – level-of-detail**). At this scale, London is split into 3 zones, where we can clearly see the northwest point has lower prevalence overall (Figure 7). We can investigate this by zooming in to London (**T6 – smooth interaction**). We zoom in to see a larger number areas (rendered by m). Not only do we find Barnet, Enfield, Hillingdon, and Hounslow to have low prevalence rates overall, but Bromley and Croydon in south London also show these signs (dementia, stroke, and cancer prevalence in particular). See Figure 7.

Case 2: Counties of the United States: Our second example explores counties in the US. We look at the average income over 40 years for each county in the United States from 1979, 1989, 1999, to 2009 in 10 year increments [48]. The US consists of over 3,000 counties.

Rendering the glyphs presents a large frequency of occlusion and therefore we use the estimated minimum size, m, to reduce the large number of glyphs to something more accessible. Starting with

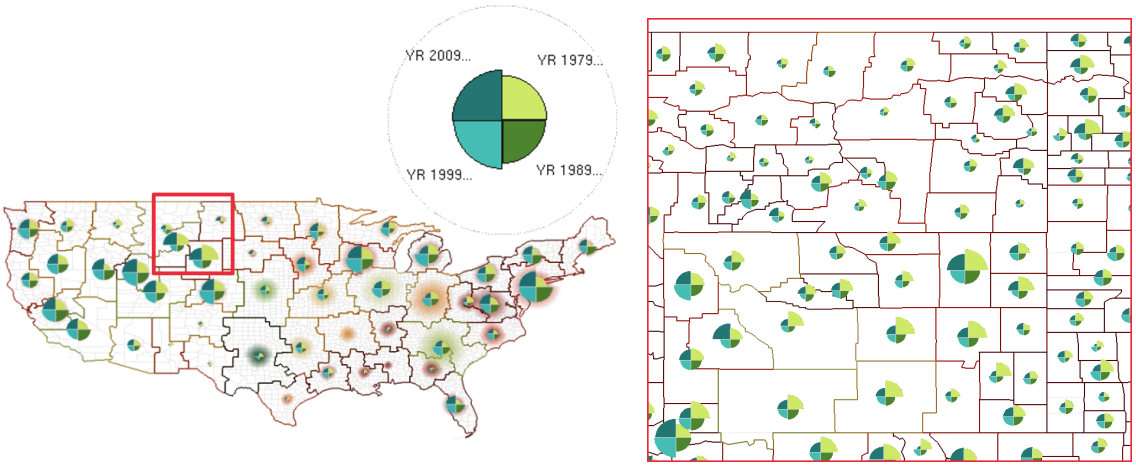


Figure 8. We discover counties around Wyoming and Montana have are higher average income in 1979 and 2019 than usual. We zoom in, and can verify this amongst particular counties. Glyph scale increased for zoomed in view.

the pie chart shows a standard distribution where the average income increases per time period (**T1 – overview**). Since each glyph represents a number of areas, we adjust the range indicator to represent areas that are rendered, and switch to a polar area chart to visualize the data (Figure 10(b)). The wheel glyph shows higher income on the east and west coasts, with the lowest value glyphs across the center of the United States. Wyoming and Montana have some uncommon behavior, where 1979 and 2009 show much stronger average income than their other variants. Zooming in, Wyoming exhibits a tendency to exhibit a higher average income in 1979 over the 40 years, independent of their standing amongst the rest of the US counties. The counties of Sublette and Teton are the exceptions to this which hold stronger mean incomes in 1999 and 2009. See Figure 8.

Case 3: Electoral Divisions of the Republic of Ireland: For our final case, we look at the electoral divisions of the Republic of Ireland. Our data set records population distribution across each division, which is split into nine groups, 0–9 years old, 10–19, 20–29...up to 80+ [49]. There are over 3,400 electoral divisions in the Republic of Ireland.

Similar to Case 2, there are a large number of electoral divisions so we immediately choose to reduce the visible areas to a perceivable number using the estimated scale glyph placement, and adjust a filter to represent a clamped range. In this example, we use bar charts to represent the data. If we look towards Munster, we can see an unusual population distribution, where the proportion of the population, 50 or above, is uncommonly high, and the proportion of people, under 50, is uncommonly low (Figure 5). Zooming in, Glanmore, Canuig, Tahilla, Derriana, Dawros, Ardea, Castlecove, and Caher seem to be the leading factors in this trend. See Figure 9.

5.3. Comparative Evaluation: Grid-Placement vs. Dynamic Placement

We evaluate user interpretation of the data against a typical grid structure for glyph placement. We use a Cartesian grid (20^2) structure that places glyphs at approximately the same size and resolution as the algorithm developed in this paper. Each area is assigned to a cell of the grid, closest to its centroid, where glyphs are derived using the same process as our algorithm. In terms of design, we try to keep both structures similar. In our algorithm, we use a thickened outline to signify the unified area the glyph presents which is not possible for the grid placement version because unit areas are arbitrarily split using a Cartesian grid. We therefore show the presented areas using a lower line width to avoid over representation. Other than this, all design elements are the same and we allow the user to adopt filters and user options identically. However, we use a standard grid structure and therefore the grid structure does not necessarily handle multiple levels of detail. Examples are shown in Figure 10.

Figure 9. After noticing a strange inconsistency in the south-west of the Republic of Ireland, we zoom in and can verify that this trend can be found amongst a selection electoral divisions. Glyph scale increased for zoomed in view.

We look at two main aspects of placement, geometric coupling, and value representation. First we examine geo-metric coupling. As the grid is uniform, in all instances, the grid allows for a larger number of glyphs, however, this can be seen as a clear positive. If we start with Figure 10(a), it is sometimes difficult to verify where a glyph is when areas reside within a corner of multiple grid slots. If we look at the central-east coast of England, identifying even large areas becomes difficult. This is because the areas are considered uniform, and therefore are distributed uniformly, as opposes to our presented algorithm which attempts to avoid this as much as possible. In examples Figure 10(b), our placement uses fewer glyphs due to the large variance in wheel glyph extents, as opposed to the grid placement that presents roughly twice as many. The grid results in the same limitation as above, with strong difficulty in understanding how values are mapped to their glyph counterparts. We run into a second problem with the density of the areas, where administrative areas make it difficult to perceive where a grid cell covers, providing little understanding of the context. Both of these problems follow on to Figure 10(c).

For value representation, the algorithms do show differences, which is to be expected, in accordance with the modifiable areal unit problem [43]. For Figure 10(a), both representations pointed to the same observation in our case study. Figure 10(b), exhibits a significant difference. The grid-placement greatly skews the value representation of the US counties, due to some grid elements covering secluded cells. The west coast contains a grid cell with San Francisco, which causes most of the other glyphs to be quite small, independent of the data range option selected. It becomes difficult to compare the two placement options. Although that is the case, both placement schemes lead to the observation found that the time period of 1979 maps to a larger segment in Wyoming. Figure 10(c) also presents the observations found in our case study, although the concentration is more spread out, which can be considered better for examinations. If we consider the ability to zoom, we feel that only the Cartesian grid representation of the Republic of Ireland can lead to a fair comparison for observation, and this is only based on trial-and-error.

6. Future Work

There are many avenues for future work we can consider. At the moment, we use the raw derived centroid as a placement strategy. although this removes a lot of density and occlusion, there is still some wasted space. We believe that by adding some overlap removal, we could use space more

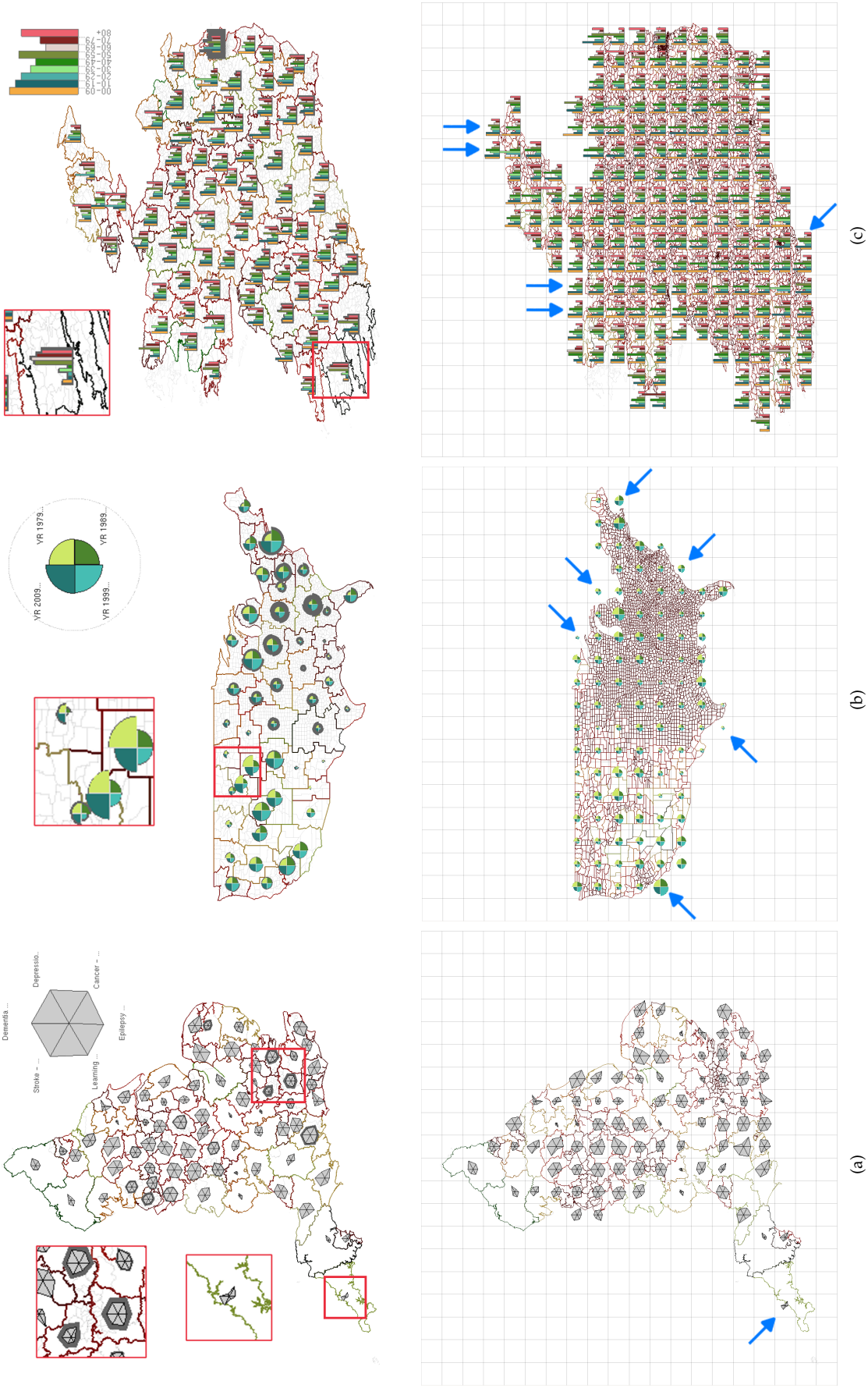


Figure 10. Glyph-placement comparison: top row – our method, bottom row – grid-based. (a) Comparison of CCGs (b) Comparison of US Counties (c) Comparison of Ireland's Electoral Divisions. Glyphs based on grid-placement are often de-coupled from the geo-space the represent. The blue arrow signifies the cause of the incomparable data values.

efficiently, whilst still avoiding any decoupling problems. Although we present some case studies, the algorithm could be more carefully compared to other glyph placement strategies with a user-study evaluation which would allow us to understand the advantages and disadvantages to a typical users' exploration process. Although we think the use of smooth transitions is a great tool for understanding variation, they may not always be necessary. We feel that there are many avenues for exploring at multiple levels of detail. For example, directly zooming to a glyphs unit area extents may not need to represent zooming, to speed up exploration.

7. Conclusion

We present a glyph placement algorithm supporting multivariate geospatial visualization at different levels of detail. We discuss how we create scale aware map, and apply the process to glyph placement. We also discuss the different glyph options and filters we have designed to support exploration of multivariate data. Finally, we evaluate the algorithm by examining the separate use cases, and compare against a pre-existing glyph placement strategy.

References

1. Public Health England. FingerTips PHE. Accessed: 2018-11-20.
2. Fairbairn, D.; Andrienko, G.; Andrienko, N.; Buziek, G.; Dykes, J. Representation and its relationship with cartographic visualization. *Cartography and Geographic Information Science* **2001**, *28*, 13–28.
3. Ward, M.O.; Lipchak, B.N. A visualization tool for exploratory analysis of cyclic multivariate data. *Metrika* **2000**, *51*, 27–37. doi:10.1007/s001840000042.
4. Ellis, G.; Dix, A. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics* **2007**, *13*, 1216–1223.
5. Borgo, R.; Kehrer, J.; Chung, D.H.; Maguire, E.; Laramee, R.S.; Hauser, H.; Ward, M.; Chen", M. "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications". *"Eurographics State of the Art Reports"* **"2013"**, pp. "39–63".
6. McNabb, L.; Laramee, R.S. Survey of Surveys (SoS)-Mapping The Landscape of Survey Papers in Information Visualization. Computer Graphics Forum. Wiley Online Library, 2017, Vol. 36, pp. 589–617.
7. Fuchs, J.; Isenberg, P.; Bezerianos, A.; Keim, D. A Systematic Review of Experimental Studies on Data Glyphs. *IEEE Transactions on Visualization and Computer Graphics* **2016**. forthcoming, doi:10.1109/TVCG.2016.2549018.
8. Ward, M.O. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization* **2002**, *1*, 194–210.
9. Tobler, W. Thirty five years of computer cartograms. *ANNALS of the Association of American Geographers* **2004**, *94*, 58–73. doi:10.1111/j.1467-8306.2004.09401004.x.
10. Nusrat, S.; Kobourov, S. Task taxonomy for cartograms. *17th IEEE Eurographics Conference on Visualization (EUROVIS - short papers)* **2015**.
11. Nusrat, S.; Kobourov, S. The State of the Art in Cartograms. Eurographics conference on Visualization (EuroVis)–State of The Art Reports. The Eurographics Association, 2016, Vol. 35, pp. 619–642.
12. Tominski, C.; Gladisch, S.; Kister, U.; Dachselt, R.; Schumann, H. A survey on interactive lenses in visualization. Citeseer, 2014, Vol. 3.
13. Tominski, C.; Gladisch, S.; Kister, U.; Dachselt, R.; Schumann, H. Interactive lenses for visualization: An extended survey. Computer Graphics Forum. Wiley Online Library, 2016.
14. Cockburn, A.; Karlson, A.K.; Bederson, B.B. A review of overview+ detail, zooming, and focus+ context interfaces. *ACM Comput. Surv.* **2008**, *41*, 2–1.
15. Luboschik, M.; Schumann, H.; Cords, H. Particle-based labeling: Fast point-feature labeling without obscuring other visual features. *IEEE Transactions on Visualization and Computer Graphics* **2008**, *14*, 1237–1244.
16. Jänicke, S.; Heine, C.; Stockmann, R.; Scheuermann, G. Comparative Visualization of Geospatial-temporal Data. GRAPP/IVAPP, 2012, pp. 613–625.

17. Rohrdantz, C.; Krstajic, M.; El Assady, M.; Keim, D. What is Going On? How Twitter and Online News Can Work in Synergy to Increase Situational Awareness. 2nd IEEE Workshop on Interactive Visual Text Analytics Task-Driven Analysis of Social Media, 2012.
18. Jo, J.; Vernier, F.; Dragicevic, P.; Fekete, J.D. A Declarative Rendering Model for Multiclass Density Maps. *IEEE transactions on visualization and computer graphics* **2019**, *25*, 470–480.
19. Guo, D. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *International Journal of Geographical Information Science* **2008**, *22*, 801–823, [https://doi.org/10.1080/13658810701674970]. doi:10.1080/13658810701674970.
20. Dorling, D. The visualization of local urban change across Britain. *Environment and Planning B: Planning and Design* **1995**, *22*, 269–290.
21. Slingsby, A.; Wood, J.; Dykes, J. Treemap cartography for showing spatial and temporal traffic patterns. *Journal of Maps* **2010**, *6*, 135–146.
22. Slingsby, A.; Dykes, J.; Wood, J. Rectangular hierarchical cartograms for socio-economic data. *Journal of Maps* **2010**, *6*, 330–345.
23. Tong, C.; Roberts, R.; Laramee, R.S.; Berridge, D.; Thayer, D. Cartographic Treemaps for Visualization of Public Healthcare Data. Computer Graphics and Visual Computing (CGVC); Wan, T.R.; Vidal, F., Eds. The Eurographics Association, 2017. doi:10.2312/cgvc.20171276.
24. Tong, C.; McNabb, L.; Laramee, R.S.; Lyons, J.; Walters, A.; Berridge, D.; Thayer, D. Time-oriented Cartographic Treemaps for Visualization of Public Healthcare Data. Computer Graphics and Visual Computing (CGVC); Wan, T.R.; Vidal, F., Eds. The Eurographics Association, 2017. doi:10.2312/cgvc.20171273.
25. Beecham, R.; Slingsby, A.; Brunsdon, C. Locally-varying explanations behind the United Kingdom's vote to leave the European Union. *Journal of Spatial Information Science* **2018**, *2018*, 117–136.
26. Nusrat, S.; Alam, M.J.; Scheidegger, C.; Kobourov, S. Cartogram visualization for bivariate geo-statistical data. *IEEE transactions on visualization and computer graphics* **2018**, *24*, 2675–2688.
27. Minard, C.J. Carte figurative et approximative des quantites de viandes de boucherie envoyees sur pied par les departements et consommees a Paris, 1858. republished in 'Palsky, G. "Des chiffres et des cartes-la cartographie quantitative au XIXe siècle, Paris, èditions du CTHS, coll." (1996)'.
28. Kahrl, W.L.; Bowen, W.A.; Brand, S.; Shelton, M.L.; Fuller, D.L.; Ryan, D.A. *The California Water Atlas*; State of California, 1978.
29. Olson, J.M. Spectrally encoded two-variable maps. *Annals of the Association of American Geographers* **1981**, *71*, 259–276.
30. Dunn, R. A dynamic approach to two-variable color mapping. *The American Statistician* **1989**, *43*, 245–252.
31. Brewer, C.; Campbell, A.J. Beyond graduated circles: varied point symbols for representing quantitative data on maps. *Cartographic Perspectives* **1998**, pp. 6–25.
32. Andrienko, N.; Andrienko, G. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*; Springer Science & Business Media, 2006. doi:10.1007/3-540-31190-4.
33. Slocum, T.A.; McMaster, R.B.; Kessler, F.C.; Howard, H.H. *Thematic cartography and geovisualization*; Pearson Prentice Hall Upper Saddle River, NJ, 2009.
34. Bertin, J. *Semiology of graphics: diagrams, networks, maps*; University of Wisconsin press, 1983.
35. Elmer, M.E. Symbol Considerations for Bivariate Thematic Mapping. PhD thesis, University of Wisconsin–Madison, 2012.
36. Kresse, W.; Danko, D.M. *Springer handbook of geographic information*; Springer Science & Business Media, 2012.
37. Tsorlini, A.; Sieber, R.; Hurni, L.; Klauser, H.; Gloor, T. Designing a Rule-based Wizard for Visualizing Statistical Data on Thematic Maps. *Cartographic Perspectives* **2017**, *0*.
38. Chung, D.H.; Legg, P.A.; Parry, M.L.; Bown, R.; Griffiths, I.W.; Laramee, R.S.; Chen, M. Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization* **2015**, *14*, 76–90.
39. McNabb, L.; Laramee, R.S.; Fry, R. Dynamic Choropleth Maps - Using Amalgamation to Increase Area Perceivability. The 22nd International Conference on Information Visualization (IV); IEEE, , 2018; pp. 284–293. doi:10.1109/iV.2018.00056.
40. Ward, M.O. Multivariate data glyphs: Principles and practice. In *Handbook of data visualization*; Springer, 2008; pp. 179–198.

41. Ropinski, T.; Preim, B. Taxonomy and usage guidelines for glyph-based medical visualization. SimVis, 2008, pp. 121–138.
42. McNabb, L.; Laramee, R.S.; Wilson, M. When Size Matters - Towards Evaluating Pereivability of Choropleths. The Computer Graphics & Visual Computing (CGVC) Conference 2018; The Eurographics Association, , 2018; pp. 163–171. doi:10.2312/cgvc.20181221.
43. Openshaw, S. The modifiable areal unit problem. *Concepts and techniques in modern geography* **1984**, *38*.
44. Nightingale, F. *Notes on matters affecting the health, efficiency, and hospital administration of the British Army : founded chiefly on the experience of the late war*; Harrison and Sons, St. Martin's Lane, London, 1858.
45. Siegel, J.H.; Farrell, E.J.; Goldwyn, R.M.; Friedman, H.P. The surgical implications of physiologic patterns in myocardial infarction shock. *Surgery* **1972**, *72*, 126–141.
46. Brewer, C.A. A Transition in Improving Maps: The ColorBrewer Example. *Cartography and Geographic Information Science* **2003**, *30*, 159–162. doi:10.1559/152304003100011126.
47. Gramazio, C.C.; Laidlaw, D.H.; Schloss, K.B. Colorgorical: creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics* **2017**.
48. U.S. Department of Commerce. USA Counties Data File Downloads. Accessed: 2018-11-20.
49. Ordnance Survey Ireland. Open Data for Census 2016 Ireland. Accessed: 2018-11-20.