

Article

On the Use of ‘Glyphmaps’ for Analysing the Scale and Temporal Spread of COVID-19 Reported Cases

Roger Beecham ^{1,*}, Jason Dykes ², Layik Hama ¹ and Nik Lomax ¹

¹ School of Geography, University of Leeds, Leeds LS2 9JT, UK; l.hama@leeds.ac.uk (L.H.); n.m.lomax@leeds.ac.uk (N.L.)

² giCentre, Department of Computer Science, University of London, London EC1V 0HB, UK; j.dykes@city.ac.uk

* Correspondence: r.j.beecham@leeds.ac.uk

Abstract: Recent analysis of area-level COVID-19 cases data attempts to grapple with a challenge familiar to geovisualization: how to capture the development of the virus, whilst supporting analysis across geographic areas? We present several *glyphmap* designs for addressing this challenge applied to local authority data in England whereby charts displaying multiple aspects related to the pandemic are given a geographic arrangement. These graphics are visually complex, with clutter, occlusion and salience bias an inevitable consequence. We develop a framework for describing and validating the graphics against data and design requirements. Together with an observational data analysis, this framework is used to evaluate our designs, relating them to particular data analysis needs based on the usefulness of the structure they expose. Our designs, documented in an accompanying code repository, attend to common difficulties in geovisualization design and could transfer to contexts outside of the UK and to phenomena beyond the pandemic.

Keywords: COVID-19; geovisualization; multivariate visualization; cartography; glyphs



Citation: Beecham, R.; Dykes, J.; Hama, L.; Lomax, N. On the Use of ‘Glyphmaps’ for Analysing the Scale and Temporal Spread of COVID-19 Reported Cases. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 213. <https://doi.org/10.3390/ijgi10040213>

Academic Editor: Wolfgang Kainz

Received: 8 February 2021

Accepted: 23 March 2021

Published: 1 April 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A key challenge in spatial analysis is how to visualise and interpret patterns over both spatial and temporal dimensions simultaneously. The increasing availability of longitudinal or time series data that also contains spatial information is driving methodological development and a number of reviews have been carried out, for example addressing the cartographic options in traditional maps [1] and interactive visual data analysis systems [2], or more foundational papers establishing the design space for spatiotemporal data visualization [3]. Work in this area is constantly evolving, driven by new data and enabled by new technologies. The most obvious and pressing of which has been the spread of the SARS-CoV-2 virus, also known as the Corona Virus or COVID-19. Due to the gravity of the pandemic, and certainly efforts to make case data easily accessible (e.g., [4]), there has been intense focus on spatiotemporal analysis. In a short period of time during the first wave of the pandemic, a wide range of visualization efforts were published and shared on-line, tracking its spread and trajectory in populations around the world [5].

Many of these analyses of area-level COVID-19 cases and deaths data attempt to grapple with a challenge familiar to geovisualization: how to compare the development of cases aggregated to area-level, whilst retaining the spatial context associated with those areas? Animation was used with great effect to communicate a sense of the pace of change and spread of cases across geographic areas [6,7] and was even used in government briefings [8]. Animation nevertheless has limits as it relies on visual memory—on the ability of users to track information between frames as time progresses [9]. One way of getting around this is to cumulatively build, or ‘reveal’, a visualization such that information in previous frames is retained. This approach is more viable for certain chart types than others. Mathieu Rajerison [6] and Michal Škop [10], for example, show the spatial progression of COVID-19 case data with animated line charts overlaid on a polygon map. Static

equivalents have previously been applied across a range of domains and with different nomenclature—value flow-maps [11], geosparklines [6] and *glyphmaps* [12], our preference.

This paper develops various *glyphmap* designs for studying the spread of COVID-19 cases data during the first wave of the pandemic for Local Authorities in England. To inform this design work, we identify several Data Requirements (DatRs) from the existing attempts to represent area-level spread in the COVID-19 cases data. These are to display case data according to:

- DatR1 **Geography**—case numbers by area displayed in an arrangement that reflects their spatial relationships.
- DatR2 **Absolute number**—of cases by area (total and/or cumulative case counts).
- DatR3 **Relative number**—of cases by area, for example expressing total and/or cumulative case numbers as a share of population size.
- DatR4 **Rate of change**—the extent to which growth in cases by area is speeding-up or slowing-down.
- DatR5 **Time elapsed**—against an absolute or relative start point in time.
- DatR6 **Case history**—case numbers by area either continuously (daily case releases) or at specific *milestones* in the disease trajectory.
- DatR7 **Cases relative to local ‘peak’**—whether the daily growth in case numbers at a time point by area has reached its fastest recorded growth rate.

Additional to these are four Design Requirements (DesRs) to which our designs should adhere if they are to be successful in supporting detailed spatiotemporal analysis:

- DesR1 **Concurrent**—all data items must be shown simultaneously to support comparison, exploration and other synoptic tasks.
- DesR2 **Discernible**—all marks must be discernible, with limited or manageable occlusion.
- DesR3 **Prioritised**—phenomena and patterns that are important must be visually salient.
- DesR4 **Estimable**—graphical techniques used to encode quantities must enable accurate estimation.

Meeting each of the design and data requirements together is challenging. If all data items are to be shown concurrently (DesR1) visual clutter is increasingly likely, making individual items difficult to discern (DesR2). Through a detailed design discussion, this paper attempts to draw attention to these difficulties in order to suggest candidate designs and make claims about the effectiveness of particular design configurations. The contributions are:

- A **survey** of recent glyphmap approaches for spatiotemporal analysis of COVID-19 cases data;
- Glyphmap **designs** for spatiotemporal analysis of cases data that meet our data and design requirements and that may transfer to other contexts, **implemented** using a high-level visualization grammar (*ggplot2* [13]);
- **encoding schematics**, a novel means of describing design candidates, closely linked to their implementation, and which help draw attention to issues of data density and encoding effectiveness;
- **claims** around the likely **effectiveness** of our **novel visualization** designs in light of shifting data analysis needs related the pandemic.

2. Background

2.1. COVID-19 Visualization and Glyphmaps

The rapid spread of COVID-19 has resulted in numerous visualizations that aim to track the virus over time, across geographic areas. Data are typically captured at regular intervals and aggregated according to administrative areas within the region of interest. The virus’s extent—it has impacted most countries and occupied the policy agenda of most governments—offers an unprecedented opportunity for innovation and learning, with a diverse community of researchers analysing the same datasets and asking similar questions.

Whilst the data are novel, the visualization challenge is familiar. Temporal patterns (that are linear) are often best enabled using standard time-series charts and aligned scales, where time progresses along a horizontal axis and the quantity of interest is represented along a vertical axis. With a 2D geospatial arrangement, horizontal and vertical position is already in use and instead temporal analysis must be effected via other means. For example, via difference representation [14], creating a single ‘change’ statistic between two points in time and displaying the measure directly as a change map [14]; animation, representing each time cut as a frame on a thematic map; introducing a third dimension on which time might be represented (e.g., space-time-cubes [15]); or by generating discrete time series graphics for each spatial unit, using established idioms such as line charts, and applying a 2D geospatial arrangement to form glyphmaps [12].

As the COVID-19 outbreak began to establish itself in March 2020, numerous maps were published and shared online displaying area-level counts of cases data in standard choropleths [16,17], proportional-symbol maps [16] and dot-density maps [18]. Colin Angus’s work [7] was particularly detailed in its attention to temporal patterns of locally-occurring peaks in English local authorities, contrasting a small-multiple heat map with authorities ordered top-to-bottom according to the recency of their peaks in cases, and a geospatial arrangement (standard map and hexagon cartogram) where change-over-time is conveyed by animating over the daily new cases data and colouring by proximity to local peaks.

As this design work continued, various glyphmaps emerged demonstrating how spatiotemporal patterns in area-level cases data could be represented simultaneously. These glyphmap designs appeared to convey the sorts of detailed, multivariate spatiotemporal structure relevant to the analysis of COVID-19 identified in our Data Requirements (DatRs). For example, Mathieu Rajerison [6] used lines to signify cumulative case trajectories in France, located at the centroid of départements, with emphasis placed on départements experiencing rapid growth in absolute cases using animation. The same encoding was used by Michal Škop to represent area-level case data for the Czech Republic, with the addition that overall case numbers are also double-encoded using colour lightness as well as height [10]. Reis Thebault and Abigail Hauslohener, of The Washington Post, published a similar visualization for US counties [19], a version of the graphic is in Figure 1. In addition to annotating counties with distinctive growth rates, the authors used colour value and thickness to encode growth rates and relative exposure (current cases as a share of the population), respectively.

The Washington Post piece is particularly impressive as a data dense visualization—lines representing over 3000 US counties and 85 time points are shown simultaneously. Many careful design decisions are made in order to emphasise certain aspects and deemphasise others. For example, in varying line thickness by relative exposure and line colour by growth rate, those counties less exposed to the virus and with slower growth rates are much less salient. It is challenging to discern complete trajectories here, but instead a typical model or expectation of these trajectories can be learnt from visually scanning the graphic. That spatial autocorrelation [20] affects the trajectories is helpful—an overall pattern of exposure can be abstracted almost pre-attentively, before eyes are then drawn to exceptions. Initially these might be towards the extreme end; tall, steep, dark and thick lines, suggesting large absolute numbers, rapid growth rates and high exposure. Secondly, interesting subtle patterns can be discerned, for example a thick and mid-dark line surrounded by lines that are generally lighter and thinner; a county that appears locally exceptional in having a comparatively high growth and exposure rate. Occlusion is nevertheless inevitable, meaning that not all marks in the visualization are **discernible** (DesR2). This might not be so problematic if the data items in the visualization are appropriately **prioritised** (DesR3)—that is, in using thickness to convey relative exposure (cases relative to population size) and colour and height growth rates, and showing the lines with a spatial arrangement, it may be that the more important patterns in case growth are indeed given appropriate emphasis. For example, the obvious outlier counties (labelled)

and the counties in the west of the US experiencing relatively low relative virus growth rates when compared nationally, but that are locally exceptional. Nevertheless, there may also be interesting patterns in the case growth trajectories occurring in the North East and Mid West that are unlikely to be **discernable** (DesR2) due to the heavy occlusion.

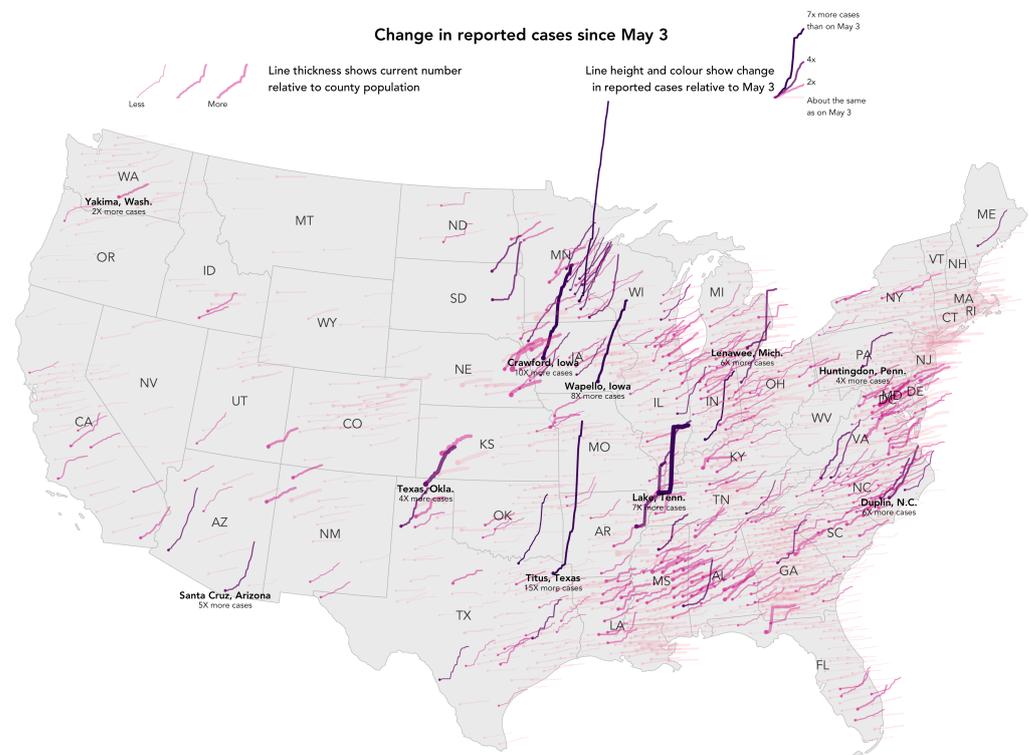


Figure 1. Map displaying growth in COVID-19 cases by US County, based on the design by Thebault and Hauslohner [19]. Our graphic uses data collated by New York Times and made available via the `covdata` R package [21]. Documented code for the graphic is in the code repository accompanying this paper.

2.2. Evaluating Design Candidates

In this paper, we develop several glyphmap design candidates for representing COVID-19 cases data and, in a similar way to the Washington Post example, pay particular attention to how the glyphs might be parameterised and layered to build data density and reveal important structure. Developing design *candidates*, where multiple solutions are generated to meet analysis requirements, is an important aspect of visual data analysis [22]. How to evaluate and select the ‘best’ candidates is not straightforward. *Upstream* evaluation involves discussing low-level design choices, applying cognitive and perceptual principles to ensure that guidelines around effective visualization design are not violated. *Downstream* evaluation involves soliciting feedback from target ‘end-users’ and selecting designs based on this discussion [23]. A category of downstream evaluation that lends itself particularly to DesR4—whether the discussed chart idioms [24] encode quantities in a way that can be accurately **estimated**—is the perception-based experiment (e.g., [25]), which is a particular challenge for geographic contexts [26]. We reserve this for future work and instead focus on the upstream encoding threat—the possibility of violating cognitive and perceptual principles—as it is especially relevant to the sorts of applied visualization projects envisaged, where graphics are encoded with multiple data items at the same time. Our *encoding schematics* provide a novel framework for structuring this design candidate evaluation. Although the data on which our designs are based is limited to English local authorities, the designs could transfer to country settings other than the UK and analysis domains other than Epidemiology.

3. Datasets and Technologies

Our designs use a single COVID-19 dataset and level of geographic hierarchy in the UK: confirmed cases data recorded for 150 Upper Tier Local Authority areas in England, released daily by Public Health England [27], where upper tier authorities comprise county council and unitary authority areas. We wish to characterise spread in the first wave of the virus and therefore make the decision to analyse case data reported up to 1 June 2020. Geographic boundary data were obtained from the Office for National Statistics's (ONS) Open Geograpy Portal [28] and population data from the ONS mid-year population estimates released at County Council and Unitary Authority level. Restricting ourselves to this geography has implications for our designs. It may be desirable to analyse spatiotemporal trajectories in the case data at lower levels of geography, for example the ONS Deaths Involving COVID-19 dataset released at Middle-layer Super Output Area level [29]. However, PHE's is a complete dataset ranging back to the first confirmed case in England in late January 2020. The 150 Upper Tier Local Authorities in England also present an interesting visualization challenge. They are administrative delineations covering both densely populated urban centres, provincial towns and more sparse rural areas—as discussed in Section 4.3, this creates challenges related to occlusion and visual clutter, which we try to address using semi-spatial layouts.

A code repository, with discussion of data, code and design choices, is published alongside this paper. Detailed in this repository is data processing code for the derived variables on which our graphics depend: 7-day rolling case counts, the local 'peak' for each local authority over the study period (cases to 1 June 2020) and relative case counts by local authority normalised according to local population size. All data graphics were programmed using the ggplot2 package in R. For many of the graphics, or design candidates, we present a static and animated alternative within the repository.

4. Designs

In this section, we describe our glyphmap designs for analysing local authority-level COVID-19 cases data in detail. We introduce a mechanism for describing our designs, which we call *encoding schematics*. These schematics are then used to document and assess the design choices made when creating glyphmaps of increasing data density. As is often the case, increased data density results in problems of occlusion and clutter, particularly in this case where there is a requirement to represent multiple features of the case data (DatR2–DatR7) for local authorities with a geospatial arrangement (DatR1). We reflect on this and layout options which involve relaxing geography, overcoming the occlusion problems at the expense of geographic precision, demonstrating the kind of *design exposition* [30,31] that we use to guide our design process (Figure 2).

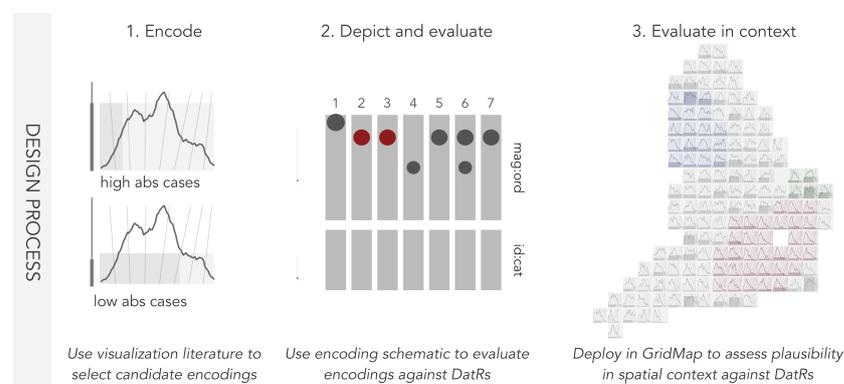


Figure 2. Graphical representation of the upstream evaluation process for our design candidates.

4.1. Describing Designs: Encoding schematics

Visualization design is ultimately a process of decision-making. Data relevant to an analysis use case must be filtered and prioritised before being matched to visual

channels [24] (Figure 3) through which components of a data graphic can be encoded. Certain visual channels are more effective at encoding certain types of data than others—and so judicious decisions must be made around which data to encode with which visual channel [24,32]. For data rich graphics where several data items are to be encoded simultaneously, as in the Washington Post example, these decisions become increasingly challenging.

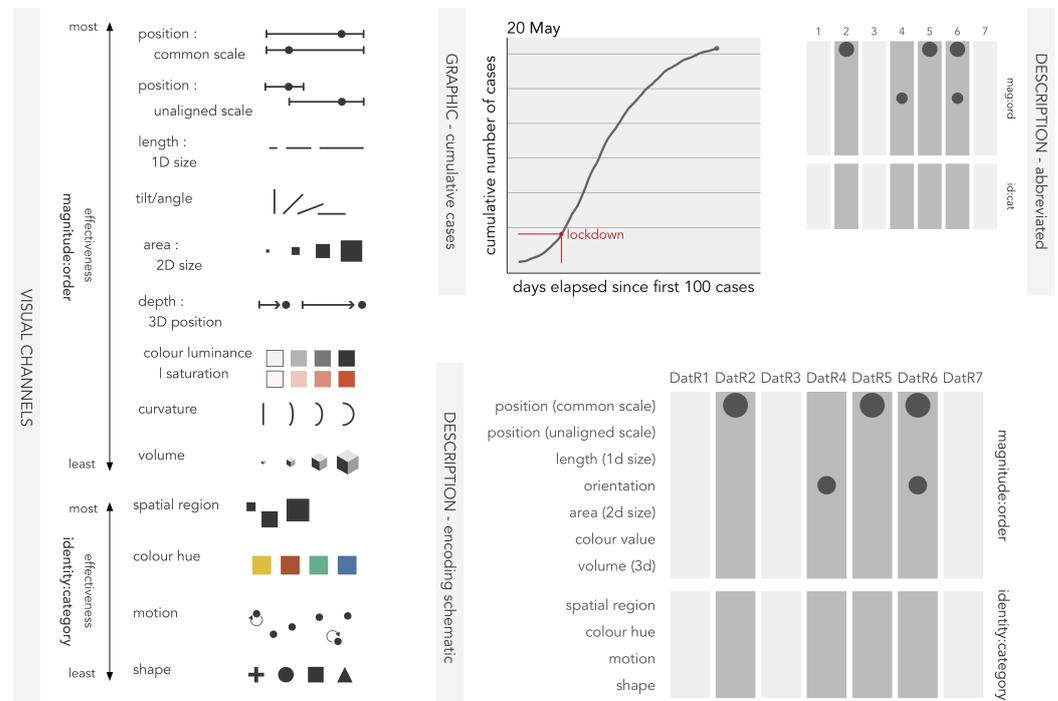


Figure 3. Left: Adapted from Munzner [24], visual channels through which data can be encoded, ordered according to effectiveness. This ordering is based on empirical work by Cleveland and McGill [33], later replicated by Heer and Bostock [34]. Right: example schematic describing a line chart of cumulative cases, and above to the right is a simplified version that we use in this paper for concise descriptions. We can quickly see from the main schematic that 4/7 DatRs are addressed (grey columns), with broadly effective encodings (high large dots) and some double encoding (columns with two dots).

In our case, there are numerous ways in which graphics might be encoded—data matched to visual channels—to meet the seven Data Requirements (DatRs) identified in the introduction. To describe and compare across the possible design candidates we propose using *encoding schematics*. A characteristic example is illustrated in Figure 3 and in this case the schematic describes the encoding for a familiar daily cumulative cases chart, featured in the figure itself. The columns translate the seven DatRs into data channels [24] and the rows identify the visual channels [24] with which these data can be encoded. So in this case, the linechart conveys DatR2 **absolute number** of cases, DatR4 **rate of change**, DatR5 **time elapsed** and DatR6 **case history**. This encoding is denoted with dark fills for each of the columns. Furthermore, identified via dots is the corresponding visual channels (rows). DatR2 **absolute number** and DatR5 **time elapsed** are encoded using position on an aligned scale—vertical and horizontal position, respectively, DatR4 **rate of change** is encoded using orientation (line steepness) and DatR6 **case history** is communicated indirectly, and can be derived using a combination of position and orientation. Notice that the arrangement of the visual channels (rows) in the matrix is deliberate. Following Munzner [24] visual channels are grouped by the types of data to which they are most appropriately applied—quantitative and ordinal, or ‘magnitude’ and ‘order’ according to Munzner [24] and categorical and nominal, ‘identity’ and ‘category’ [24]—and are then ordered by their effectiveness based on graphical perception literature (e.g., Figure 3).

Each design candidate presented in the paper is accompanied by an encoding schematic. This allows high-level differences between design candidates to be quickly communicated and invites consideration of the combination of encodings used when generating complex data graphics. For example, data density is implied by the number of columns with darker fills; a sense of encoding effectiveness can be read by the vertical position of the dot within the table and this is double encoded with dot size—the larger and higher the dot within the matrix the more effective, following Munzner [24], the encoding; and encoding efficiency can also be quickly inferred from scanning the columns—more than one dot implies double encoding. Clearly, for designs attempting to encode many data items concurrently, strict adherence to this schema might not guarantee successful design. Many dots higher up in the matrix implies high data density *and* encoding effectiveness, but may result in designs that are cluttered and unintelligible. However, the schemas help draw attention to this—they provide a structured framework for evaluating the encodings used in our designs. As the graphics are also generated using a library underpinned by *grammar of graphics* thinking [13] (`ggplot2`), the schematics are also closely linked to implementation, clarifying the mapping of data- to visual- *aesthetics* intrinsic to `ggplot2` specifications.

4.2. Charting Idioms: Lines and Ridge Contours

We explore two chart idioms for designing to the seven DatRs: *line* and *ridge contour* charts (Figure 4). The line chart requires little explanation. DatR5 **time elapsed**, the number of days since the first 100 cases was recorded, is encoded along the horizontal axis; the cumulative number of daily cases (DatR2) is encoded along the vertical axis; and a line connecting daily cumulative case counts is drawn in temporal order. The chart can be static and display the full case history (DatR6) or designed to animate over the cases data, as demonstrated via the frames in Figure 4. The *ridge contour* charts attempt to encode loosely the same data properties as the lines. They were proposed as an abstraction for emphasising comparison across particular aspects of our Data Requirements. For example, whilst line charts are the ‘obvious’ approach to encoding the full temporal trajectory of the virus, comparison of **absolute** (DatR2) and **relative** (DatR3) case counts or emphasising particular milestones in the disease trajectory (DatR6 **case history**) is also important. There is precedent in cartography for using ridge symbolisation for encoding quantities [35] and the different ridge shapes that result from varying ridge width and height according to separate data items—thin and short, thin and tall, short and wide, tall and wide—might provide convenient visual short-hands. In our ridge charts, linear time varies along the horizontal axis and frequency along the vertical. Rather than a single line connecting points in temporal order, a separate line is drawn for each frame (release of cases data), similar to the ‘lockdown’ annotation in the line chart, but connecting positions on the horizontal and vertical axes to contrive a triangle or ridge shape. Case history is therefore encoded via animating over the ‘current’ frame, which is made bold and also through the case milestones that persist through the animation. The case milestones are a form of *visual benchmark* used frequently in cartographic contexts [36], and appear at regular intervals—every 1000 cases in this instance. Milestones located close together imply a fast **rate of change** (DatR5) and milestones further apart imply a slow rate of change; it is for this reason that we name them *contours*.

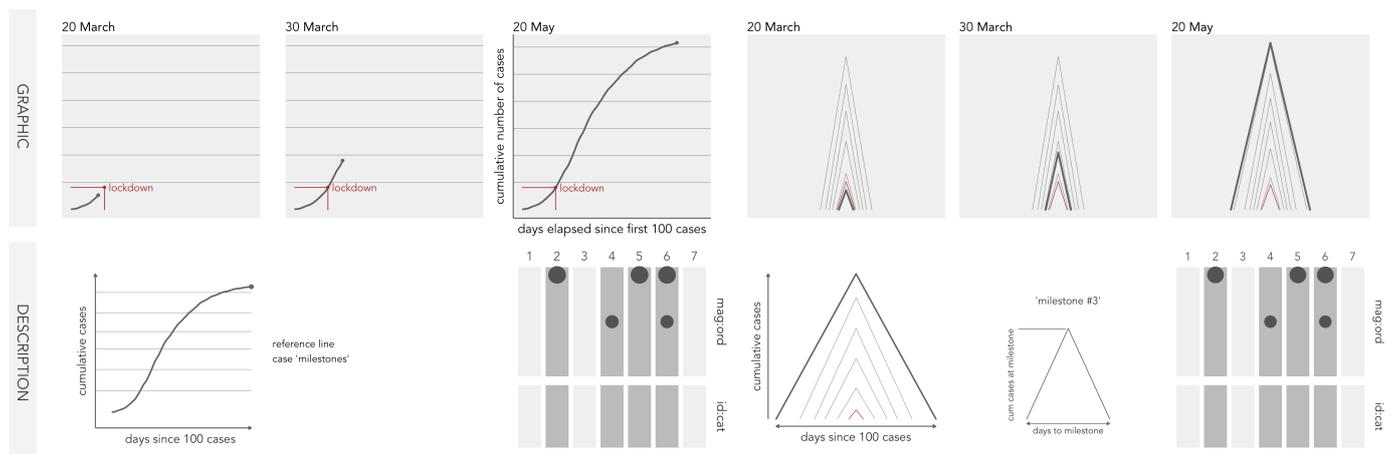


Figure 4. Example line and contour ridge encodings applied to data for the region of London, accompanied with descriptions via encoding schematics. Selected frames from an animation are displayed. An animated equivalent is available at the paper’s [github repository](#).

4.3. Geospatial Arrangements

There are several options for incorporating spatial context within our designs. In the top row of Figure 5 are three candidates: an ‘exact’ arrangement, with local authorities positioned according to their true geographic location (at local authority centroids), a continuous area cartogram layout with local authorities distorted according to population size [37]; and a semi-spatial ordering with local authorities of regular size and geometry (grid squares) but with an approximate spatial arrangement (e.g., [38,39]). Each is presented using the ridge contour design and encoding described in Figure 4 and with accompanying encoding schematics.

The ‘exact’ arrangement has the obvious benefit of being highly recognisable, enabling authorities to be easily located and perhaps regional grouping and comparisons to be more accurately and quickly performed. The obvious deficiency is the cluttering and occlusion in more densely populated parts of the country and particularly London—this layout clearly violates DesR2 **Discernible**. The cartogram is a substantial improvement, but does not entirely solve the problems of clutter and occlusion. Additionally, since aspects of population size are intrinsic to our data graphics (e.g., Figure 6) there is not a strong case to distort local authorities according to population size in the same way as one might an electoral cartogram of voting outcomes. The grid layout, a geospatial *small multiple with gaps (smwg)* [38], is the least recognisable. Departing from England’s familiar geometry means an additional overhead in terms of learning the layout. Additionally, due to the fact that regular squares are used, not all adjacency relationships between local authorities are preserved [38–41]. This fact is likely to interfere with judgements around spatial dependence, which as demonstrated by the Washington Post graphic [19], is an important factor when monitoring the spread and development of COVID-19. There are, though, several additional benefits conferred from the use of regularly-sized grids. Firstly, the grids allow more visually complex and detailed re-designs. In Figure 10, we use the full grid space to superimpose several chart types that encode directly cumulative cases and new daily reported cases, this greatly helps with meeting DesR1 **Concurrent**, though compromising DatR1 **geography**. Secondly, the grid layout may also better support positional judgements for the two key quantitative measures—time elapsed (horizontal position) and cumulative cases (vertical position). When glyphs are spatially arranged, this positional encoding is a secondary and inferior one (unaligned scale), as represented by the encoding schematics where the encoding for DatR2 **absolute number** and DatR5 **time elapsed** is bumped down one row. For the *smwg* grid layout, however, aligned comparisons of ridge heights is almost possible along the rows and, perhaps less obviously, of ridge widths along the columns. This slight difference is represented in the encoding schematic

by making the mark outlines for DatR2 and DatR5 light grey and positioning them on the row representing *position* (*common scale*). This elevates *smwgs* above the cartogram and ‘real’ geographies on the effectiveness of encoding it affords (*DesR4 Estimable*), and we opt for the *smwg* layout for the remainder of the paper.



Figure 5. Candidate geospatial arrangements for local authorities: left, arranged according to ‘real’ location and authorities sized according to physical geography; middle, physical geography is relaxed and authorities sized according to population using rubber sheet distortion algorithm [37]; right, authorities are of fixed size and spatially arranged using layout algorithm in Meulemans et al. [38]. Each is accompanied with an encoding schematic. The light-coloured dots for the *smwg* layout denote that the positional encoding of ridge width and heights is partially on an *aligned* scale. An animated map morphing between ‘real’ and *smwg* layouts is in the paper’s [github repository](#).

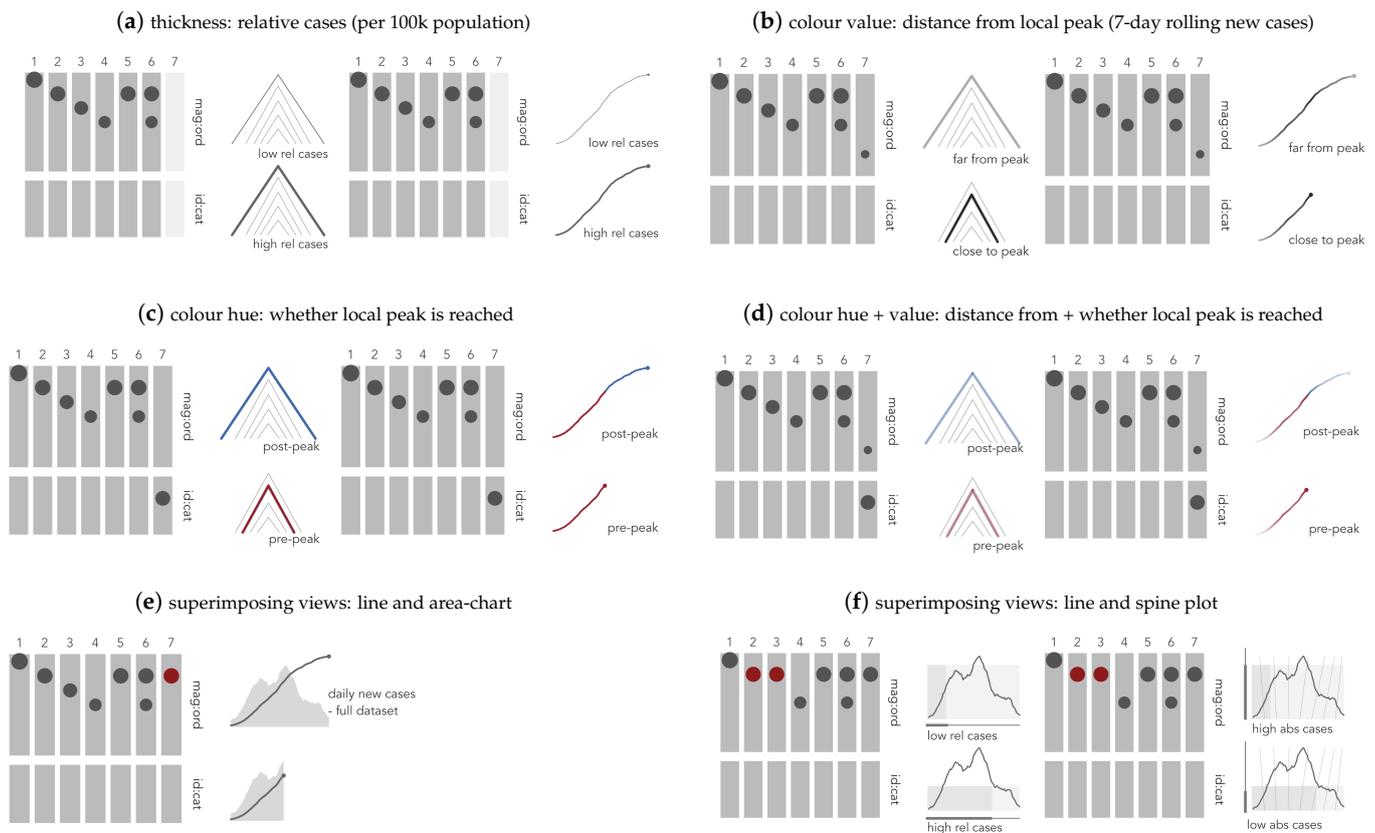


Figure 6. Design candidates for building data density. Ridge contour and line equivalents are represented with the same subset of data. Each design is accompanied by a schematic. Data density can be inferred by the number of filled columns, so from design Figure 6b all seven DatRs are represented. Encoding effectiveness can be inferred from the position and size of dots in each schematic—as per Munzner [24] larger dots higher in the matrix suggest greater encoding effectiveness.

4.4. Increasing Data Density

The designs presented in Figure 5 are already data dense, with position and orientation used to meet five of the seven DatRs. The graphics can be further parameterised in order to meet the remaining DatRs—DatR3 **relative number** and DatR7 **cases relative to ‘peak’**—and using the remaining visual channels identifiable from the empty rows in the encoding schematics.

Figure 6 presents several design candidates that address this challenge with additional encodings selected through considering the schematics. In each, DatR3 **relative number** of cases is encoded using line thickness—length (1D size) according to the schematic, and although less effective than position, an appropriate visual channel for communicating quantities. Figure 6b–d encode aspects of DatR7 **cases relative to ‘peak’** using colour. Here a local ‘peak’ is used, which is the point in time when the largest reported number of daily cases, expressed as a 7-day rolling average, is recorded. In Figure 6b distance in reported 7-day rolling cases to/from this peak is mapped to a continuous scale and encoded using colour value. The darker the colour, the closer the current point in the animation is to this peak—the bold line for the ridge contours and the most recent observation (and dot) for the lines. So in the two rows demonstrating two time slices in Figure 6b, the ridge in the first row is light grey (current new cases recorded are much less than the peak) and the ridge in the second row is dark grey (current new cases recorded are almost at the peak). Note that from the ordering in Figure 3, colour value is a less effective visual channel for encoding quantities and the encoding schematic reflects this—a small dot appears low down in the DatR7 column. In Figure 6c, DatR7 **cases relative to ‘peak’** is treated as a categorical variable and encoded using colour hue, an appropriate visual channel for

categorical data and reflected in the DatR7 column of the schematic. Where the current point in time is pre-peak, the ridge or line is coloured red, where it is post-peak, it is coloured blue. In Figure 6d, distance from the peak is mapped to a diverging scheme—colour hue to distinguish whether the current point in time is pre- or post-peak and colour value according to distance from this peak. Again, the encoding schematic is updated to reflect this with two dots now appearing in the DatR7 column.

Figure 6e is slightly different in that DatR7 **cases relative to ‘peak’** is encoded directly with the addition of a background area-chart of the daily new cases data, represented with a red dot in the schematic to denote an additional ‘layer’. This is an effective (positional) encoding, and for which there is not an obvious equivalent in the ridge graphics. However, this sort of ‘overloaded’ [42] view where two separate chart types (line and area charts) are superimposed on top of each other adds visual complexity. In Figure 6f we try to reduce this complexity by replacing the cumulative cases line with a spine plot [43] where height varies according to **absolute** number of cases (DatR2) and width according to **relative** number of cases (DatR3)—this is represented with red dots in the schematic. In the example, the local authority in the top row has recorded larger case counts than that in the bottom row, but relative to population size a much higher share of the population in the bottom row has been infected. A global scaling is applied here, so if the maximum infection rate across local authorities was 800 *per* 100k population, local authorities with an infection rate close to this would show a spine plot with a darker fill occupying almost the full horizontal width; if the maximum case count across local authorities was 10k, then local authorities with case counts close to this would show a spine plot that extends across a cell’s full vertical height.

Previous experiments with overloaded maps suggest that when presented as a full glyphmap these two views are likely to be perceived in concert [44]. A consequence of removing the cumulative cases line, however, is that the sense of aggregate-level growth, speed-up and slow-down, in cases over the observed time period is lost. We attempt to capture the essence of this, whilst not overwhelming the graphics in terms of visual complexity and saliency, by annotating the chart with faint milestones. Different from the ridge contours these milestones are sampled at regular time intervals—every 10 days since the first 100 cases are recorded. The encoding requires initially a little more interpretation: two horizontal axes are introduced, with time progressing along the bottom horizontal axis and cumulative case counts along the top axis. A local scaling is used and so when milestone lines are angled to the left (\) the rate at which cases are accumulating over time is *slower* than the period average; to the right (/) cases are accumulating *faster* than the period average; and vertical (|) cumulative cases are *in line with* the period average. These reference markers are added to the plot background and should be read concurrently with the daily cases area-chart. Since these milestones capture a sense of accumulating speed-up and slow-down, there is autocorrelation in the line angles; they become most effective when displayed as small multiples to support comparison across local authorities, as in the full glyphmap in Figure 10.

Each of the design candidates in Figure 6 are plausible given visualization design theory (as validated through the schematics), but moving through Figure 6a–f, the designs become increasingly visually complex. In Figure 6d, three data channels are mapped to the lines and ridges simultaneously (line thickness, colour hue and colour value) and in Figures 6e,f, two separate chart types (line and area charts) are superimposed on top of each other. Whether or not the distinct data items can genuinely be perceived when represented as full glyphmaps is difficult to establish empirically (c.f. [44,45]).

Absent from our designs is the use of 3D (depth). Whilst Munzner [24] gives depth a low effectiveness ordering (see Figure 3), there is precedent in Cartography for encoding temporal data items along a z-axis (e.g., space-time-cubes [15]). Since the temporal element is so important to the analysis of disease spread—the line graphics in particular follow characteristic shapes—there is not a strong case for exploring representations of time using depth (3D position). An interesting future activity may instead be to build interaction to support *rotation* of our designs in 3D space [46,47]. For example, estimation

of quantities encoded via height and width is necessarily hindered due to the 2D geospatial arrangement—the glyphs in our glyphmaps are *unaligned*. Selective and flexible rotation of the ridge graphics in particular may help with the perception of these quantities.

In the section that follows we qualitatively compare the ridge contours and lines, and their design candidates, and aim to make recommendations matching certain design configurations to data analysis needs.

5. Analysis

In this section, we present full glyphmaps of local authority case data based on the design candidates in Figure 6. To structure this discussion we identify three themes. *Overall case extent* (Section 5.1) is a sort of composite of DatR2 **absolute number** and elements of DatR6 **case history**—and that we use to describe the effectiveness of designs for representing the overall magnitude and collateral damage of the virus given the amount of time it has been established in a local authority. *Change and case history* (Section 5.2) is used to evaluate the effectiveness of designs at together representing DatR4 **rate of change**, DatR5 **time elapsed** and DatR6 **case history**. *Re-prioritising daily signatures and peaks* (Section 5.3) reflects on design configurations in Figure 6e,f that give greater emphasis to daily case counts and patterns around local peaks.

5.1. Overall ‘Case Extent’

When conceiving of the ridge contours, we felt they might provide a unique and efficient visual short-hand for conveying case extent—absolute number of cases relative to time that the virus has been present in a local authority. For example, ridge contours that are thin and short suggest the virus is *establishing* and *growing*, thin and tall that the virus is *established* but still *growing*, wide and tall that the virus is *established* but has probably peaked with case numbers *slowing*, wide and small that the virus may *not* be *established* and that case numbers may be *slowing* (Figure 7). Once the encoding is learnt, it is possible to identify some of these categories of virus extent particularly in the animated version of Figure 8, which applies the encoding described in Figure 6a as full glyphmaps. The symbolisation used in the ridges (emerging peaks) also draws attention to the height encoding (case frequency) and the density in the milestone contours is read almost pre-attentively, which again supports judgements around virus extent. The occlusion that the encoding introduces serves to emphasise local authorities that contain the largest case counts—Kent, Lancashire, Essex—at the expense of authorities positioned in rows immediately above those authorities. This sort of interference between data items could easily be avoided in our *smwg* layout by ensuring that ridges and lines do not extend beyond their own grid cells. However, we wish to provide additional emphasis to local authorities with particularly high case counts and so allowing ridges to encroach on neighbouring cells is a deliberate design decision. Since line thickness varies according to **relative number** of cases (DatR3), the occlusion effect is even greater for those local authorities with populations particularly exposed to the virus—in Lancashire and the north west for example. The same is nevertheless true of the more familiar line equivalent, where patterns in virus presence, and additional information on growth, can be read more directly from line heights and slopes.

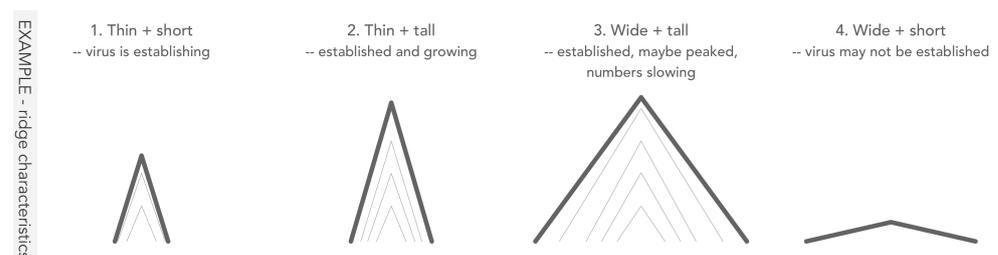


Figure 7. Four characteristic shapes that we envisaged being detected in our ridge contour graphics.

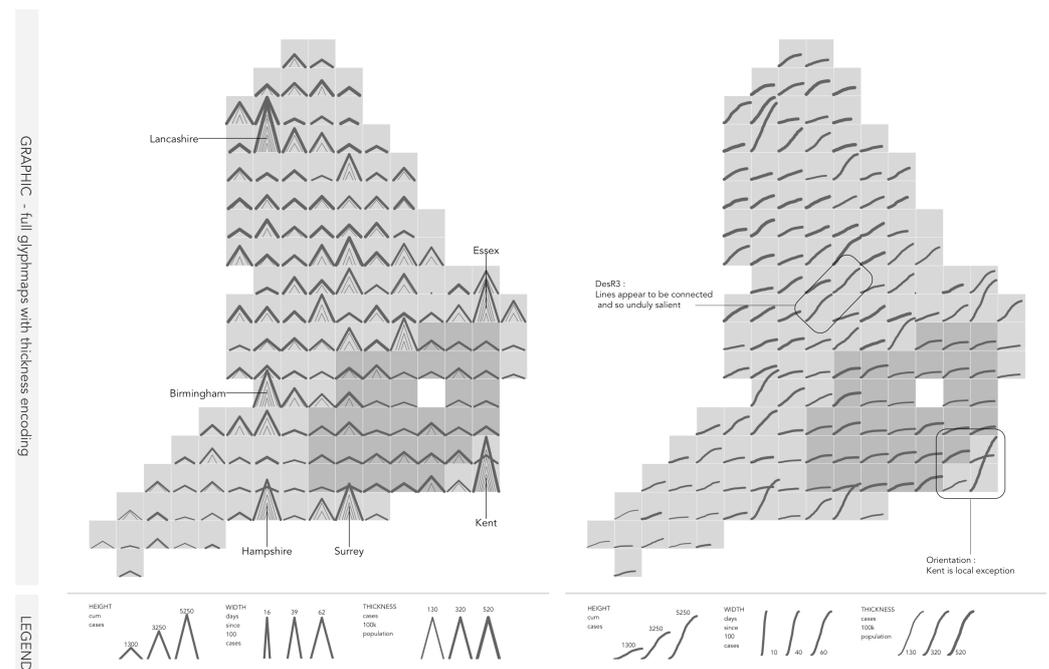


Figure 8. Full glyphmaps for ridge contours and lines, with thickness encoding applied as in Figure 6a. An animated equivalent is available at the paper’s github repository.

Considering the requirements identified in the introduction, not all data items are being shown **concurrently** (DesR1) in Figure 8 as there is no explicit encoding of **local ‘peaks’** (DatR7) in daily new cases. Individual marks in both the ridges and lines are **discernible** (DesR2) and the quantities encoded using position (time elapsed, cumulative cases) and size (line thickness for relative cases) are **estimable** (DesR4). However, visually scanning across the two graphics there are occasionally unhelpful (unintended) artefacts in the lines chart where, due to similarities in slopes between neighbouring local authorities (grid cells in the *smwg*), lines do not always look discrete (annotated in Figure 8). These appear unduly salient and therefore the line version risks violating DesR3 **prioritised**—phenomena that are important must appear visually salient. Given this, and the fact that the symbolisation of the ridges (emerging peaks) seems to give special priority to case counts, our recommendation is for the ridge contours over the lines when representing overall case extent in a spatial context.

5.2. Change and Case History

Different from the ridge contours, then, the lines communicate DatR4 **rate of change**, DatR5 **time elapsed** and DatR6 **case history** in a direct way. They result in characteristic shapes that are expressive and can be very quickly parsed. The pattern of reported cases having flattened is the dominant shape for local authorities in London and many in the south of the country (Figure 8) as of 1st June 2020, the last release of new cases data used in our graphics. This flattening happens later for local authorities in the midlands and north of England, with cases still rising reasonably sharply for many local authorities. Additionally, since there is spatial autocorrelation in case trajectories, with neighbouring local authorities sharing similar characteristic shapes, attention is drawn to where this is not the case—where case numbers are large and trajectories locally exceptional, for example Kent in the south east (annotated in Figure 8).

Colour is varied in the graphics to further encode case history—DatR7 **cases relative to local ‘peak’**. In Figure 6, we identify three encoding options: ‘distance from local peak’ (7-day moving average in new cases) represented on a continuous scale using colour value (Figure 6b), on a categorical scale using colour hue (Figure 6c) and on both a continuous and categorical scale using a diverging scheme (Figure 6d). In Figure 9, each of these is

demonstrated for the ridge contour charts. Animating over variations in colour of the ridges helps draw attention to spatial/regional aspects of change in new case numbers—the shifting of local peaks around the country in a similar way to Colin Angus’s animated choropleth [7]. However, processing this sort of compound visualization [45], in which thickness, shape (height+width) and colour are varied simultaneously and then animated, is challenging. Where the colour encoding is applied to the ridge marks themselves, change is sometimes difficult to detect—the graphics in Figure 9 may risk breaking DesR4 **estimable**. Applying the colour encoding to the background spatial units (in this case the *smwg* grid squares demonstrated in Figure 9) may help. A distracting ‘flicker’ effect is nevertheless introduced into the animation and it may be that this graphical formulation is beyond most perceptual capabilities. In the line version, the additional encoding of colour gives greater emphasis to the duration of local ‘growth’ periods—long sections of red reinforcing periods of sustained growth in cases. This works reasonably effectively for static visualizations. Unlike with the ridges there is not a strong case (or need) for animating over the coloured lines as little additional emphasis is introduced via the animation. The main effect when animating the ridges—that of regionally shifting peaks—is less easy to detect when animating the lines. If geographic patterns in change and case history is of principal concern, we advocate using the lines over the ridges, optionally colouring lines according to a data property of interest (e.g., ‘distance from local peaks’).

5.3. Re-prioritising Daily Signatures

Our two main designs, the lines and ridge contours, and also our seven DatRs, were heavily informed by COVID-19 glyphmaps released immediately after lockdown—the point at which the UK government imposed restrictions on movement and social interaction—that were particularly concerned with conveying absolute numbers. Of principal interest here was where case counts were greatest, and secondly, the pattern of exponential growth in these areas. Whilst the geography of absolute and relative case counts remains important, displaying daily case counts explicitly enables detailed information around local virus presence—the timing, duration and character of local peaks—becomes increasingly relevant to the more geographically targeted restrictions that characterised subsequent and most likely future waves of infection.

Figure 10 illustrates this alternative emphasis. A formal description of the encoding is in Section 4.4, but all seven DatRs are variously encoded with greater priority placed on DatR7 **cases relative to local ‘peak’**, achieved by representing daily new case counts (with 7-day smoothing) directly via an area-chart. DatR2 **absolute number** and DatR3 **relative number** of cases are communicated via a spine chart, which appears in the background of the *smwg* grid cells. Furthermore, are the reference milestones clarifying whether local case trajectories are slowing down (\) or speeding up (/). From this it is possible to identify the early peaks and then steep drop-off in new cases after lockdown in the area charts for local authorities in London (annotated). When visually scanning the area-charts and the spine plots simultaneously, there is also a sense that the duration of peaks is greater for local authorities in the north west where the largest relative number of cases is reported later and where peaks in daily new case counts are more prolonged.

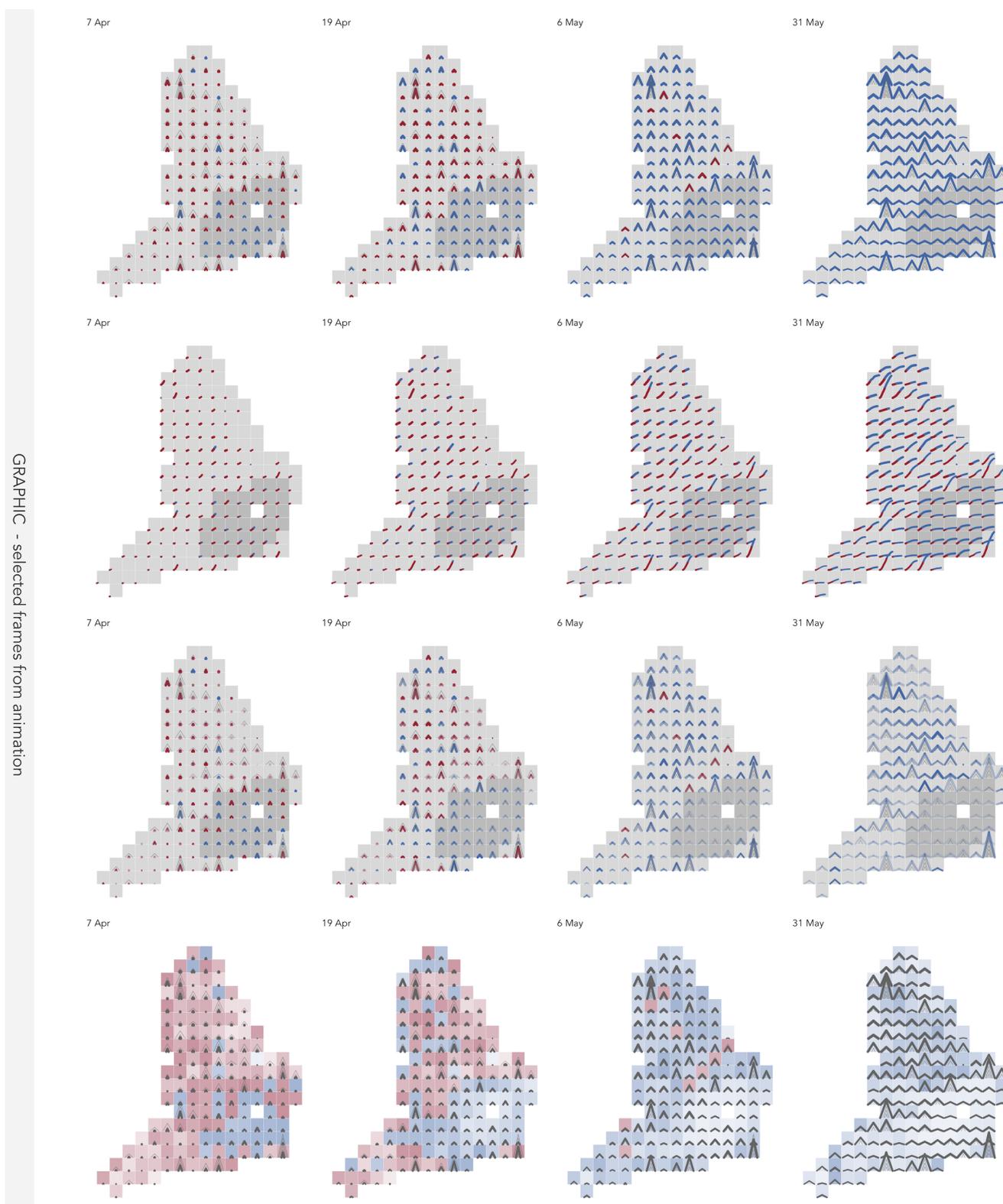


Figure 9. Frames from animated glyphmaps: in the top two rows the thickness and colour hue encoding is applied to ridges and lines, respectively, as in Figure 6c; in the bottom two rows the thickness, colour hue *and* colour value (lightness) encoding is applied to the ridges, as in Figure 6d, but in the bottom row area backgrounds are varied rather the ridge marks themselves. An animated equivalent is available at the paper's [github repository](#).

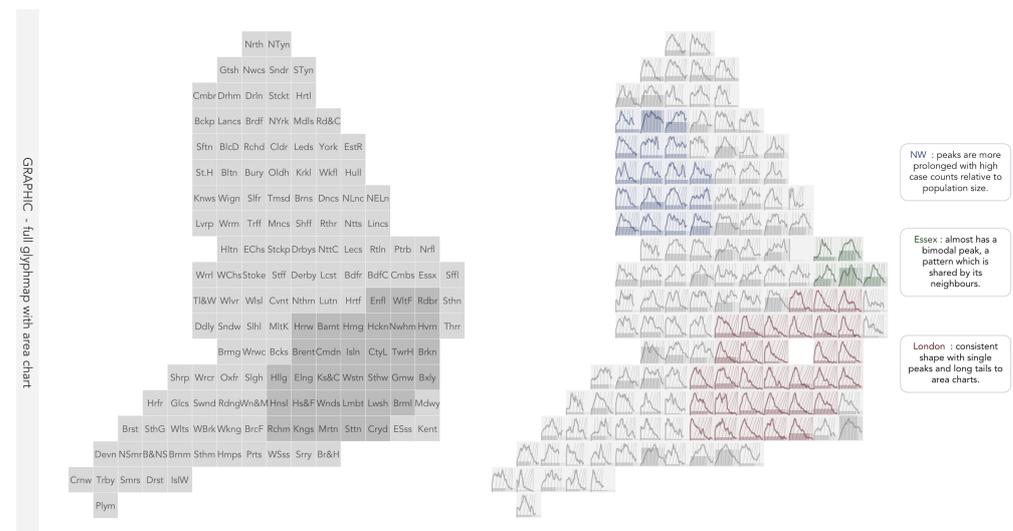


Figure 10. Full glyphmap with area-chart of daily new cases (with 7-day smoothing) and spine plot of absolute and relative cases superimposed. An animated equivalent is available at the paper’s [github repository](#).

An important difference in Figure 10 compared with the other design candidates is that, rather than simply parameterising existing ridge and line marks using colour or thickness, two distinct chart layers are introduced—lines and area-charts. It is very difficult to establish empirically whether or not these sorts of *overloaded* views [42] can genuinely be perceived simultaneously (c.f. [31,44]). The 1D structure in the area-charts of daily cases data is in itself reasonably rich. However, from visually examining the figure, we can make the case for geographically arranging even the more visually complex area-charts: that doing so reveals patterns in the **geography** (DatR1) of virus presence and history that would be very difficult to discern using non-visual means, simply from scanning across the shapes of the area-charts.

One way of demonstrating this is to generate a map line-up [48,49] of decoy area charts, whereby the observed data for each local authority are randomly permuted around locations in the *smwg*, and inserting the ‘real’ dataset amongst these decoys. This map line-up is presented in Figure 11; a slight update from Figure 10 is that a global start point for each local authority (7 March 2020) is used rather than a locally-varying start (from when the first 100 cases were reported in each authority). Not only does this demonstrate the strong uni-modal, right skew in the daily cases data for local authorities in London, but that there is also regional autocorrelation in the timings of peaks in the daily cases data (the 1D pattern of area-chart peaks) and to an extent the nature and duration of peaks (the 1D shape of the area-charts).

The alternative emphasis provided by Figure 10 is important. It conveys useful detail around the **rate of change**, **history** and nature of **local growth** trajectories (DatR4–DatR7), but also **absolute** (DatR2) and **relative** (DatR3) exposure, that is prescient at a time when targeted responses to re-infections at particular local authorities and regions are to be evaluated. The graphic seems to comply with the four Design Requirements and, according to the accompanying schematic in Figure 6f, uses an effective encoding. That spatial autocorrelation structure can be discerned even in the detailed 1D distribution of daily cases data (as demonstrated in Figure 11) further validates our approach of representing detailed case trajectories with a geographical arrangement. Clearly, though, this design is only feasible where the spatial units are regularly sized grid cells, as in our *smwg*, and sufficient space is available for a legible overloaded glyph in each cell; it would not transfer well to the US county data and the larger numbers of geographic areas used in the Washington Post graphic [19].

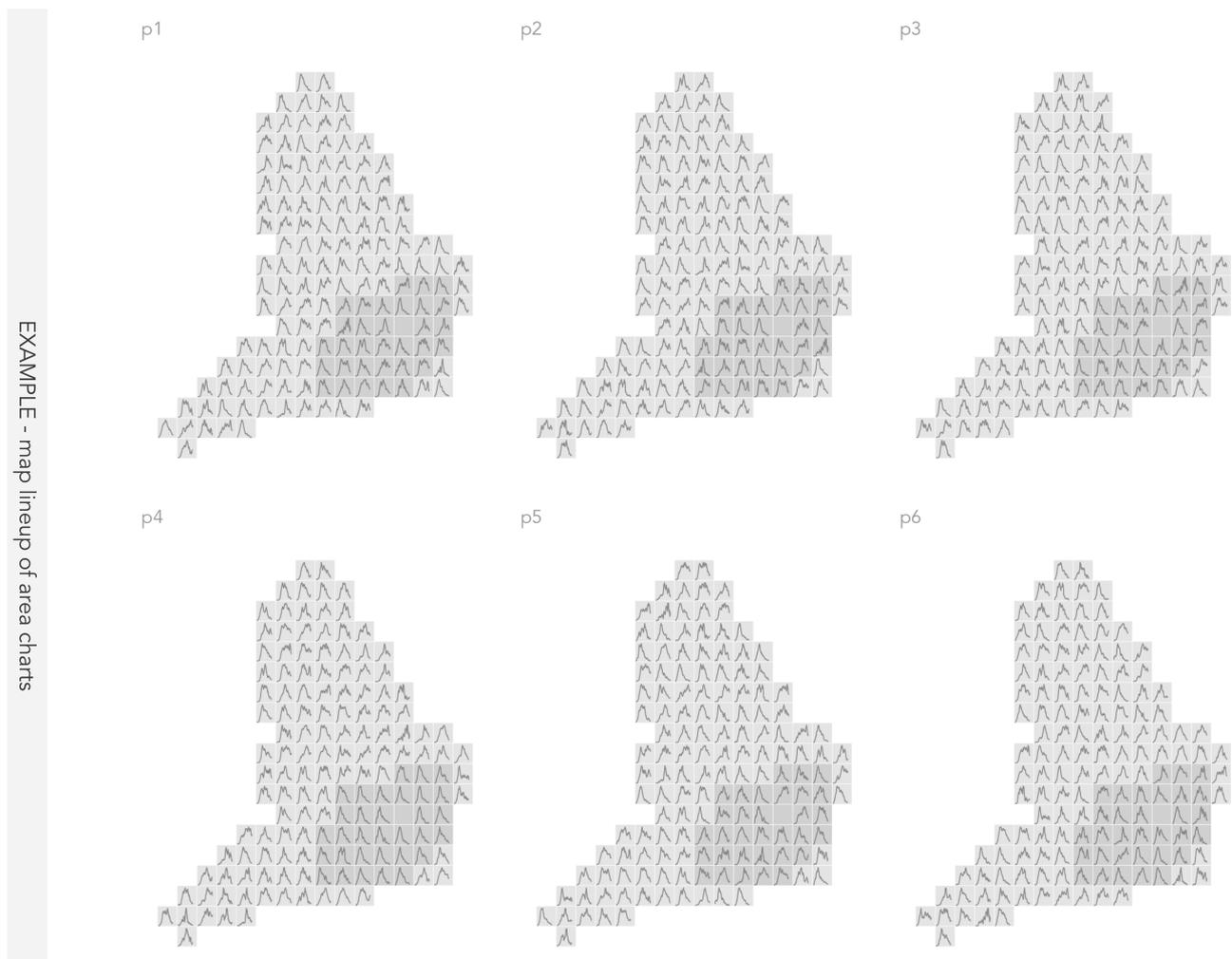


Figure 11. A map line-up test [48,49] of daily cases area-charts in which the ‘real’ dataset (p4) is presented alongside five decoy plots generated by randomly permuting the observed cases data around local authorities.

6. Conclusions

This paper adds both to applied and cartographic [17,50] literature analysing COVID-19 and its spread and, more generally, to approaches in geovisualization aimed at visually analysing multivariate geospatial structure. That datasets related to the pandemic have been shared rapidly and widely, with analysts focussed on these same data and challenges in an intense way, presents a unique opportunity for learning and transfer between research disciplines. We contribute to this effort in a number of ways.

First, we provide a review of the rapidly developing visualization efforts analysing the spread of COVID-19 over geographic areas, with a special focus on glyphmaps. The ambition in this visualization activity was to encode many data items within geospatial context and from this we abstract several data and design requirements necessary for analysing the extent, magnitude and nature of the virus’s spread. Arranging spatial units according to a relaxed geospatial layout gives us the space to show multivariate graphics as interpretable glyphmaps. We used candidate layouts and glyph designs to simultaneously convey the temporal trend in COVID-19 cases, locate and compare peaks in different parts of the country, to summarise details around temporal trajectories (multi-modal peaks) and to encode information on absolute and relative case counts.

Second, are the *encoding schematics*. Visualization design is ultimately a process of decision-making; after carefully abstracting, organising and prioritising data and design requirements, decisions must be made about how best to leverage visual systems (visual marks and channels [24]). This becomes challenging in geovisualization where multi-

ple data items must be incorporated within a geospatial arrangement, and the *encoding schematics* usefully provide a menu of data and visual encoding channels and therefore a framework for navigating the design space for our candidate glyphmap designs. That the encoding schematics are underpinned by an empirically-informed encoding hierarchy means they are a quick and low-cost way of comparing across and validating the multiple candidate designs.

A third substantive contribution is the designs themselves. We present a range of candidate designs that expose spatiotemporal structure in virus extent and history in England with varying success. This work serves to highlight the substantial spatial differences in both the speed and the magnitude of infection across different local authority areas. The designs are described, validated and critiqued in a structured way and in light of established principles, using the encoding schematics. We hope that these designs and this discussion may transfer—supporting COVID-19 visualization design in other geospatial contexts or for visual analysis of other epidemiological phenomena. That the designs are implemented using a high-level visualization grammar (`ggplot2`), and accompanied with a code repository, may foster this activity.

Finally, we make targeted recommendations linking candidate designs to data display and analysis needs. We do not offer a single design recommendation and it may be necessary to flexibly vary designs according to data analysis needs—to prioritise and bring certain data items into focus, a theme that we have previously identified for multivariate geovisualization [44]. Through our design exposition and discussion, we do however present a process for guiding visualization design leading to novel data graphics that successfully characterise detailed structure in the daily cases data for UK local authorities, and that reveals a striking geography. This activity seems prescient even to the current moment in the pandemic, enabling us to see the varying spatiotemporal character of infection and to design appropriate visualizations as data and needs evolve.

Author Contributions: Conceptualization, Roger Beecham, Jason Dykes, Layik Hama, Nik Lomax; methodology, Roger Beecham, Jason Dykes, Layik Hama, Nik Lomax; software, Roger Beecham; validation, Roger Beecham, Jason Dykes, Layik Hama, Nik Lomax; formal analysis, Roger Beecham, Jason Dykes, Layik Hama, Nik Lomax; investigation, Not Applicable; resources, Not Applicable; data curation, Roger Beecham; writing—original draft preparation, Roger Beecham, Jason Dykes, Layik Hama, Nik Lomax; writing—review and editing, Roger Beecham, Jason Dykes, Layik Hama, Nik Lomax; visualization, Roger Beecham, Jason Dykes, Layik Hama, Nik Lomax; supervision, Not Applicable; project administration, Roger Beecham, Nik Lomax; funding acquisition, Roger Beecham, Nik Lomax. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the ‘Digital Twins: Urban Analytics’ theme within that grant & The Alan Turing Institute.

Data Availability Statement: A repository containing data, code and further discussion behind our work is at <https://github.com/rogerbeecham/covid-19-datavis> (accessed on 24 March 2021).

Acknowledgments: Discussions with, and inputs from, the following colleagues have shaped this work: Chris Rooney, Genetec; Aidan Slingsby, University of London; Jo Wood, University of London.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PHE	Public Health England
ONS	Office for National Statistics
DatR	Data Requirement
DesR	Design Requirement.

References

- Vasiliev, I. Mapping time. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **1997**, *34*, 1–51. [CrossRef]
- Andrienko, N.; Andrienko, G.; Gatalsky, P. Exploratory spatio-temporal visualization: An analytical review. *J. Vis. Lang. Comput.* **2003**, *14*, 503–541. [CrossRef]
- Bach, B.; Dragicevic, P.; Archambault, D.; Hurter, C.; Carpendale, S. A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes. *Comput. Graph. Forum* **2017**, *36*, 36–61. [CrossRef]
- Centre for Systems Science and Engineering. COVID-19 Data Repository. 2020. Available online: <https://github.com/CSSEGISandData/COVID-19> (accessed on 20 March 2021).
- Paszto, V.; Burian, J.; Macku, K. COVID-19 data sources: Evaluation of map applications and analysis of behavior changes in Europe’s population. *Geografie* **2020**, *125*, 171–209. [CrossRef]
- Rajerison, R. Geosparklines: A Twitter Map Story. 2020. Available online: <https://datagistips.hypotheses.org/514> (accessed on 20 March 2021).
- Angus, C. CoVid Plots and Analysis. 2020. Available online: https://figshare.shef.ac.uk/articles/dataset/CoVid_Plots_and_Analysis/12328226 (accessed on 20 March 2021).
- UK Government. Coronavirus: Downing Street Briefing. 2020. Available online: <https://youtu.be/WXQlg2cR1Wo?t=363> (accessed on 9 October 2020).
- Harrower, M.; Fabrikant, S. The role of map animation in geographic visualization. In *Geographic Visualization: Concepts, Tools and Applications*; Dodge, M., McDerby, M., Eds.; John Wiley & Sons: Chichester, UK, 2008; Chapter 5, pp. 49–65.
- Škop, M. Map-Sparks. 2020. Available online: <https://observablehq.com/@michalskop/map-sparks> (accessed on 9 October 2020).
- Andrienko, N.; Andrienko, G. (Eds.) *Interactive Visual Tools to Explore Spatio-Temporal Variation*; AVI '04; Association for Computing Machinery: New York, NY, USA, 2004; Volume 240.
- Wickham, H.; Hofmann, H.; Wickham, C.; Cook, D. Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics* **2012**, *23*, 382–393. [CrossRef]
- Wickham, H. A layered grammar of graphics. *J. Comput. Graph. Stat.* **2010**, *19*, 3–28. [CrossRef]
- Monmonier, M. Strategies For The Visualization Of Geographic Time-Series Data. *Cartogr. Int. J. Geogr. Inf. Geovisualization* **1990**, *27*, 30–45. [CrossRef]
- Kraak, M. The space-time cube revisited from a geovisualization perspective. In *Proceedings of the ICC 2003: Proceedings of the 21st International Cartographic Conference—ICC 2003, Durban, South Africa, 10–16 August 2003*; International Cartographic Association: Wellington, New Zealand, 2003; pp. 1988–1996.
- Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf. Dis.* **2020**, *20*, 533–534. [CrossRef]
- Juergens, C. Trustworthy COVID-19 Mapping: Geo-spatial Data Literacy Aspects of Choropleth Maps. *J. Cartogr. Geogr. Inf.* **2020**, *70*, 155–161. [CrossRef]
- Fieldsend, H. Watch COVID-19 Spread throughout England-Resource Centre: Esri UK & Ireland. 2020. Available online: <https://resource.esriuk.com/blog/watch-covid-19-spread-throughout-england/> (accessed on 9 October 2020).
- Thebault, R.; Hauslohner, A. COVID-19’s Deadly New Surge Is in Rural America as States Re-Open-The Washington Post. 2020. Available online: <https://www.washingtonpost.com/nation/2020/05/24/coronavirus-rural-america-outbreaks/?arc404=true> (accessed on 9 October 2020).
- Legendre, P. Spatial autocorrelation: Trouble or new paradigm? *Ecology* **1993**, *74*, 1659–1673. [CrossRef]
- Healy, K. *Covdata: COVID-19 Case and Mortality Time Series*; R Package Version 0.5.2; 2020. Available online: <https://kjhealy.github.io/covdata/> (accessed on 20 March 2021).
- van Wijk, J. Views on Visualization. *IEEE Trans. Vis. Comput. Graph.* **2006**, *12*, 421–32. [CrossRef]
- Munzner, T. A Nested Process Model for Visualization Design and Validation. *IEEE Trans. Vis. Comput. Graph.* **2009**, *15*, 921–928. [CrossRef]
- Munzner, T. *Visualization Analysis and Design*; CRC Press: Boca Raton, FL, USA, 2014.
- Talbot, J.; Setlur, V.; Anand, A. Four Experiments on the Perception of Bar Charts. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 2152–2160. [CrossRef]
- Peraya, V.; Pietriga, E.; Bezerianos, A. A comparison of visualizations for identifying correlation over space and time. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 375–385.
- Public Health England. Coronavirus (COVID-19) in the UK. 2020. Available online: <https://coronavirus.data.gov.uk/> (accessed on 20 March 2021).
- Office for National Statistics. ONS Open Geography Portal. Available online: <https://geoportal.statistics.gov.uk/> (accessed on 20 March 2021).
- Office for National Statistics. Deaths Involving COVID-19 by Local Area and Deprivation. Available online: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsinvolvingcovid19bylocalareaanddeprivation> (accessed on 20 March 2021).
- Wood, J.; Kachkaev, A.; Dykes, J. Design Exposition with Literate Visualization. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 759–768. [CrossRef]

31. Beecham, R.; Dykes, J.; Rooney, C.; Wong, W. Design Exposition Discussion Documents for Rich Design Discourse in Applied Visualization. *IEEE Trans. Vis. Comput. Graph.* **2020**. [[CrossRef](#)]
32. Roth, R.E. Visual Variables. In *International Encyclopedia of Geography*; American Cancer Society: Atlanta, GA, USA, 2017; pp. 1–11.
33. Cleveland, W.; McGill, R. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Am. Stat. Assoc.* **1984**, *79*, 531–554. [[CrossRef](#)]
34. Heer, J.; Bostock, M. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI '10, Atlanta, GA, USA, 10–15 April 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 203–212.
35. Brewer, C.; Campbell, A.J. Beyond Graduated Circles: Varied Point Symbols for Representing Quantitative Data on Maps. *Cartogr. Perspect.* **1998**, *29*, 6–25. [[CrossRef](#)]
36. Harrower, M.A. Visual Benchmarks: Representing Geographic Change with Map Animation. Ph.D. Thesis, The Pennsylvania State University, State College, PA, USA, 2002.
37. Dougenik, J.A.; Chrisman, N.R.; Niemeyer, D.R. An Algorithm To Construct Continuous Area Cartograms. *Prof. Geogr.* **1985**, *37*, 75–81. [[CrossRef](#)]
38. Meulemans, W.; Dykes, J.; Slingsby, A.; Turkay, C.; Wood, J. Small Multiples with Gaps. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 381–390. [[CrossRef](#)]
39. Meulemans, W.; Sondag, M.; Speckmann, B. A Simple Pipeline for Coherent Grid Maps. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 1236–1246 [[CrossRef](#)]
40. Eppstein, D.; van Kreveld, M.; Speckmann, B.; Staals, F. Improved grid map layout by point set matching. *Int. J. Comput. Geom. Appl.* **2015**, *25*, 101–122. [[CrossRef](#)]
41. Meulemans, W.; Sondag, M.; Speckmann, B. Generating tile maps. *Comput. Graph. Forum* **2017**, *36*, 435–445.
42. Elmqvist, N.; Fekete, J. Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines. *IEEE Trans. Vis. Comp. Graph.* **2010**, *16*, 439–454. [[CrossRef](#)]
43. Hummel, J. Linked bar charts: Analysing categorical data graphically. *Comput. Stat.* **1996**, *11*, 23–33.
44. Beecham, R.; Rooney, C.; Meier, S.; Dykes, J.; Slingsby, A.; Turkay, C.; Wong, W. Faceted Views of Varying Emphasis (FaVVEs): A framework for visualising multi-perspective small multiples. *Comput. Graph. Forum* **2016**, *35*, 241–249. [[CrossRef](#)]
45. Szafir, D.A.; Haroz, S.; Gleicher, M.; Franconeri, S. Four types of ensemble coding in data visualizations. *J. Vis.* **2016**, *16*, 11. [[CrossRef](#)] [[PubMed](#)]
46. Knust, C.; Buchroithner, M. Principles and Terminology of True-3D Geovisualisation. *Cartogr. J.* **2014**, *51*, 191–202. [[CrossRef](#)]
47. Edler, D.; Huber, O.; Knust, C.; Buchroithner, M.; Dickmann, F. Spreading Map Information over Different Depth Layers—An Improvement for Map-Reading Efficiency? *Cartogr. Int. J. Geogr. Inf. Geovisualization* **2014**, *49*, 153–163. [[CrossRef](#)]
48. Beecham, R.; Dykes, J.; Meulemans, W.; Slingsby, A.; Turkay, C.; Wood, J. Map LineUps: Effects of Spatial Structure on Graphical Inference. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 391–400. [[CrossRef](#)]
49. Wickham, H.; Cook, D.; Hofmann, H.; Buja, A. Graphical Inference for Infovis. *IEEE Trans. Vis. Comput. Graph.* **2010**, *16*, 973–979. [[CrossRef](#)]
50. Kent, A. Mapping and Counter-Mapping COVID-19: From Crisis to Cartocracy. *Cartogr. J.* **2020**, *57*, 187–195. [[CrossRef](#)]