

9. HORIZON CHARTS: DRAWING INSIGHT FROM LARGE TIME-SERIES DATASETS

9.1 Introduction

This paper provides an overview of a data visualisation technique called horizon charts. The approach, originally devised by data visualisation software company [Panopticon](#),¹ has been around approximately five years but is not widely applied. The horizon chart can appear complex at first sight, but this paper explores the purpose of the approach and describes the scenarios where it might be a valuable tool for drawing insight from large time-series datasets.

In summary, the horizon chart is a variation on the area chart, modified to deal with both positive and negative values. Rather than presenting negative values beneath the x-axis, the negative area is mirrored on to the positive side and then coloured differently to indicate its negative polarity. The result is a chart that occupies a single row of space, which helps to accommodate multiple stories onto a single display and facilitates comparison to pick out local and global patterns of change over time.

This paper draws heavily from a previous review of the horizon chart approach by Stephen Few.² His paper provided an extremely comprehensive assessment of the general merits of this technique, and the primary purpose of this paper is to consider the technique in a local application, specifically health data for the West Midlands region. Specifically, this paper illustrates how these charts can be used to identify unusual behaviours within large time-series datasets and understand patterns amongst many cases over time and the paper goes on to consider both the strengths and weaknesses of the approach, and highlight where it is an appropriate methodology to employ.

9.2 Methodology

Horizon charts can be overwhelming at first glance, but once the underlying principles and construction have been understood the simplicity of the approach and ease of interpretation can be appreciated. Figure 1 presents an example of the horizon chart.

The remainder of this section describes the process undertaken to produce the chart presented in Figure 9.1. It uses actual local health data, the rate of unplanned hospitalisations per 100,000 for chronic ambulatory care sensitive conditions (adults) across districts/boroughs and unitary areas within West Midlands Region (see References for a link to the raw data). Data with similar attributes can be used. Horizon charts are ideal to display time series data with a relatively large number of comparable cases, the example presented here is particularly useful at illustrating the nuances of the technique.

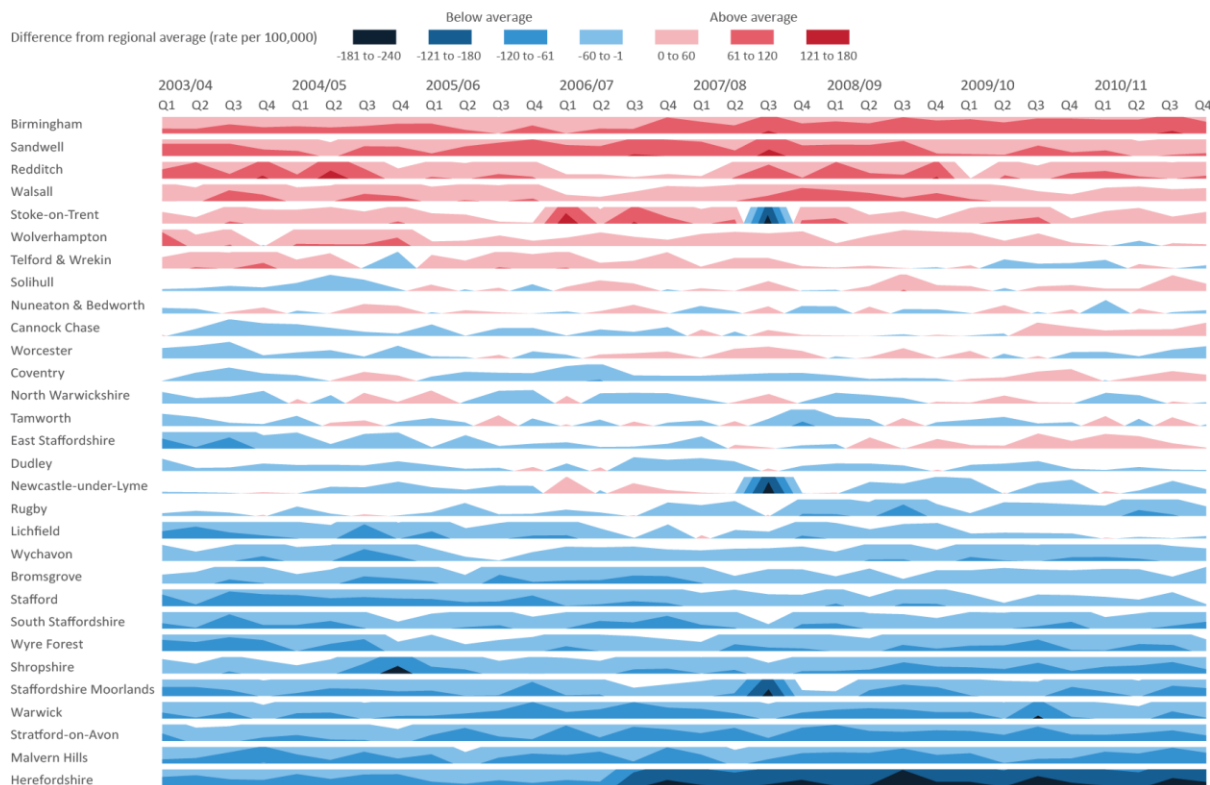
The technique demonstrated here can be created using an add-in developed for Microsoft Excel called Sparklines.³ This add-in is relatively straightforward to use, once the parameters are understood. Once the formula is established for one case (e.g. a local authority area) it can be simply copied and pasted for all other cases. A chart like the ones presented in this paper could be constructed within half an hour.

However, the stages displayed in this illustration have been created by hand using Excel and Adobe Illustrator to help explain the methodology in more detail. For this exercise we are using the West Midlands data for unplanned hospitalisation for chronic ambulatory care sensitive conditions which are published as indicator 2.3 of the NHS Outcomes Framework.⁶ An extract of the data is presented

in Figure 9.2. The data are arranged into rows of cases, which in this example are the local authority areas, and columns of dates, in this case years disaggregated into quarters.

Figure 9.1: Example horizon chart using West Midlands data for unplanned hospitalisation for chronic ambulatory care sensitive conditions

Unplanned hospitalisation for chronic ambulatory care sensitive conditions (adults), 2003/04 - 2010/11



Source: NHS Information Centre for health and social care <https://indicators.ic.nhs.uk>

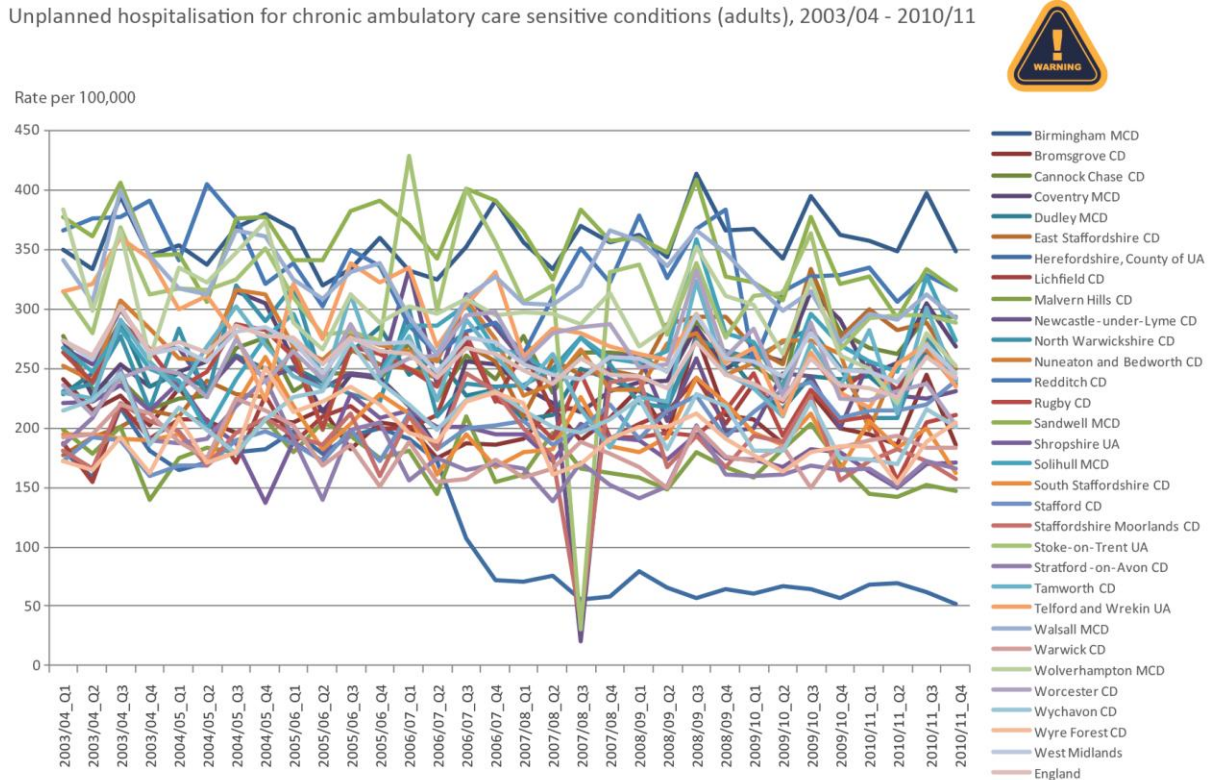
Figure 9.2: Source data in Excel spreadsheet format

Org Code	ONS Code	Organisation	2003/04_Q1	2003/04_Q2	2003/04_Q3	2003/04_Q4	2004/05_Q1	2004/05_Q2	2004/05_Q3	2004/05_Q4	2005/06_Q1	2005/06_Q2
15 00CN	E0800025	Birmingham MCD	350.7	333.8	395	344.6	353.6	337.8	368.4	379.8	368.2	319
16 47UB	E0700234	Bromsgrove CD	241.3	215.3	227.6	201.6	233.9	205.5	190	206.8	205	219
17 41UB	E0700192	Cannock Chase CD	277.3	231.9	241.3	216.2	225.1	226.7	266.8	276	230.8	2
18 00CQ	E0800026	Coventry MCD	273.8	227.8	253.8	234.8	246.8	256.6	314.7	304.5	256.8	21
19 00CR	E0800027	Dudley MCD	229.1	246.5	289.5	235.4	246.1	233.5	261.3	249.5	251.1	23
20 41UC	E0700193	East Staffordshire CD	177.1	192.5	201.5	214.3	209.7	240.3	228.8	226.3	268.3	19
21 00GA	E0600019	Herefordshire, County of UA	180.6	158.5	217.8	180.6	164.4	170.8	179.8	181.7	197.2	17
22 41UD	E0700194	Lichfield CD	187	154	220	191.2	207	207	170.7	227.1	186.8	205
23 47UC	E0700235	Malvern Hills CD	198.5	177.9	201.4	139.7	174.7	183.4	179	206.7	179.1	204
24 41UE	E0700195	Newcastle-under-Lyme CD	267.6	253	299.9	266.6	267.9	232	259.7	242.2	250.4	233
25 44UB	E0700218	North Warwickshire CD	231.3	237.9	277.2	216.1	283.5	225.5	320.1	290	319.6	239
26 44UC	E0700219	Nuneaton and Bedworth CD	252.8	240	307.3	284.1	258.4	257.2	315.8	312.3	268.1	257
27 47UD	E0700236	Redditch CD	366.4	376.1	378	391.6	341.7	404.7	376.1	320.7	338.3	302
28 44UD	E0700220	Rugby CD	263.2	238.2	289.7	266.2	233.2	247.8	287.1	282.7	259.8	19
29 00CS	E0800028	Sandwell MCD	377.6	361.3	406.3	345.5	345.9	306	376.9	378.3	341.7	34
30 00GG	E0600051	Shropshire UA	221.3	222.1	235.5	214.6	238.2	204.8	186.6	137.1	187.9	

It might be possible to make some basic observations by means of a simple scan of the data. However, we have 1,024 values to consider (32 areas x 32 points in time), and it is obviously difficult to understand the full complexity of the story this way. So, we naturally consider more visual ways to explore the data.

Where the interest is in examining trends over time and comparing areas with each other, a line chart would be a typical approach. Running the data through Excel and producing a default line chart provides us with Figure 9.3.

Figure 9.3: The default line chart



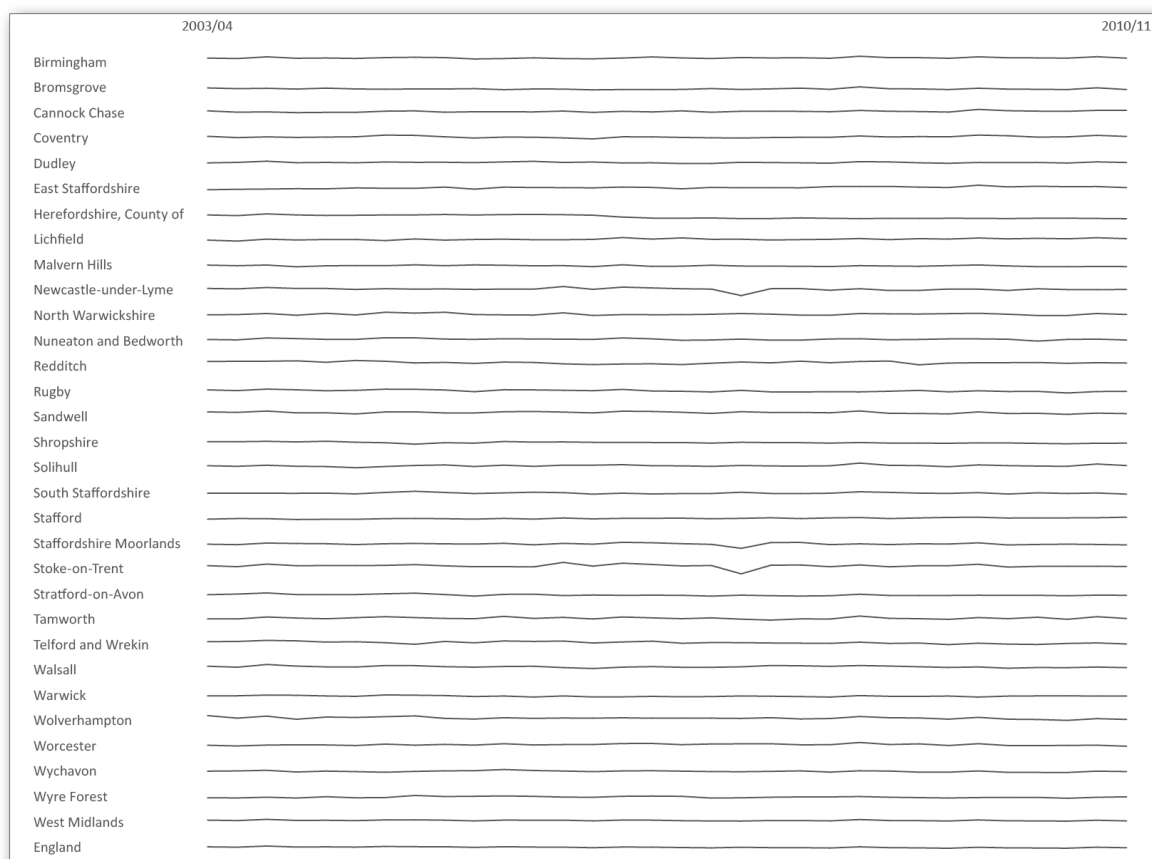
Source: NHS Information Centre for health and social care <https://indicators.ic.nhs.uk>

Clearly this chart is difficult to decipher and is of limited value. All we can say is that the figures tend to range between 150 and 400 per 100,000 and fluctuate year on year. There are some outliers that catch the eye, but it is very hard to identify which areas these relate to given the proximity of the lines and the limited number of colours that can be used.

It is virtually impossible to follow a specific area over time, identify which areas the lowest and highest rates relate to or compare an individual area to the regional average. This is where the Horizon Chart can add value, and we now go on to explore the process behind producing these diagrams.

First, in order to try and understand more easily what is happening in each area, we separate the data into individual line charts (Figure 9.4). This method is sometimes referred to as 'small multiples', a label usually attributed to Edward Tufte.^{4,5} As we can see, 32 line charts arranged in rows are not ideal. It is difficult to pick out key trends and there is still no way to examine the relative experience of each area compared to, for example, the average across the whole region.

Figure 9.4: Line charts for all areas



The available space does not allow us to produce individual charts with enough height to understand the changes taking place; the lines are too flat. How can we present and explore the data more effectively, in a way that helps us see patterns more clearly?

Let us examine an individual area in more detail; Figure 9.5 presents the data for Solihull.

Figure 9.5: Data for an individual area

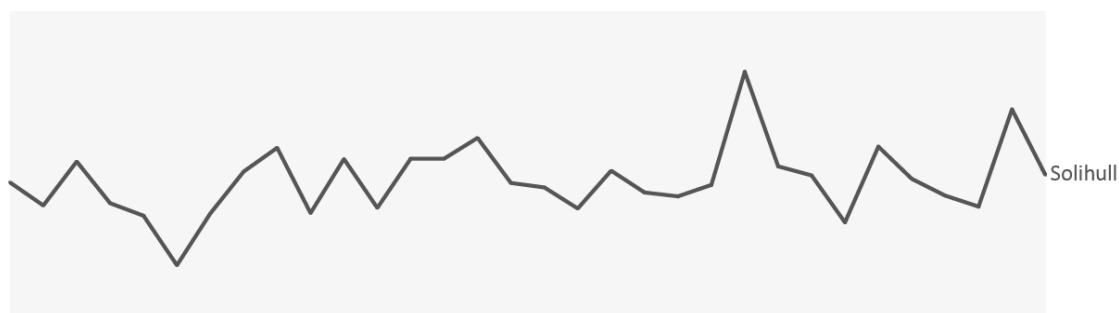


Figure 9.6 compares the figures in Solihull with the average for the West Midlands region. We start to see how local figures can vary from the average, but again 32 of these would still be incredibly hard to interpret.

Figure 9.6: An individual area compared with the regional average

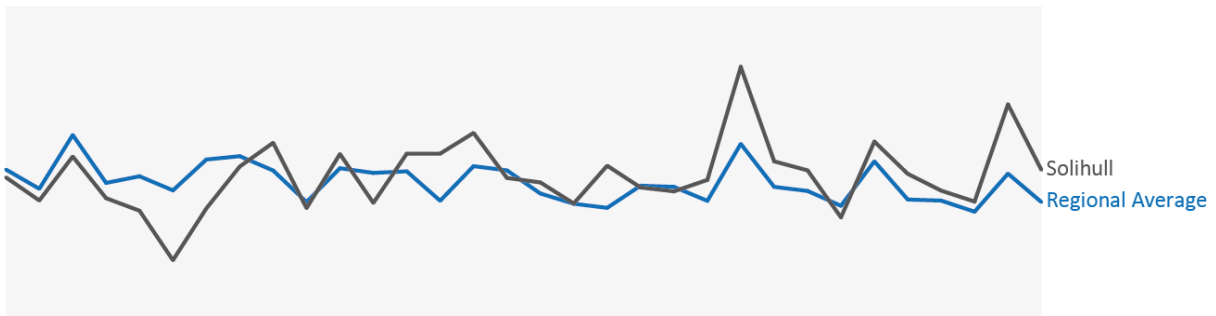
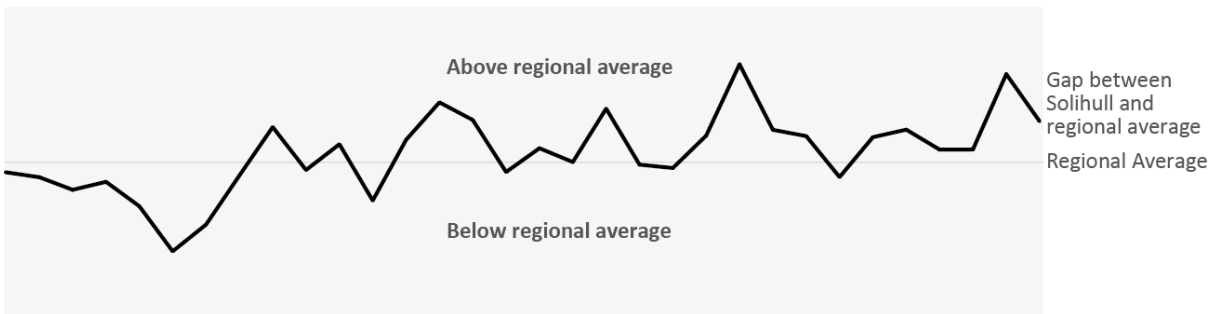


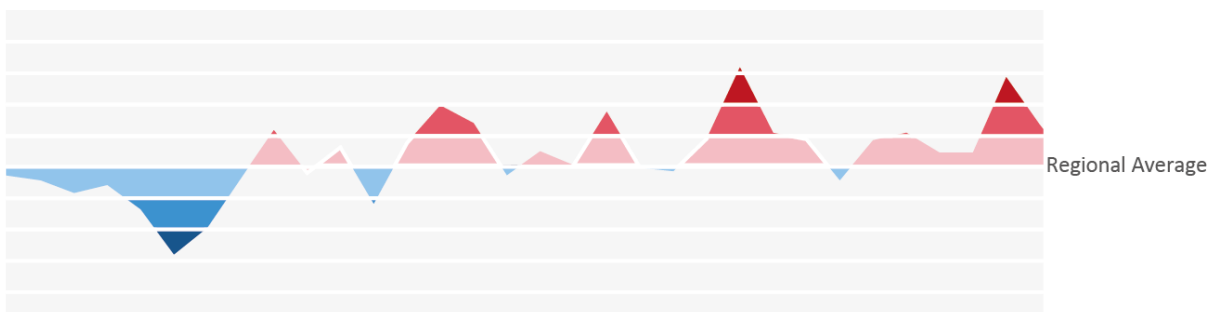
Figure 9.7 simplifies things by looking at the difference between Solihull and the regional average. We are now able to begin identifying the trend taking in place in Solihull in a regional context. Namely, after falling below the average for a short period at the beginning of our study period, the rate in Solihull then returned to levels similar to the regional average, remaining at or above the average for the majority of the period in question.

Figure 9.7: Illustrating the gap between an individual area and the regional average



Colour has been added in Figure 9.8 to draw attention to instances where the local figures for Solihull are above or below the regional average. Red denotes a local rate above the average, blue represents periods where Solihull's rate is below the average. The intensity of the colour makes larger variations stand out more easily; rather than varying colour intensity continuously, it has been varied in discrete steps. Given the complicated nature of a graphic consisting of 32 individual charts, applying steps in colour intensity will make it easier to identify notable changes in data values.

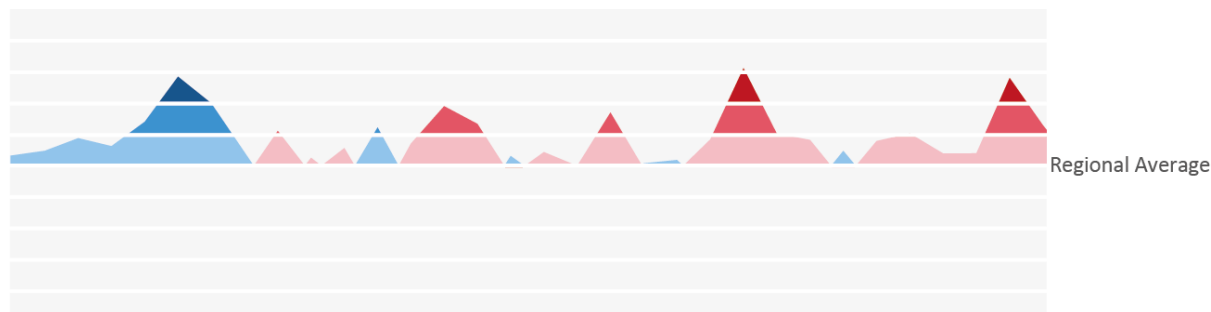
Figure 9.8: Adding colour to make the differences more visible



It is now much easier to identify extreme values, both above and below the average, when all 32 charts are eventually put together.

Given that there will still be a limitation to the height of the final diagram, it is helpful to make the best possible use of the vertical space. We do this by inverting the negative (in this case, blue) values and placing them above the horizontal axis. This provides us with Figure 9.9.

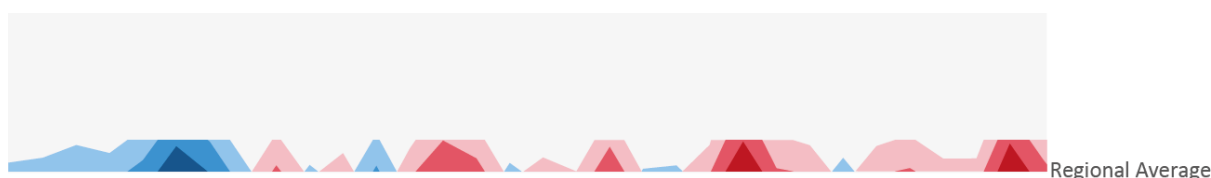
Figure 9.9: Inverting the negative values to utilise the available space more effectively



This works because of course at any one point in time the value can only be either above or below (or equal to) the regional average; there is no risk of any data being hidden. The slight downside is that the reader now needs to understand that some values are above the average while others are below; this is why the choice of colour is important. This example uses red because, intuitively, people associate red with a ‘warning’ of poor or weaker performance. In this case, red is interpreted as a higher than average rate of unplanned hospitalisations per 100,000 for chronic ambulatory care sensitive conditions. The darker the red, the further away from the average is the local rate.

Having improved the use of the vertical space, we can go one stage further and ‘collapse’ the colour bands so that they all sit on the same baseline. The darker colours sit at the front so that they can be seen. Applying this approach to the Solihull data gives us Figure 9.10.

Figure 9.10: Collapsing the colour bands to make best use of the space



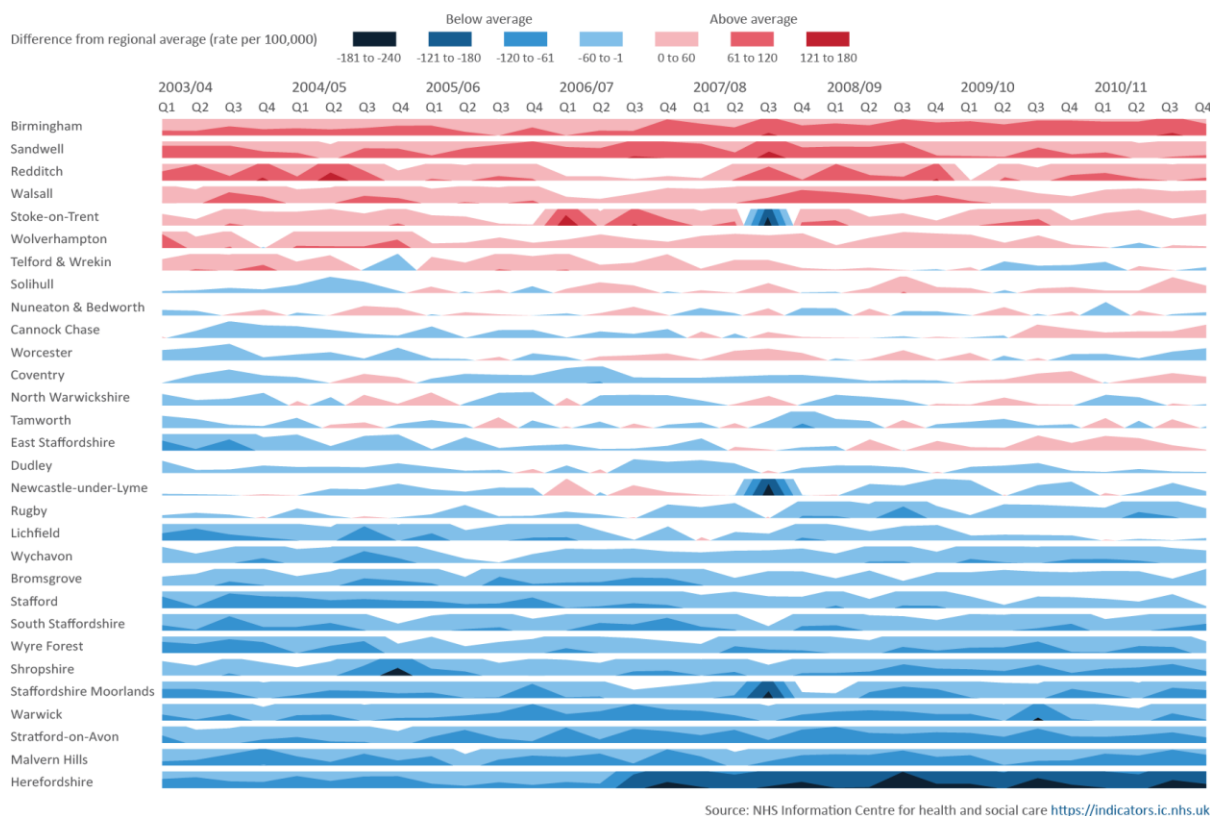
The vertical space gained from this approach allows us to either heighten the display of these values and/or increase the number of rows of charts on the screen. There are some trade-offs with this approach, which are discussed in Section 9.4.

Returning to the full dataset, Figure 9.11 presents information for all of the local authority areas in a single chart. Note that the thresholds for determining the final set of colour bands are broader than those used in Figure 9.9; using the same thresholds would have produced too many shades of red and blue for those areas with more extreme values and the graph would have been harder to interpret. It is advised that around six to eight colour bands are used (three or four for both positive and negative values). In this example, each shift in colour band represents an extra 60 (per 100,000) above or below the regional average.

The areas have been sorted from high to low in terms of average distance from the regional average over the full time period. This means that those areas that are generally above the regional average feature towards the top of the chart. Therefore we have a predominance of red towards the top of the chart and blue towards the bottom, giving a horizon effect. This is where the name of the chart comes from.

Figure 9.11: The final horizon chart

Unplanned hospitalisation for chronic ambulatory care sensitive conditions (adults), 2003/04 - 2010/11



Putting image size aside for a moment (it would work more effectively in landscape and full screen, and you can explore an online version [here](#)), it has certainly become easier to pick out some of the key messages. Once we are familiar with the methodology a scan of Figure 9.11 can reveal the following:

- Birmingham has, across the time period in question, the largest above average deviation from the regional trend.
- In addition to Birmingham, Sandwell, Redditch and Walsall remain above the regional average throughout the entire period.
- Conversely, Herefordshire has the lowest average rate and, particularly towards the end of the period, is considerably below the regional average (denoted by the very dark shades of blue). The chart prompts the question ‘what has caused this step change?’.
- Other areas, such as North Warwickshire and Tamworth, remain close to the regional average (denoted by the lighter shades and smaller peaks).
- In overall terms, we see more blue than red. This tells us that fewer areas tended to be above the average than below it. There could be two reasons for this; either the highest rates are more extreme, relative to the average rate, than the lowest rates or the higher rates tend to be associated with the more populous areas therefore skewing the average towards these higher rates. We see that the highest rates belong to areas like Birmingham, Sandwell, Walsall, Stoke and Wolverhampton. These are indeed the larger metropolitan boroughs and we can say that their values will be having a greater impact on the regional average than those at the foot of the chart.

- Our eye is also drawn to a number of strange looking values in quarter three of 2007/08. The charts for three areas seem to spike with large negative values. The areas in question are Stoke, Newcastle-under-Lyme and Staffordshire Moorlands. Is it a coincidence that these areas are all in the broader definition of Staffordshire? Is there a data accuracy issue for that particular quarter? These outliers will have been hard to detect from a scan of the raw data, but the horizon chart immediately draws attention to the issue.

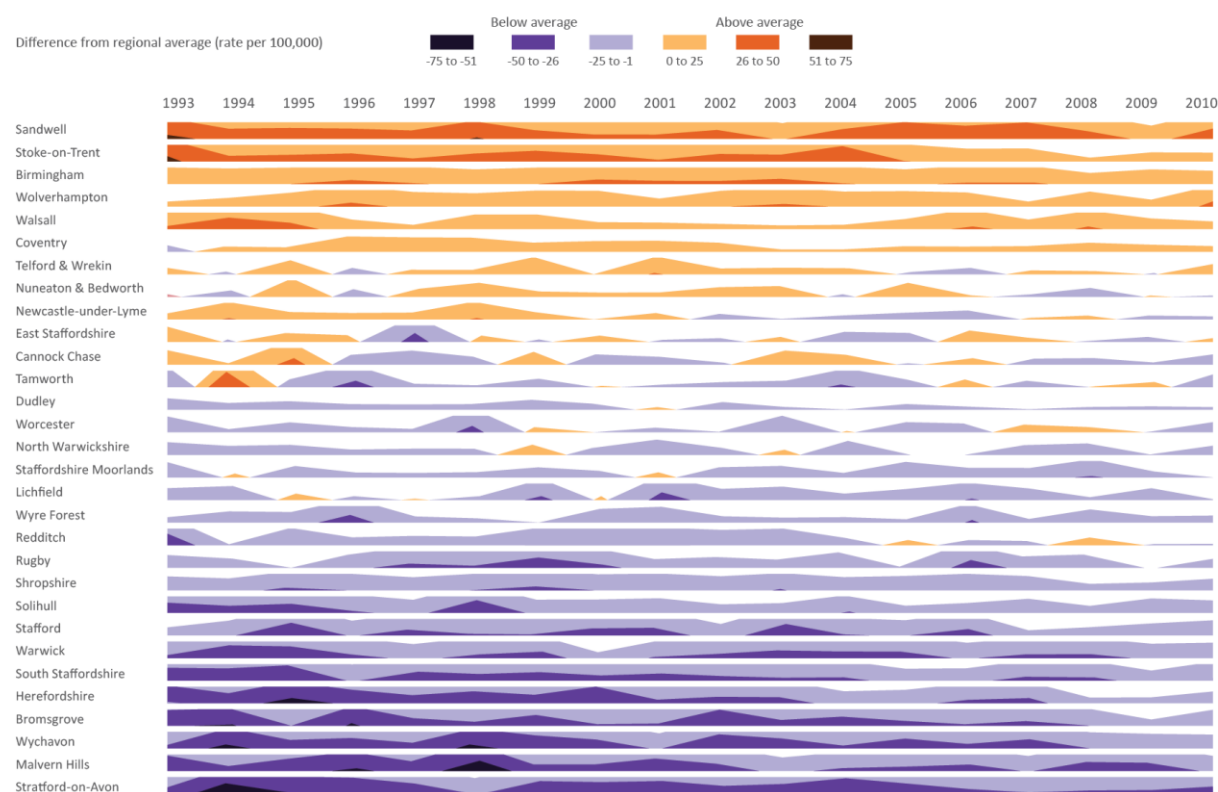
9.3 Practical local examples

To illustrate the approach applied to other datasets, a further three examples are provided here. All use health data covering the West Midlands region. Although we argue in Section 4 that the red-blue colour scheme is the most effective way of presenting the horizon chart, we have used different colour schemes in these three further examples to help distinguish between them.

Figure 9.12 presents data on avoidable deaths (mortality from causes considered amenable to health care) over the period 1993 to 2010.⁷ We see some interesting patterns emerge. For example the higher prevalence of dark shades towards the start of the time period indicates that, over time, there has been an underlying trend towards the regional average, with less extreme variation since the turn of the century. We also see that the metropolitan areas again feature towards the top of the diagram, specifically parts of the Black Country, Stoke and Birmingham.

Figure 9.12: Avoidable deaths in horizon chart form

Avoidable Deaths (mortality from causes considered amenable to health care), 1993 - 2010



Source: NHS Information Centre for health and social care <https://indicators.ic.nhs.uk>

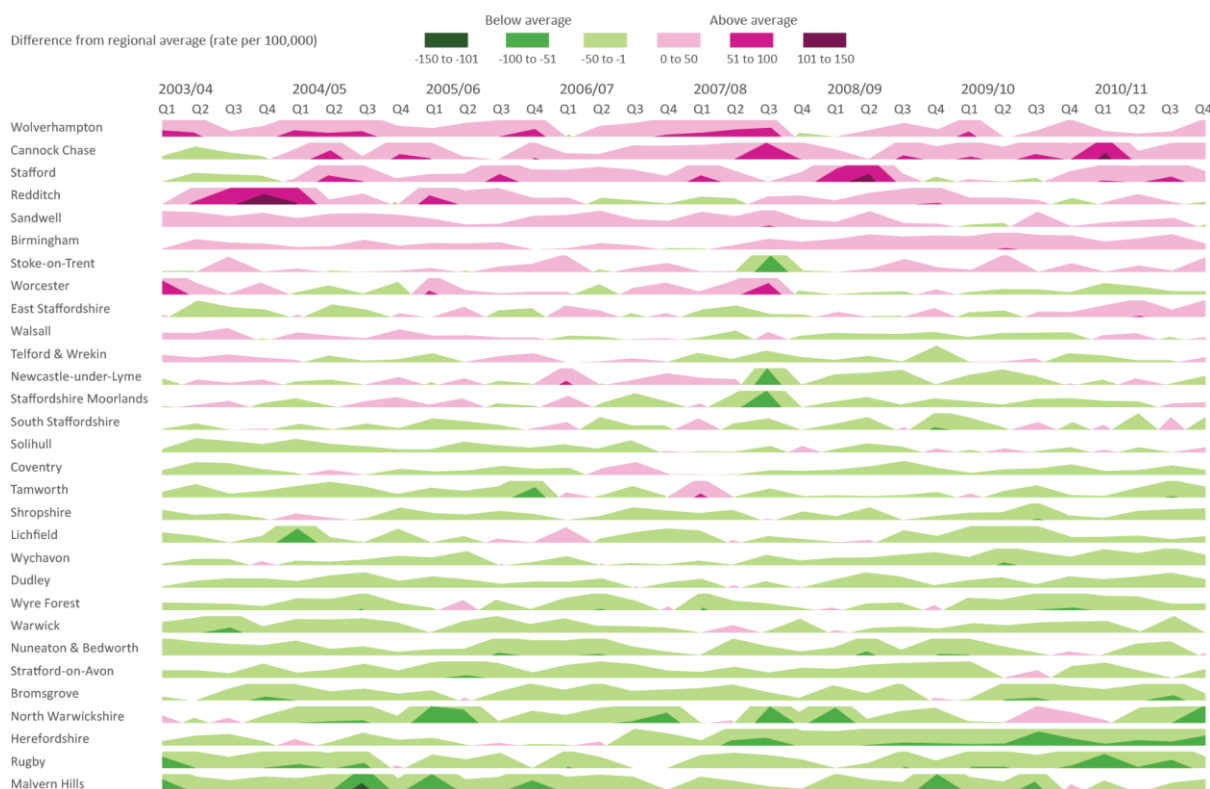
(Click [here](#) to see a full screen version)

Next, we have presented data on unplanned hospitalisation for asthma, diabetes and epilepsy in the under-19 population (Figure 9.13).⁸ Again, the horizon chart approach allows us to generate some

quick insight from around 1,000 statistics. For example, there are fewer instances of the darkest shades; this means values vary less from the average than we have seen in our other examples. There are fewer extreme values. Also, we see that individual local authority areas are less likely to remain above or below the regional average throughout the entire study period. Even those areas with the highest average rates (Wolverhampton, Cannock Chase, Stafford and Redditch) all fall below the regional average on at least one occasion (denoted by the green within their individual charts). Conversely, those areas with the lowest rates (including Malvern Hills, Rugby, Herefordshire and North Warwickshire) all rise above the regional average at some point. We can conclude that these health conditions vary less across the region than the other datasets we have examined, and that individual areas are less likely to display consistently strong or weak performance (in relative terms) on this indicator.

Figure 9.13: Hospitalisation for asthma, diabetes and epilepsy (<19 yrs) in horizon chart form

Unplanned hospitalisation for asthma, diabetes and epilepsy in the under 19s, 2003/04 - 2010/11



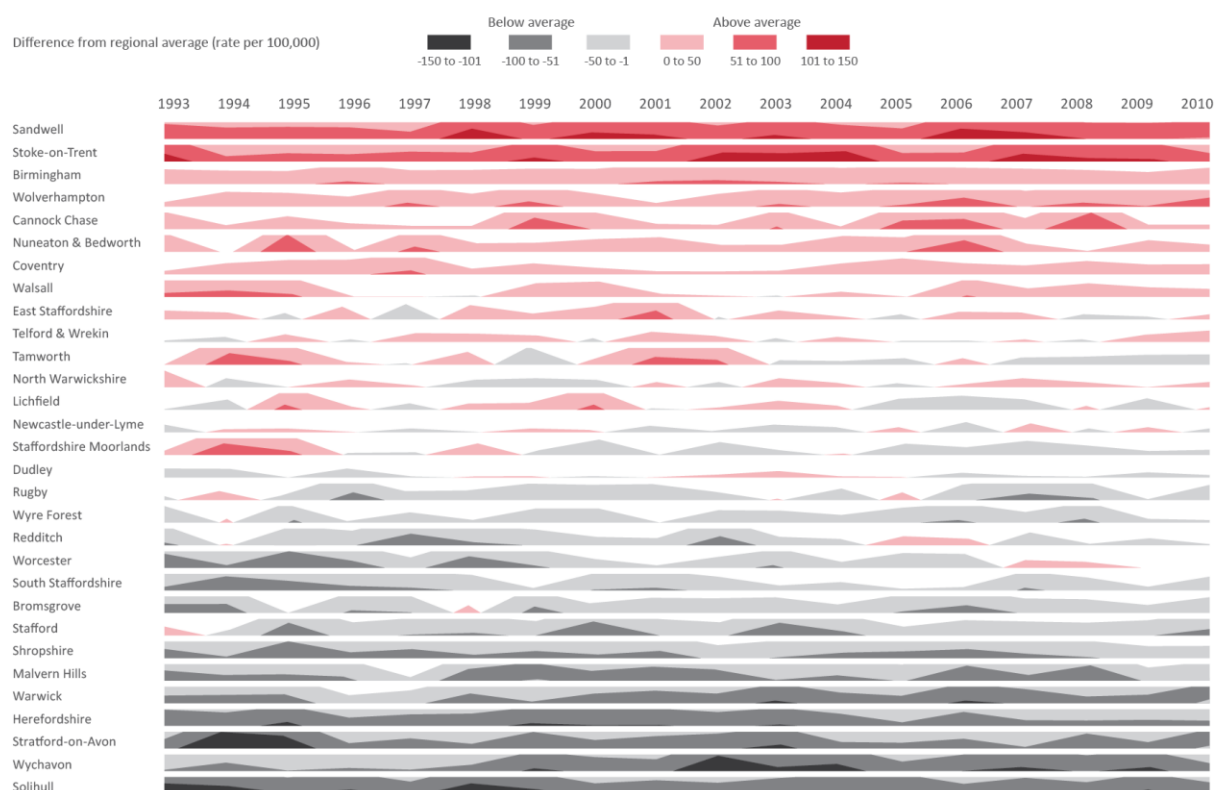
Source: NHS Information Centre for health and social care <https://indicators.ic.nhs.uk>

(Click [here](#) to see a full screen version)

Finally, we examine mortality rates (all ages, all causes) across the region over the past two decades.⁹ Figure 9.14 illustrates the various trends. Again, the highest variations from the regional average appear in metropolitan areas, notably Sandwell, Stoke, Birmingham, Wolverhampton and Walsall. Areas that tend to have the lowest rates include Stratford-on-Avon, Malvern Hills and Wychavon. This data would also work well in mapped form, demonstrating the differences in terms of urban/rural, but of course we would require a separate map for each year if we wished to examine trends over time as well.

Figure 9.14: Mortality rates (all ages, all causes) in horizon chart form

Mortality from all causes, all ages, 1993 - 2010



Source: NHS Information Centre for health and social care <https://indicators.ic.nhs.uk>

(Click [here](#) to see a full screen version)

9.4 Discussion and further applications

Although this methodology has been around for a few years, it is not widely applied and the purpose of this paper is to both make readers aware of the approach and discuss some of its strengths and weaknesses.

Firstly, let us consider the benefits of the horizon chart:

- It presents large amounts of data in a digestible way, once the method is understood. Readers may need time to absorb the approach the first time they see it, but subsequent views should prove immediately enlightening. The examples used in Sections 2 and 3 each present around 1,000 individual data points in a single chart.
- It is easy to identify extreme values and patterns. There is some perceptual effort required in understanding the chart, but this is outweighed by the benefit of being able to see a large volume of time series data.
- It enables an easy understanding of how a large number of cases (in our example, local authority districts) compare against a benchmark (in our example, the regional average). It can be applied in many other scenarios. For example, the individual cases could be comparable geographical areas, age groups, ethnic groups or socio-economic cohorts. The benchmark could be a national or local average, an overall population average or even a specific parameter that needs to be used for context. The horizon chart is an ideal tool to

display information for performance measures and other metrics where interpretation of variation between areas or groups over time would be beneficial

- A more generic point is that the 'classic' horizon chart, using a red-blue colour system is accessible to almost all readers. Often, data visualisations use a red-green approach to represent good/strong versus poor/weak performance. Around ten per cent of the population has a red-green colour deficiency, and the red-blue approach provides a useful and intuitive alternative.

Now, the limitations:

- The chart does not illustrate actual trends, only the *relative* performance of individual cases against a benchmark/average. We are not able to tell whether the actual rate of unplanned hospitalisations for chronic ambulatory care sensitive conditions is going up or down during the period in question. The purpose of the horizon chart is identify cases (areas) of notably high or low performance against an average, and whether cases are moving away from or towards the average over time. The chart illustrates relative rather than absolute change.
- It is not possible to see the actual values (in this case, rates per 100,000). Again, this is because the chart is being used to quickly identify key trends and patterns rather than the detail; this can be explored in the raw data once some areas of interest have been identified.

9.5 References

- 1 Panopticon Software Inc; c2013. *Home page*. <http://www.panopticon.com> (accessed 28 Jan 2013).
- 2 Few S. Time on the Horizon. *Visual Business Intelligence Newsletter*. Jun/Jul 2008. Available at: http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf
- 3 Rimlinger F. *Sparklines for Excel*. Weblog. <http://www.sparklines-excel.blogspot.co.uk/> May 2008. (accessed 28 Jan 28 2013).
- 4 Tufte E. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.
- 5 Tufte E. *Envisioning Information*. Cheshire, CT: Graphics Press, 1997.
- 6 NHS Information Centre for health and social care. *NHS Outcome Framework Indicators - Dataset: 2.3.i Unplanned hospitalisation for chronic ambulatory care sensitive conditions (adults)*. Leeds: NHSIC; 2012. Available at: <https://indicators.ic.nhs.uk>; indicator page ID P01383.
- 7 NHS Information Centre for health and social care. *Compendium of population health indicators - Dataset: Mortality from causes considered amenable to health care: directly standardised rate, <75 years, annual trend, MFP*. Leeds: NHSIC; 2012. Available at: <https://indicators.ic.nhs.uk>; indicator page ID P00362.
- 8 NHS Information Centre for health and social care. *NHS Outcome Framework Indicators - Dataset: 2.3.ii Unplanned hospitalisation for asthma, diabetes and epilepsy in under 19s*. Leeds: NHSIC; 2012. Available at: <https://indicators.ic.nhs.uk>; indicator page ID P01384.

- 9 NHS Information Centre for health and social care. *Compendium of population health indicators - Dataset: Mortality from all causes: directly standardised rate, all ages, annual trend, MFP*. Leeds: NHSIC; 2012.
Available at: <https://indicators.ic.nhs.uk>; indicator page ID P00345.

9.6 Further information

The charts produced in this paper were initially created using the Sparklines add-in for Excel and then modified in Adobe Illustrator. The Sparklines tool and a user manual can be accessed here: Available at: www.sparklines-excel.blogspot.co.uk

Kirk A. *Data Visualisation: a successful design process*. Birmingham: Packt Publishing, 2012
Available at: <http://www.packtpub.com/data-visualization-a-successful-design-process/book>

Heer J, Kong N, Agrawala M. Sizing the Horizon: The Effects of Chart Size and Layering on the Graphical Perception of Time Series Visualizations, *ACM Human Factors in Computing Systems (CHI)*; 4-9 Apr 2009:1303-12
Available at: <http://vis.berkeley.edu/papers/horizon/>

9.7 Author

Spencer Payne, Corporate Research Manager, Warwickshire Observatory.

Version 1.0
18 March 2013

