Exploring Digital Welfare data using GeoTools and Grids

Hodkinson, S.N., Turner, A.G.D.

School of Geography, University of Leeds

June 20, 2014

Summary

As part of the Digital Welfare project [1] a Java GIS library called GeoTools [2] has been used to automate the production of numerous maps. The extended abstract outlines this work and provides some detail of the geographical analysis involved. The main source data are: client enquiry data obtained from the main advice giving agencies in Leeds Local Authority District (LAD); and, housing and council tax benefit claimant data for Leeds LAD. The primary data are spatially referenced by residential postcodes either at the postcode unit or postal sector level. The data have been explored and are being analysed for specific purposes that are outlined.

[1] http://www.geog.leeds.ac.uk/people/a.turner/projects/DigitalWelfare/
[2] http://www.geotools.org/

KEYWORDS: Java, GeoTools, Leeds, Welfare, Automation, Mapping, Postcode, Census, Advice, MAUP

1. Introduction

Java programs are being developed in order to reproduce (in a few simple steps) spatial generalisations and geographical maps from postcoded benefit claimant and Advice Leeds client data. The programs make use of GeoTools (an open source Java library that provides tools for geospatial data) and Grids (an open source Java library for manipulating 2D square celled raster data) and are similarly made available as open source (Turner, 2015).

Customised exploratory data analysis tools are being developed based on this work. The hope is that these will prove useful in both helping to understand the geography of benefits claimants and advice seekers and also in strategically reorganising services to support citizens in an era of seemingly increasingly constrained budgets. This work is being done in collaboration with Leeds City Council and Advice Leeds and is focussed at this stage on the Leeds Local Authority District (LAD).

Section 2 provides some more detailed context for this work. Section 3 focuses on the input data. Section 4 presents a selection of Advice Leeds client data geographical maps that illustrate and revisit the Modifiable Areal Unit Problem (MAUP). Section 5 is for discussion. Section 6 concludes and outlines some of the next steps we are planning to take.

2. Context

This research is operating on a pro-bono basis given support from the School of Geography at the University of Leeds. For a time it was partially funded as part of the Digital Welfare Project and RCUK Digital Economy Communities and Culture Network+ under the theme of Communities and Culture. The work is also based on research conducted over a number of years on other projects that evolved methodology and software, skills and knowledge.

There are a combination of reasons for using Java, GeoTools and Grids for this work. An important

one is familiarity and a desire for greater familiarity, but these are also capable, functional and reasonably stable and sustained technological developments. GeoTools is a widely used platform for web based geographical information systems that seeks to implement standards for interoperability. GeoTools is adaptable and can readily be extended in bespoke ways to develop custom interactive exploratory geographical data analysis tools. Additionally GeoTools and Grids are open source and whilst Java is not fully open source it is widely available. One of the main advantages of going down the free and open source software route is that the resulting mapping technology can be readily used and deployed on any IT infrastructure without the need for expensive software licenses.

The Digital Welfare Project aimed to look at the impacts of contemporary welfare policy changes on both service users and service providers in both welfare and general advice contexts. The project had a focus on critically exploring digitalisation and its effects on advice service users, advice service providers, and benefits claimants. What digitalisation refers to in this context is to do with how access to benefits and advice and the interface between service users and providers is becoming more online and computerised and less face to face and human. So, the Digital Welfare Project focussed on something different to what is presented here, but it provides very important context both to the work presented here and for additional work we are undertaking to try to better understand the distribution and demand for advice services, and provide useful information for service reorganisation.

The place of study is in and around Leeds which hosts GISRUK 2015. Leeds is a metropolitan Local Authority District (LAD) area of around 552 km2. The Leeds LAD is one of five metropolitan LADs that comprise the county of West Yorkshire in the broader region of Yorkshire and the Humber in the North of England in the UK. Arguably, Leeds has a broader city region that extends beyond both the LAD boundary and the boundary of West Yorkshire. Our investigations have revealed that many citizens seeking advice from Advice Leeds are not resident within the Leeds LAD.

3. Input Data

In the broader research we are undertaking we not only use data from Advice Leeds, but also data provided by Leeds City Council on welfare benefits claimants and residents in social housing. Although these data have names and addresses, the data we are supplied with has had this removed, but in most cases, the unit postcode is part of each record.

The maps shown in this extended abstract are based on the following: Advice Leeds client enquiry data at postcode unit level; postcode sector and unit postcode boundaries; 2011 UK Human Population Census boundaries (at Output Area, Lower Layer Super Output Area, Middle Layer Super Output Area levels) and the ONS Postcode Directory look up (ONSPD).

The Advice Leeds client enquiry data at postcode unit level includes data from Leeds Citizens Advice Bureau (CAB), Leeds Chapeltown CAB and Leeds City Council Welfare Rights Unit. Data from other Advice Leeds organisations are not included. The postcode sector and unit postcode boundaries were obtained via Edina Digimap. The census boundaries and the ONSPD were downloaded from the Office for National Statistics (ONS) website.

The Advice Leeds client data are sensitive even though the data have had names and full addresses of individuals removed. However, the aggregated and mapped data presented here are not considered to be sensitive and it is thought that these can be disseminated openly. The maps are more for illustrative purposes than for studying the distribution of Advice Leeds clients in detail. Readers familiar with the geography of Leeds might recognise some geographical patterns in the maps. In particular, it is likely that the Advice Leeds client distributions correlate in places with commonly used measures and indices of deprivation. It is intuitive that these things correlate and indeed we have considered developing new measures of deprivation based on data we have been provided with.

4. Initial Mapping

Figures 1 to 16 are effectively maps of the same Advice Leeds postcode unit client count data for 2012. Values of zero are shown as white. The data have been classified using an equal interval

classification which divides up ranges into equal intervals based on the minimum and maximum values for each unit at each resolution. The classified data are displayed using 8 colours of red; with lower values coloured in lighter reds and the higher values coloured in darker reds. The maps shown in all the figures have the same shaped green outline which represents the Leeds LAD boundary. The data shown are for an annual time period and they represent unique clients according to client references in the data. (There may be clients represented multiple times if they are served by multiple parts of the Advice Leeds service in the time period). There are a number of reference points depicted on each of the maps. These are some of the Advice Leeds locations where citizens have been or were able to seek advice face to face.

To produce the maps, the unit postcode data were attributed to points and either these points were used in a point-in-polygon type method, or the ONSPD was used to look up which census geography a postcode was predominantly in, or the postcode was truncated into a postcode sector code. Whichever of these methods was used, the aggregated data are pretty much the same with the exception being at the boundaries where the unit areas overlap the boundary of the Leeds LAD.

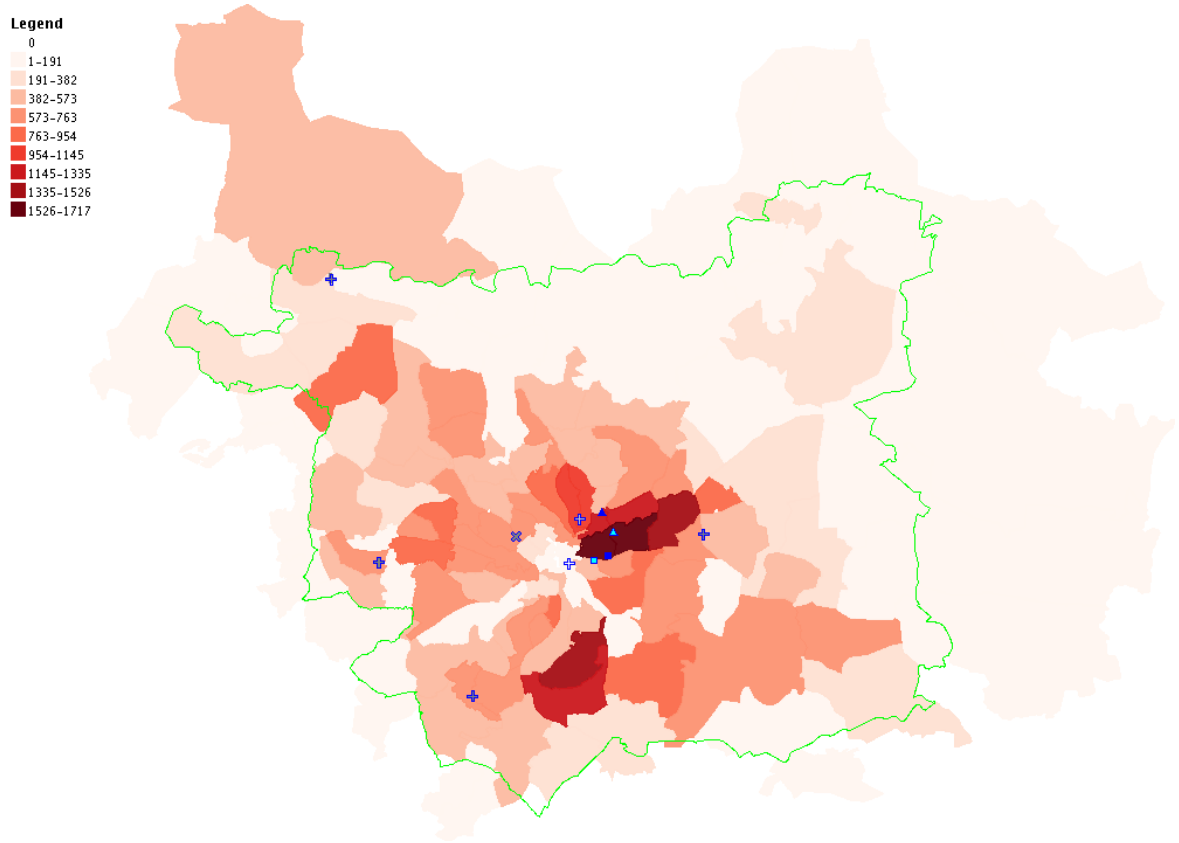

Figure 1 Map of Advice Leeds Client Count at Postcode Sector Resolution

Figure 1 presents the client count data for all postcode sectors that intersect with the Leeds LAD. Figure 2 presents the client count data for Middle Layer Super Output Areas (MSOAs). Figure 3 presents the client count data for Lower Layer Super Output Areas (LSOAs). Figure 4 presents the client count data for Output Areas (OAs). Figure 5 presents the client count data for unit postcodes. The spatial resolution of the data is generally increasing in the maps shown from Figures 1 to 5. Figures 6 to 9 show square celled raster representations of the counts at 400, 200, 100 and 50 metre resolutions respectively. Although the pattern in the maps is similar, there are differences, and collectively the maps in Figures 1 to 9 illustrate various aspects of the so called Modifiable Areal Unit Problem (MAUP) detailed by Openshaw (1984). In short the MAUP cautions inference as patterns shown in maps are dependent on the resolution of the maps and the choice of boundaries, so the less regular the units are, the more concern there is with regard spatial bias.

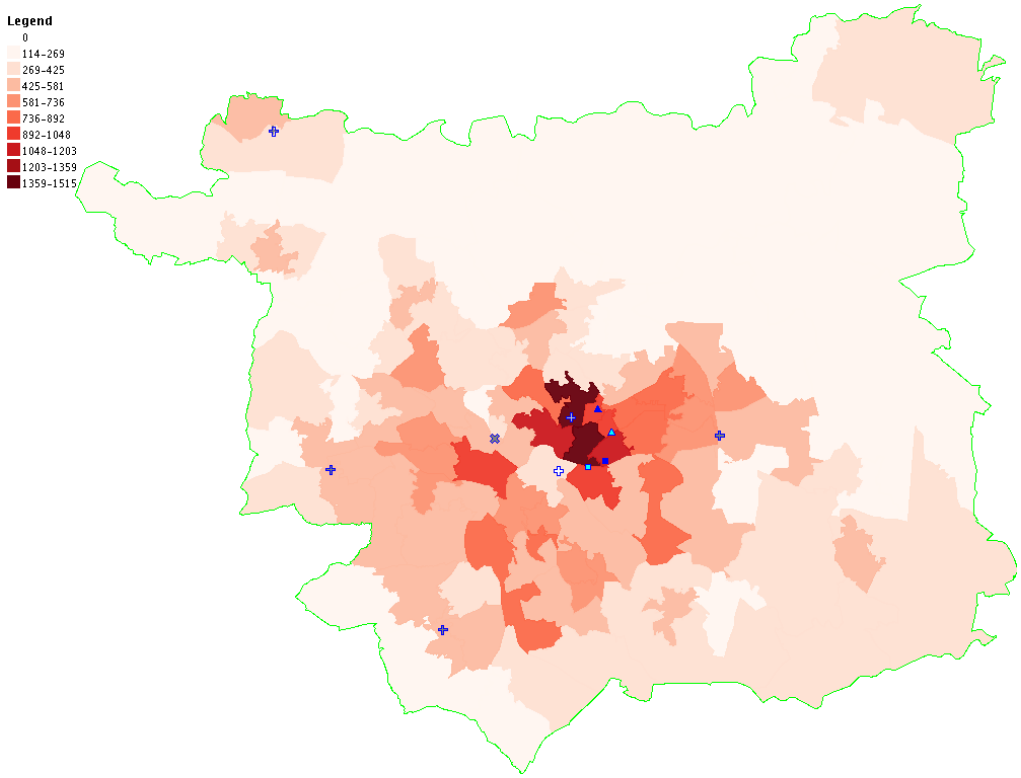

Figure 2 Map of Advice Leeds Client Count at MSOA Resolution

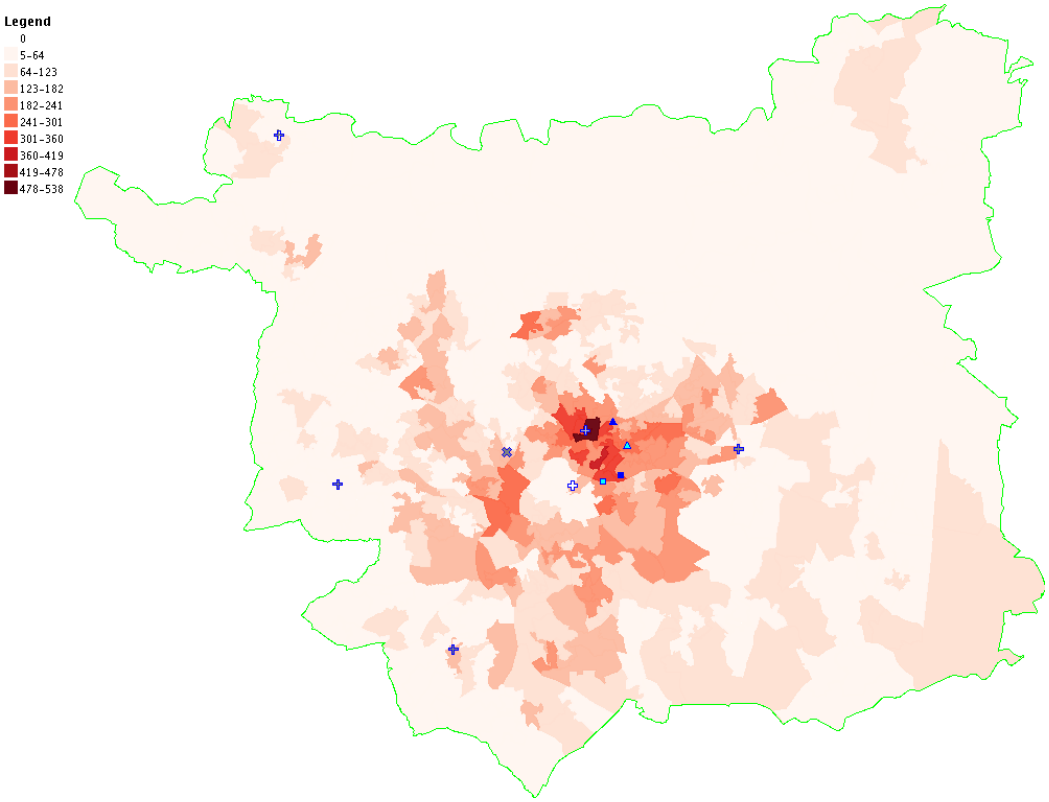

Figure 3 Location of Lancaster University

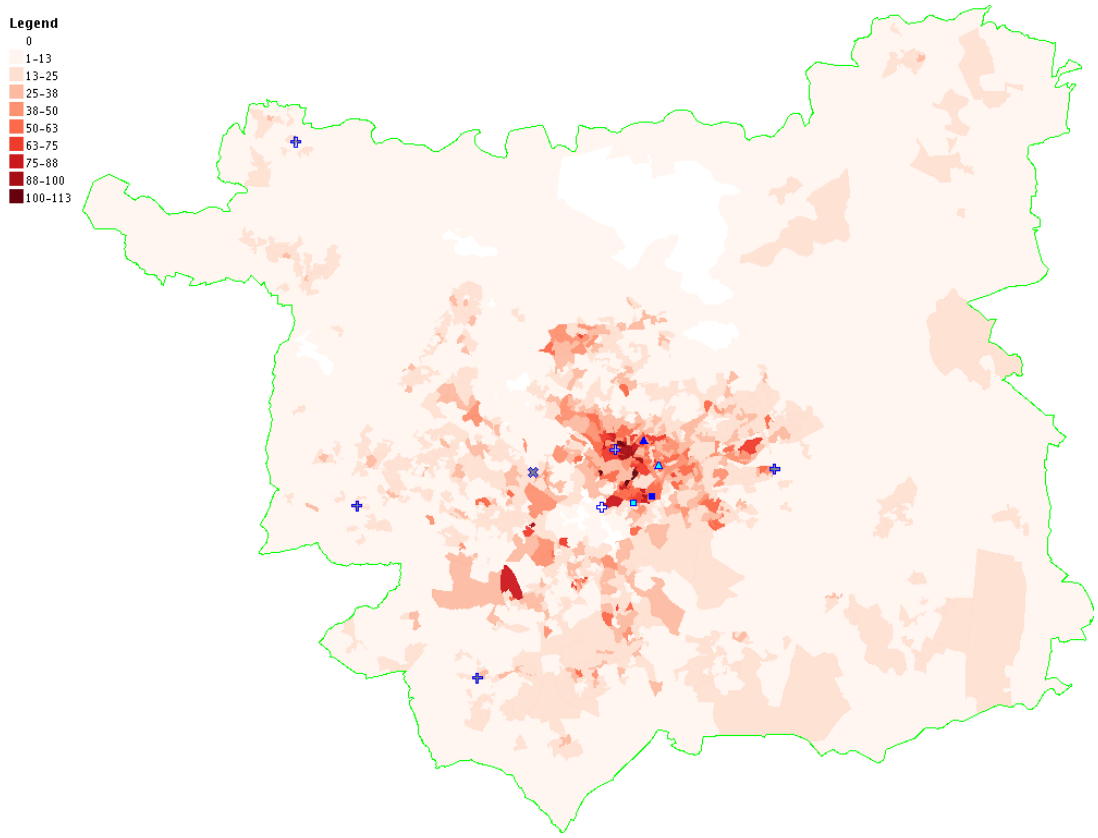

Figure 4 Map of Advice Leeds Client Count at OA Resolution

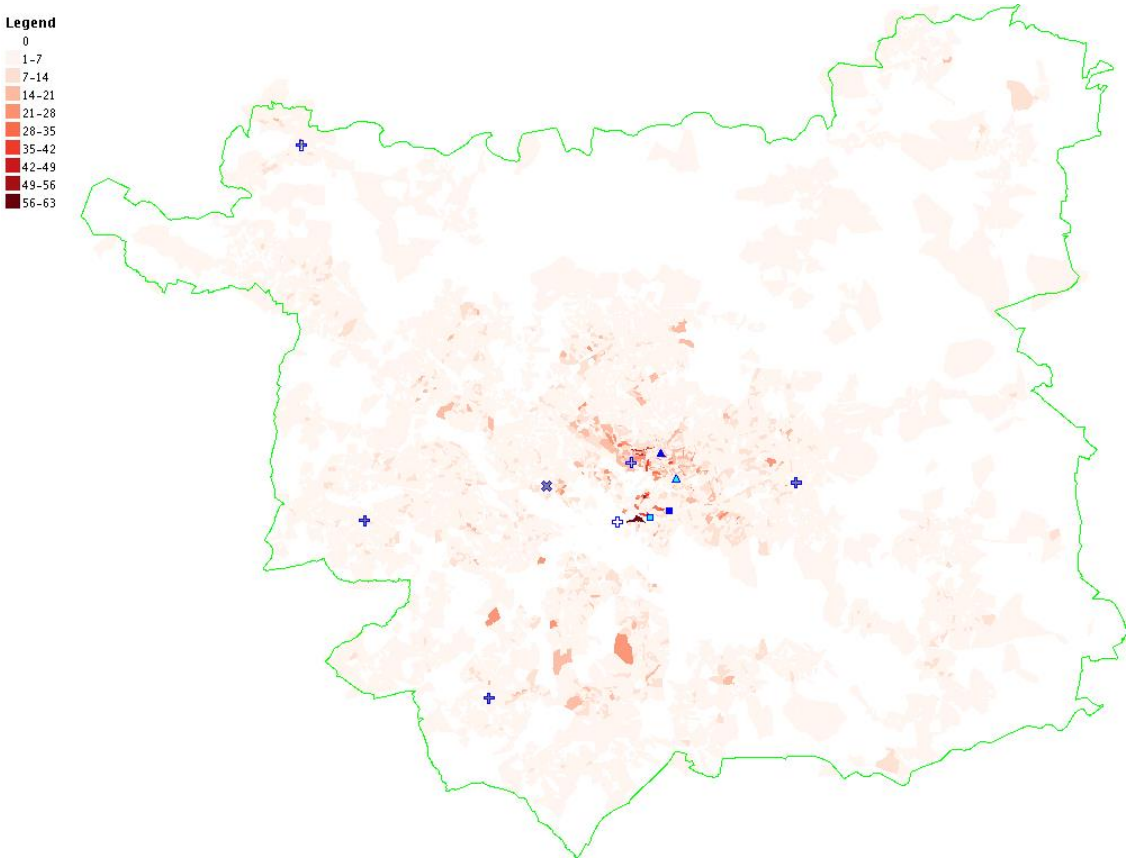

Figure 5 Map of Advice Leeds Client Count at Postcode Unit Resolution

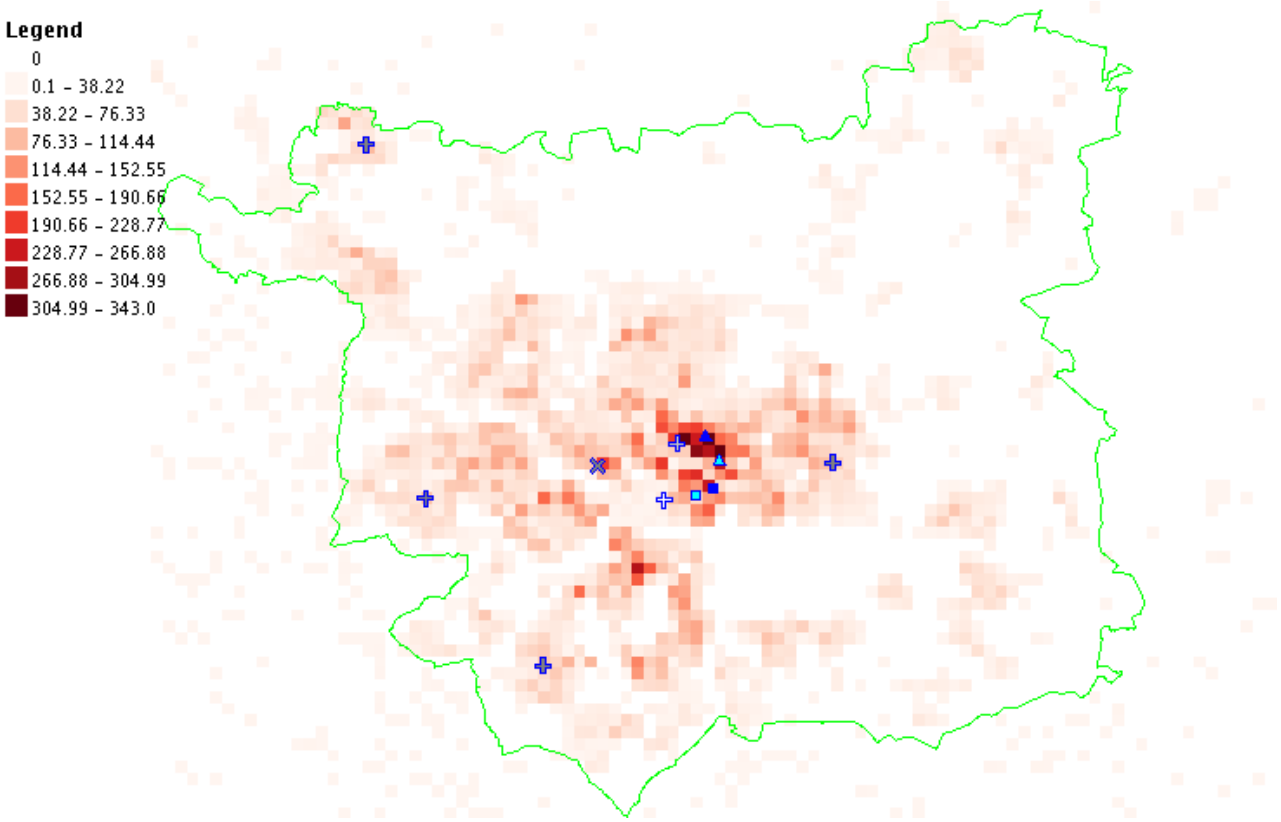

Figure 6 Map of Advice Leeds Client Count at a 400 metre Grid Resolution

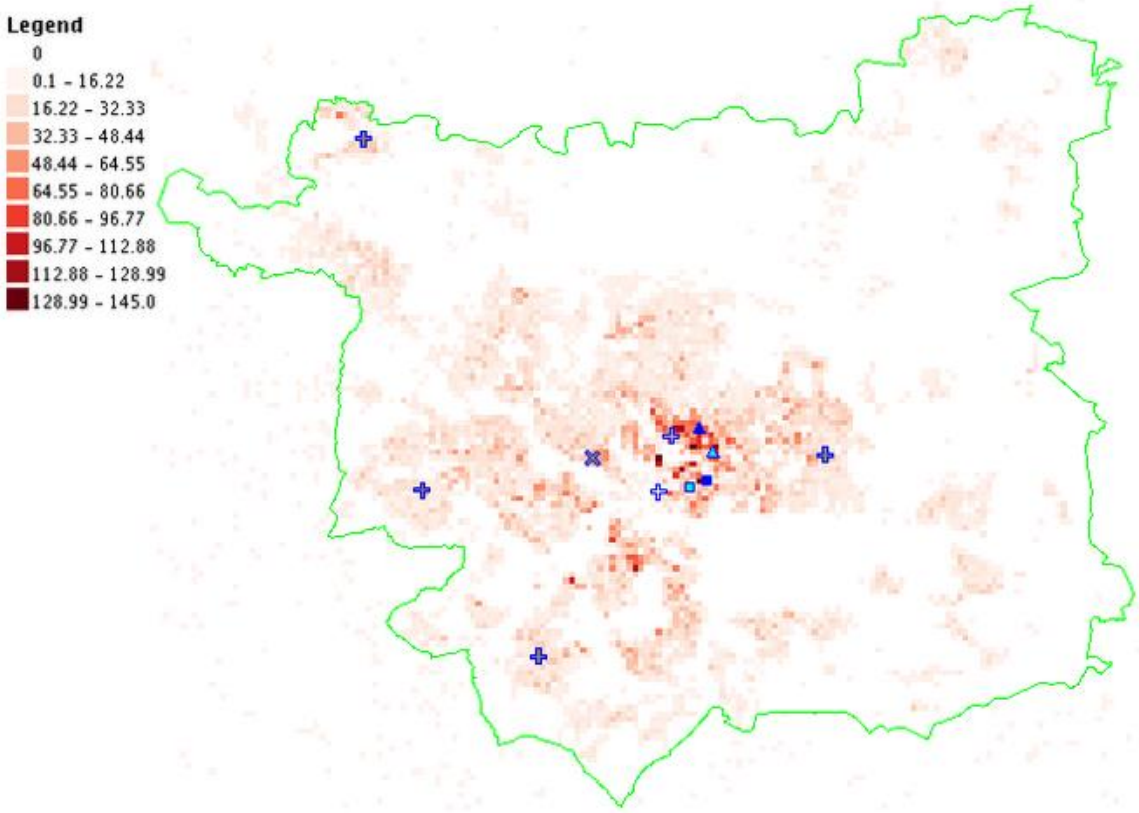

Figure 7 Map of Advice Leeds Client Count at a 200 metre Grid Resolution

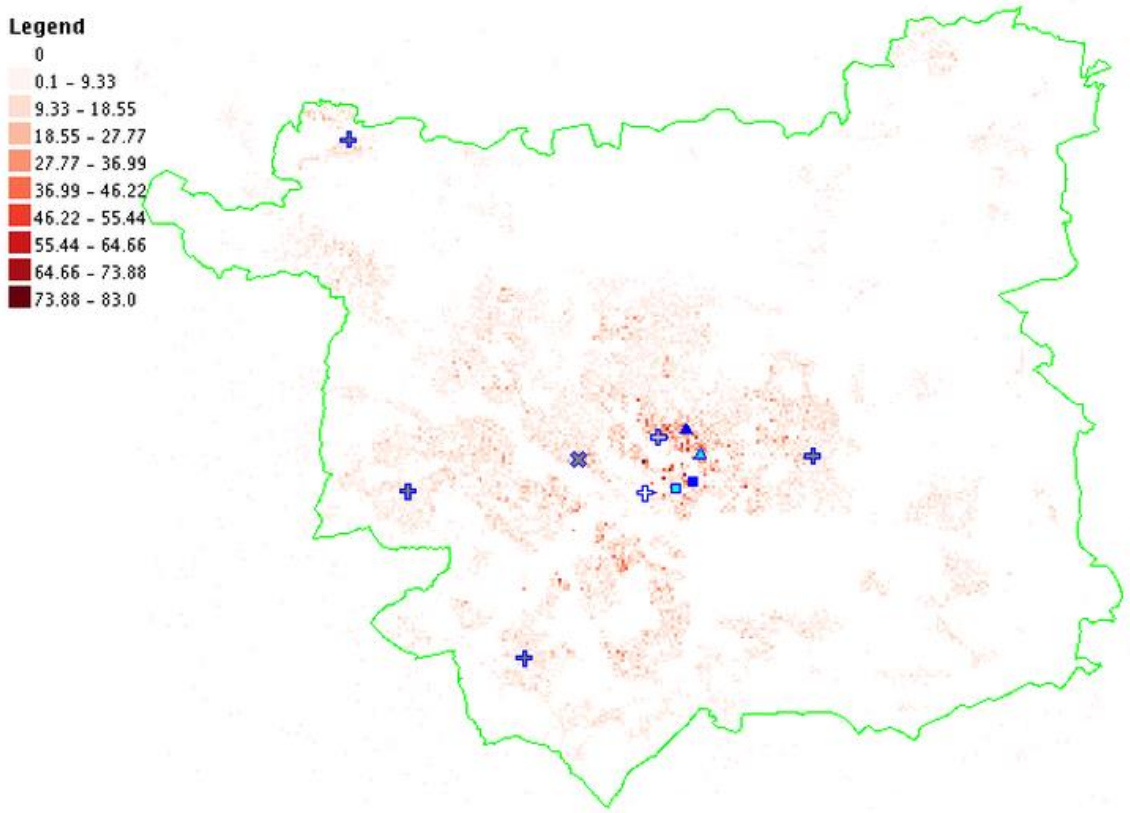

Figure 8 Map of Advice Leeds Client Count at a 100 metre Grid Resolution

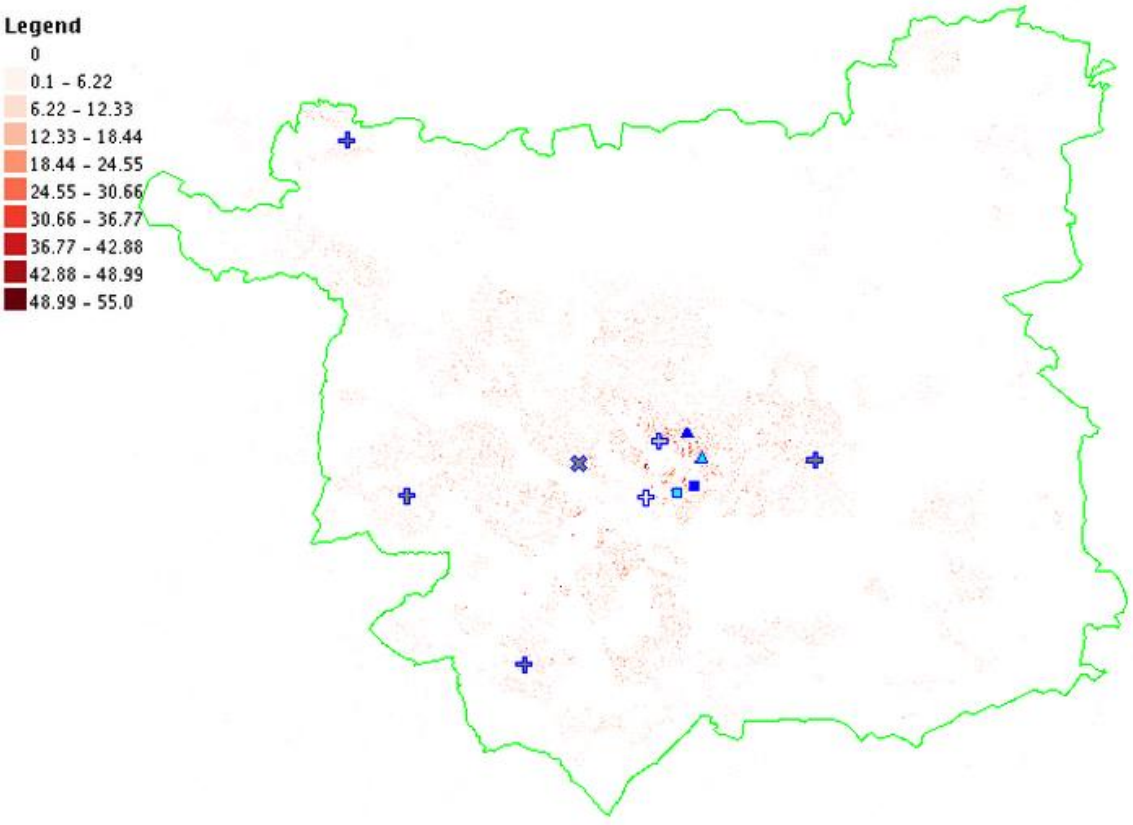

Figure 9 Map of Advice Leeds Client Count at a 50 metre Grid Resolution

Figures 6 to 9 show raster data where all the cells are the same size. Because the cells are all the same size, the maps in these figures show the same pattern as densities. Figure 6 shows values for a grid with 400 by 400 metre square cells and has a largest client count value of 343. This value is between those of the LSOA and OA resolutions with largest values of 538 and 113 respectively (as shown in Figures 3 and 4). Figure 7 shows values for a 200 by 200 metre square cells and has a largest client count value of 145 which is more similar to the OA resolution. Figure 8 and 9 show values for a 100 by 100 metre and 50 by 50 metre grids and has a largest client count value of 83 and 55 respectively. These are above and below the unit postcode resolution maximum value show in Figure 5 to be 63.

For the postcode sectors, MSOAs, LSOAs, OAs and postcode units, the count data can generally be converted into density data by dividing by the area of each unit. Respectively, the density data are shown in Figures 10 to 14. Given a quick visual inspection, it should be clear that Figures 6 to 14 appear more similar in pattern than do Figures 1 to 9.

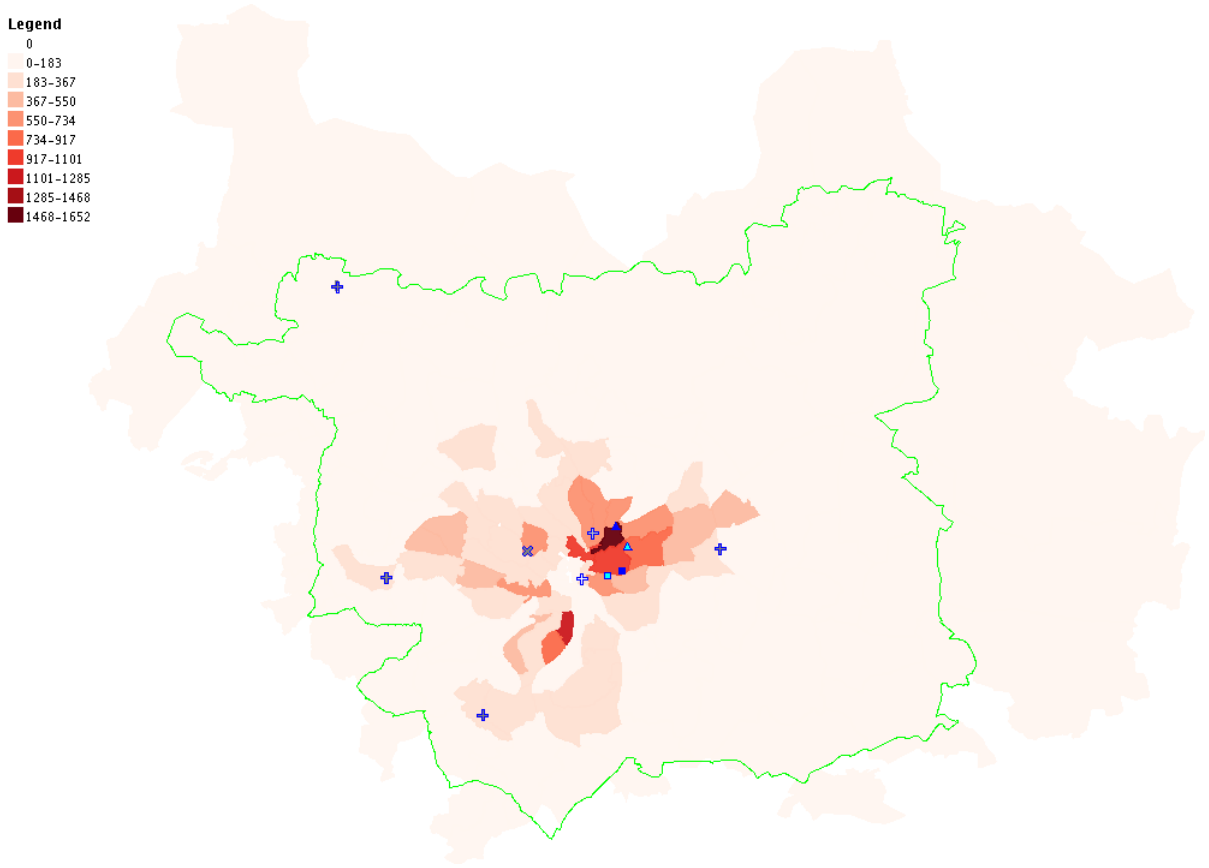

Figure 10 Map of Advice Leeds Client Density at Postcode Sector Resolution

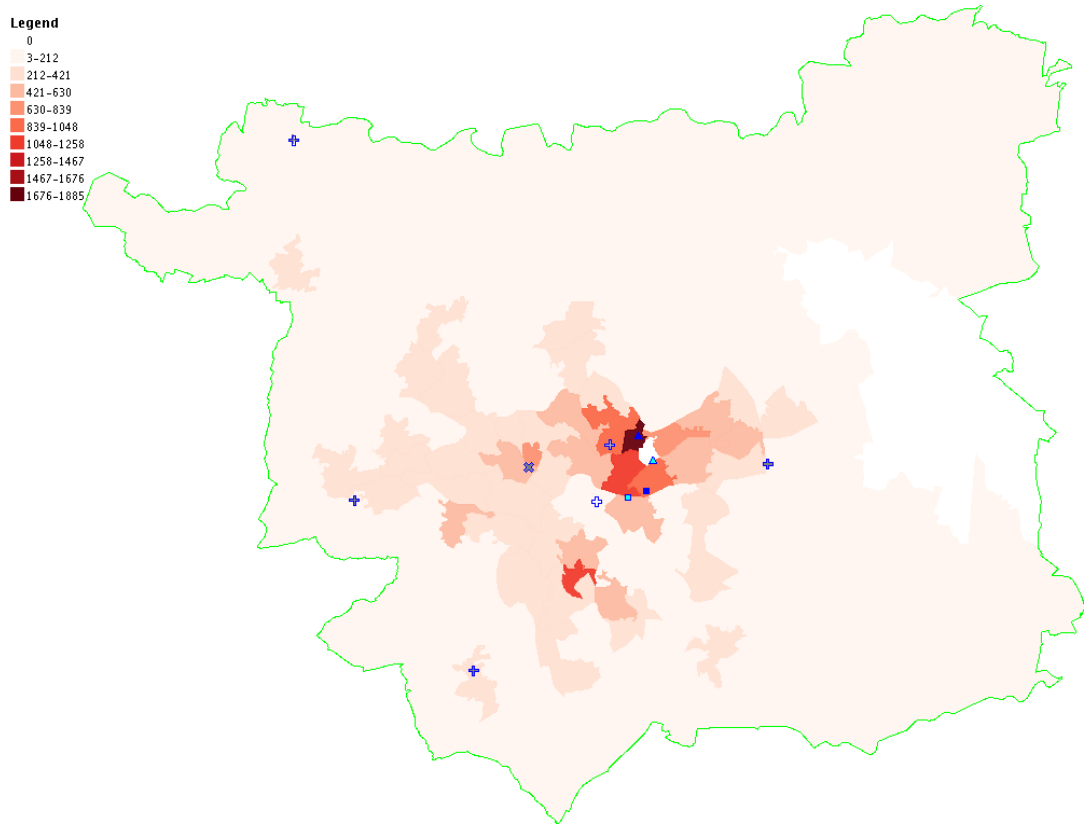

Figure 11 Map of Advice Leeds Client Density at MSOA Resolution

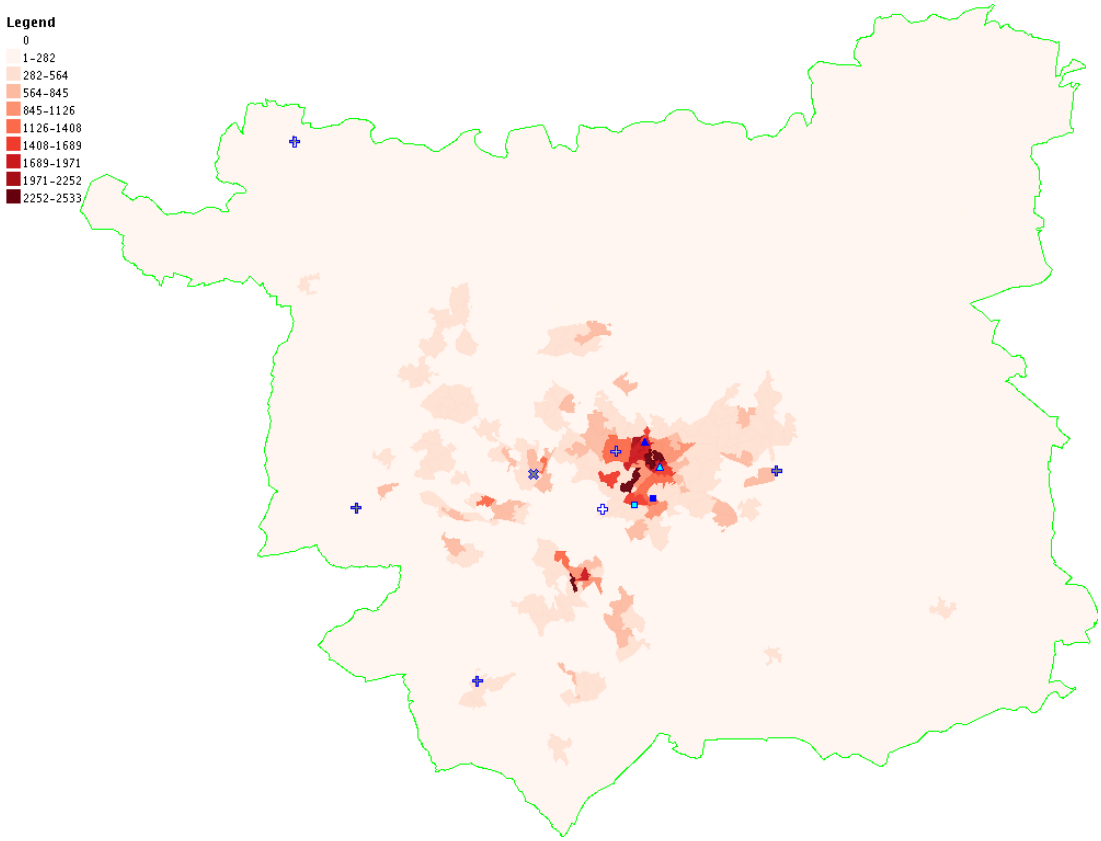

Figure 12 Map of Advice Leeds Client Density at LSOA Resolution

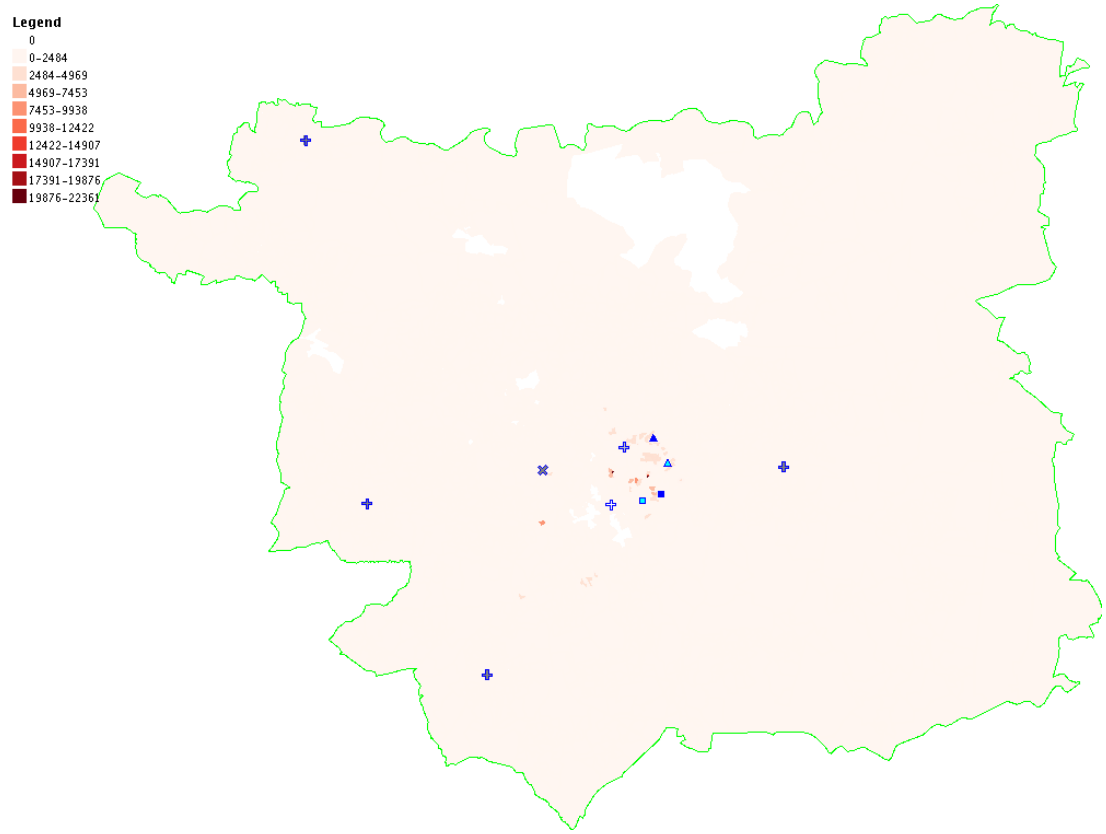

Figure 13 Map of Advice Leeds Client Density at OA Resolution

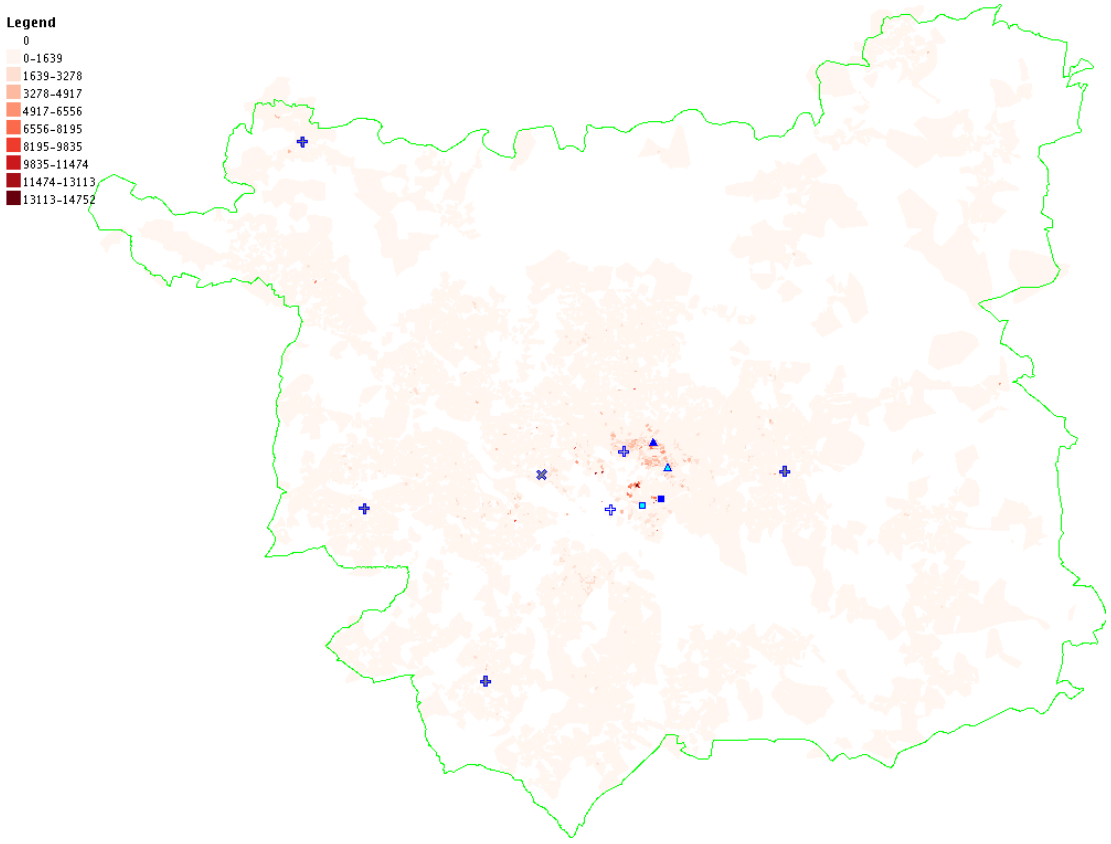

Figure 14 Map of Advice Leeds Client Density at Postcode Unit Resolution

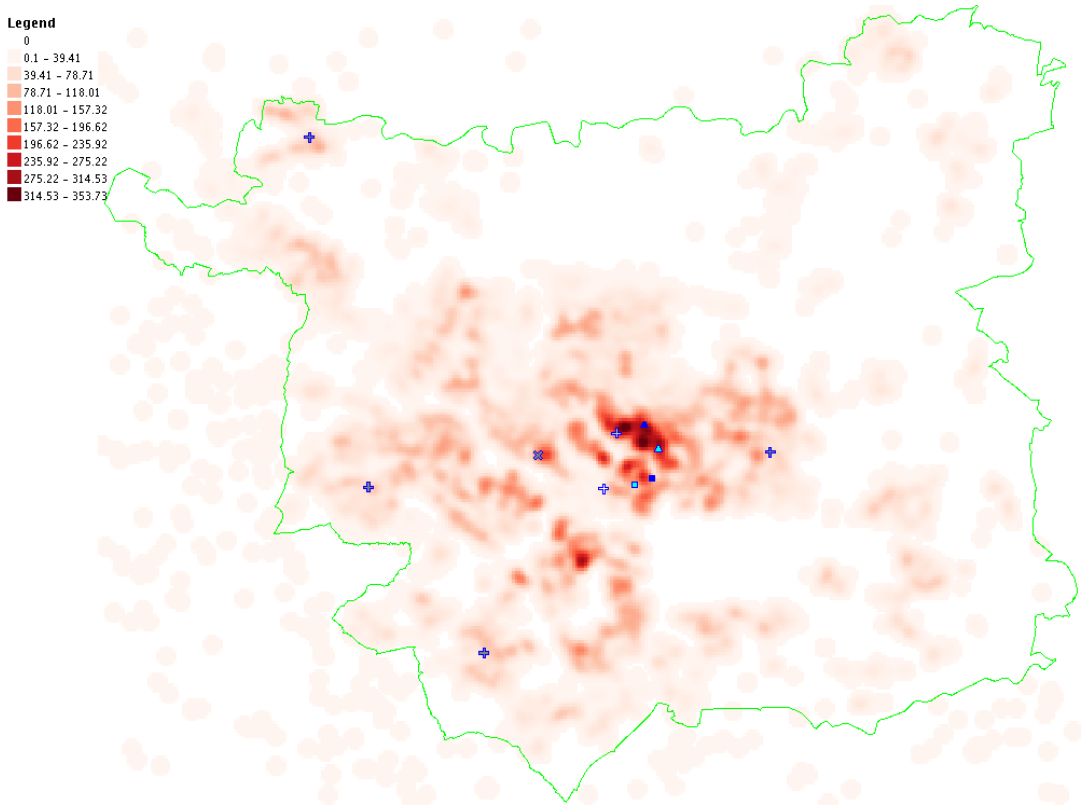

Figure 15 Map of Advice Leeds Client Density Generalised across a range of scales from 100 metres to 400 metres

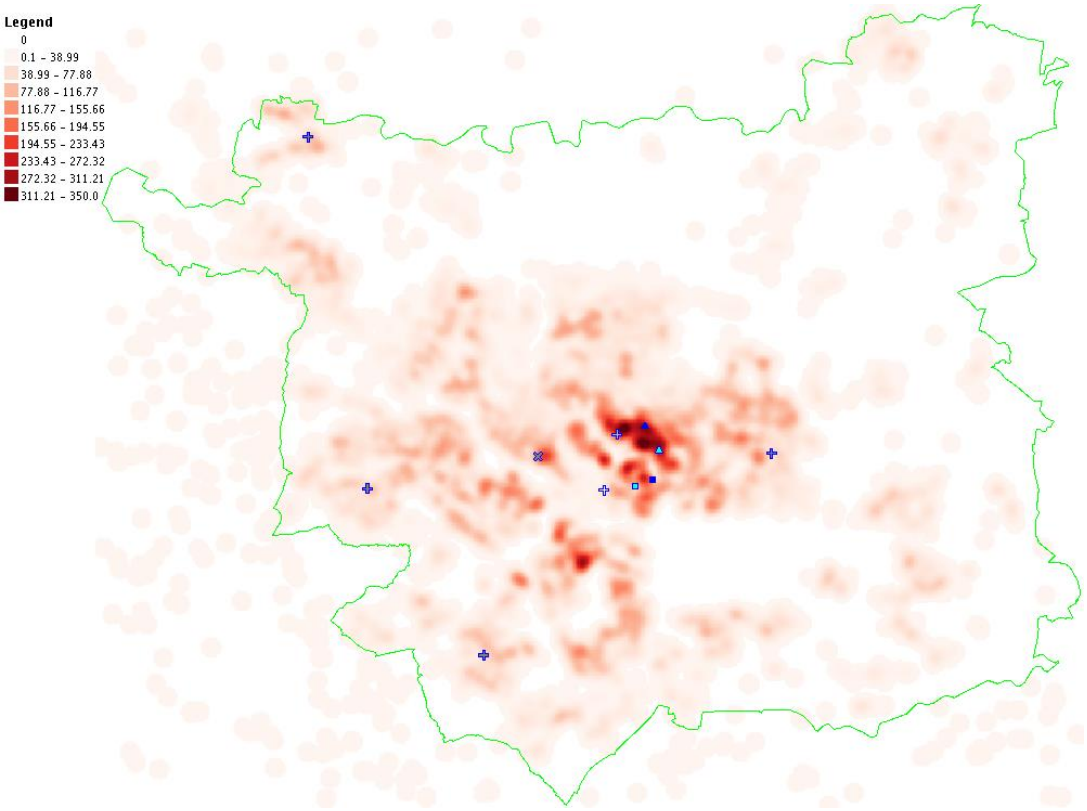

Figure 16 Map of Advice Leeds Client Densities Generalised across a range of scales from 50 metres to 400 metres

The maps reveal that the distribution of Advice Leeds clients is highly concentrated. In general they are concentrated in the more inner city areas - areas where large numbers of people reside, and with significant proportions of these being what we might term as ‘deprived’ and that we might expect to seek advice from Advice Leeds.

Detail is somewhat lost in the maps at high levels of spatial resolution as the areas become very small. However, the maps displayed could be generated much larger and with a zoomable slippy interactive maps, even the most detailed resolution can be explored at this scale to get a good idea of the distribution of Advice Leeds clients.

Generalising the raster data across a range of scales helps to reveal the distribution more clearly at scale. A detailed method for doing this is described by Turner (2000). Figure 15 shows the 100 metre resolution initial raster displayed in Figure 8. The values within a 400 metre radius being generalised back to each cell using a distance weighted function. Figure 16 shows the 50 metre resolution initial raster displayed in Figure 9. Similarly values within a 400 metre radius being generalised back to each cell using a distance weighted function. The pattern in Figures 15 and 16 are very similar. Both figures are produced to make the point that it is not necessary to go any smaller than a 100 metre initial resolution at the Leeds LAD scale for this kind of map in order to see the distribution clearly.

5. Discussion and conclusion

For a discussion of the results and a conclusion please see the full paper which is available online via the following URL:

http://bit.ly/GISRUK2015

6. Acknowledgements

This work has been and continues to be supported by the University of Leeds School of Geography. Some seed corn funding was provided by the ESRC funded NCRM Node Talisman which allowed collaborations with Advice Leeds to be started. Further funding was provided to continue this project under the EPSRC Connected Communities + Network as the Digital Welfare Project. This project would not be possible without the data and collaboration with Advice Leeds and Leeds City Council.

7. Biography

Stuart Hodkinson is a Lecturer in Critical Urban Geography. His main research focus is on the 'new urban enclosures' with a specific interest in the politics, policies and day-to-day realities of housing privatisation, urban regeneration and state-led gentrification in the UK.

Andy Turner is a researcher specialising in computational geography and research ethics. In addition to working on research projects, he is involved in teaching and administration. Andy is a highly skilled Java programmer and has a background in mathematics and e-Research.

References

Openshaw, S. (1984) The modifiable areal unit problem. Norwich: Geo Books. https://alexsingleton.files.wordpress.com/2014/09/38-maup-openshaw.pdf

Openshaw, S., Turner, A.G.D. (2001) 'Disaggregative Spatial Interpolation'. Paper presented at The GIS Research UK 9th Annual Conference, University of Glamorgan, Wales, UK.

Turner, A.G.D. (2015) Digital Welfare Project Source Code Web Page: http://www.geog.leeds.ac.uk/people/a.turner/src/andyt/java/projects/DigitalWelfare/

Turner, A.G.D. (2014) Digital Welfare Project Web Page: http://www.geog.leeds.ac.uk/people/a.turner/projects/DigitalWelfare/

Turner, A.G.D. (2000) Density Data Generation for Spatial Data Mining Applications. Paper

presented at The 5th International Conference on GeoComputation, University of Greenwich, UK.