Escaping the pushpin paradigm in geographic information science: (re)presenting national crime data

Alex Singleton* and Chris Brunsdon*,**

*Department of Geography and Planning, University of Liverpool, Liverpool L69 3BX
Email: Alex.Singleton@liverpool.ac.uk

**National Centre for Geocomputation, National University of Ireland, Maynooth

Revised manuscript received 27 November 2013

In 2011 the Home Office released the police.uk website, which provided a high-resolution map of recent crime data for the national extents of England, Wales and Northern Ireland. Through this free service, crimes were represented as points plotted on top of a Google map, visible down to a street level of resolution. However, in order to maintain confidentiality and to comply with data disclosure legislation, individual-level crimes were aggregated into points that represented clusters of events that were located over a series of streets. However, with aggregation the representation of crimes as points becomes problematic, engendering spurious precision over where crimes occurred. Given obvious public sensitivity to such information, there are social imperatives for appropriate representation of crime data, and as such, in this paper we present a method of translating the ‘point’ crime events into a new representational form that is tied to street network geography; presenting these results in an alternate national crime mapping portal http://www.policestreets.co.uk.

Key words: neogeography, GIS, policing, Geoviz

Introduction

The ‘street-level’ mapping on the police.uk website was launched by the Home Office in 2011 to provide a public portal for high-resolution spatial data pertaining to occurrences of recorded crime. These data were consolidated from across different Police Force areas and were presented both as areal aggregates for a zonal geography and also as ‘street-level’ point data. The data were made available to download as comma separated files (CSV), and the point data also presented within a public online mapping system. The police.uk website emerged within the context of a wider political discourse surrounding the transparency and accountability of public services. Furthermore, previous efforts to disseminate statistics about crime had drawn criticism, with independent review citing that the public lacked confidence in these measures, and that the geographic granularity should be improved (Statistics Commission 2006). More generally, as argued elsewhere (Chainey and Tompson 2012), the availability of data related to recorded crime are especially pertinent given their role for community empowerment through a bottom up approach to police accountability; and as such, can go some way to addressing the ‘reassurance gap’ in policing, which is where the public fear of being a victim of crime is misaligned with the probability that such events might occur (Millie and Herrington 2005; Quinton and Tuffin 2007). The imperative of resolving these issues was further amplified by the 2008 Policing Green Paper, which mandated that all police forces would be required to publish crime maps for their localities (Home Office 2008), a point later clarified as being required at ‘street level’ (Home Office 2010). It has been argued elsewhere that, although of policy utility, a theoretical rationale is lacking to underpin those decisions related to the purpose and design of police.uk (Chainey and Tompson 2012). The concern of this paper is not to replicate such discussions about the purpose, merit and ethical constraints of online crime mapping, which are adequately explored in the literature elsewhere (Weisburd et al. 2009; Chainey and Tompson 2012); however, instead we present a critique of how the police.uk website translates geographic data about crimes into information, and outlines a potential solution to mitigate some of the issues we identify. As
others have argued (Veregin 2011), there is a need within the GIS and cartographic literature to address effective communication in online mapping, not simply the technicalities of creating such representations (Field 2008; Field and O’Brien 2010).

Mapping with a pushpin paradigm
At the launch of the revised police.uk website in 2011, ‘street-level’ mapping of crime was enabled, which has generated considerable public interest with 53 million visits to date (Home Office 2013). Users of the website can search by unit postcode (similar to a zip code in the USA) and return results mapped on a backdrop of OpenStreetMap for a proximal location to the searched postcode (see Figure 1a). Over time the data and interface have been refined, for example, a switch from Google Maps to an open source alternative in line with wider changes in contemporary online cartography (Gale 2013; Virrantaus et al. 2009), introduction of the current status of criminal investigation for crime events, increasing the volume of points used to encapsulate non-street locations.
(e.g. train stations), and enabling tools that allow custom search locations to be drawn on the map. Interaction with the map follows navigation typical of ‘slippy maps’ popularised since the launch of Google Maps in 2005 (Batty et al. 2010; Haklay et al. 2008; Haklay 2013; Miller 2006). Such interfaces have enabled users to both pan and zoom by clicking and dragging on the map without having to refresh the web browser window; the ‘slippy’ term refers to the impression given of a window to a large and seamless map. As highlighted in Figure 1a, crimes are represented as points, which at higher levels of zoom are clustered into aggregate symbols representing multiple composite points, an example of both model and graphic generalisation (Gaffuri 2011). In previous iterations of the website, scaled ‘Radar’ symbology was used to convey that a point purported to an area (Chainey and Tompson 2012), whereas in the latest website these have now been replaced by a circle with a transparent border. In both cases, a numerical reference is added to indicate the number of aggregated points, in addition to the scaling of the symbols over a series of different sizes. As the map is progressively zoomed in, the density of points increases, and the symbol size decreases as the number of individual crimes aggregated by the points is reduced up to a maximum ‘street level’. At this scale, further levels of zoom create no further disaggregation of the points.
leaving a representation such as that visible in Figure 1b where multiple crimes are associated with single points.

In the latest version of the website, between zoom levels crime points are also animated to aggregate or disaggregate into new symbol locations, increasing symbol size by the volume of aggregated points. This could be argued as an improvement on previous iterations of the website where the aggregation or disaggregation of points between zoom levels was less explicit. However, there is also concern, that such automated generalisation could lead to ‘selective attention’ (MacEachren 1995; Lloyd 2005) on specific areas of the map; a concern particularly problematic where changes in zoom impact the spatial location of the point symbols alongside scaling. Clicking a point reveals a dialog box specifying ‘Crime on or near road name’, and a listed disaggregation of the number of crimes by their associated category. At launch, the points placed on the police.uk crime maps represented centroid locations of street segments, with individually geocoded crimes occurring on or near those centroids assigned to the aggregate points. More recently, these street network points have been extended to include a series of additional points of interest that encapsulate other places where crimes may occur, for example, sports and recreation areas, hospitals or shopping centres.

Although the majority of crimes are geocoded with more precise locations, for example, a house that was burgled, or the location where a person was mugged; representation of these data at an individual level on a public map would contravene UK data protection legislation, given the obvious risk of individual disclosure. However, under different legislative constraints, individual-level crime data can be displayed on public maps. An example is illustrated on the US property website ‘Trulia.com’, which shows crime data disaggregated to the individual level; another is ‘CrimeMapping.com’. However, without changes to UK law, there is a requirement for websites such as police.uk to implement some degree of disclosure control. With both generalised and non-generalised crime mapping, issues of data quality underlie representations (McGranaghan 1993), in terms of the completeness (MacDonald 2002), precision and accuracy of the geocoding (Cope 2004; Wallace 2009). In using such systems, trust is placed by users that the data used for the representation are a true and accurate reflection of the actual geography of crime.

The disclosure control procedures for aggregation of geolocated individual crime events onto the centroid points rely on a measure of nearness. This was derived by the developers of the police.uk website using a Thiessen Polygon procedure that initially was drawn around all street segment centroids to create an initial zonal geography. As noted earlier, more recent enhancements to the website have introduced ‘points of interest’ locations for reporting on certain crime events, and these centroids are now also incorporated into this procedure. Within these zones, the number of address points were counted, and where these counts fell below a threshold (less than eight addresses), the zones were merged. Thus, more sparsely populated areas would typically contain larger zones. An approximation of this final zonal geography is illustrated in Figure 2. This was created by taking the unique centroid locations of crimes that were reported in the police.uk CSV street-level data, and then from these points, creating a set of Thiessen polygon zones. Individual crimes would be assigned to the centroid of their enclosing zone, and this forms a method of disclosure control. However, arguably, given that the occurrences of crimes can be mapped back into a zonal geography, the disclosure controls for ‘rare’ (either categories, or for crime events in typically low crime areas) have enhanced disclosure risk with this system. With other sensitive data such as the Census, manipulation of ‘rare’ events through record swapping with adjacent zones would occur to further mitigate disclosure potential; and it is not clear the extent to which such procedures have been implemented in this context. In some sense, this is also an issue with the requirement of ‘street-level’ mapping, when data disclosure makes the reality of such representations very challenging in a UK context.

However, the point mapping representation adopted by police.uk is problematic, purporting a spurious level of precision about a crime ‘event’ (actually a collection of events) (Chainey and Tompson 2012), and echoes more general concerns from the cartographic literature about the use of dot maps when displaying derived data (Krygier and Wood 2005). Furthermore, given that the choice of crime data representation is known to elicit variable responses related to an individual’s fear of crime (Groff et al. 2005), using an appropriate representation is critical if an underlying objective of providing these data is to address the reassurance gap. With point-based representation, there is a significant danger users will misinterpret crime events as occurring at exactly those points displayed on the map. Furthermore, given that the points represent a series of streets, if there were features not located within the streets that may be linked to higher rates of crime, then crime events close to such features would be displaced to the centroid location. To an extent this issue is reduced with the more recent addition of non-street points for some of these environmental features (e.g. pubs).

Given these issues, the use of points on police.uk may appear an odd choice, where introductory Geographic Information Science (GIS) courses typically comprise some component on appropriate choices of visualisation (DiBiase et al. 2006; Monmonier 1996). Concerns from
within the GIS community were rapidly voiced when the website was launched through various online blogs. However, it is important to note that this website is representative of developments in the wider geoweb that have emerged predominantly outside of academic geography (Crampton 2009; Haklay et al. 2008), and at a time when it has also been argued that geographers were exhibiting a declined interest in mapping (Dodge and Perkins 2008; Herb et al. 2009; Wheeler 1998).

Mapping along lines

As discussed earlier, a constraint on the development of the police.uk website was the requirement for crimes to be reported at street level, which in a literal sense would be very difficult to achieve under current UK data disclosure legislation. Precursor websites to police.uk, such as the Metropolitan Police crime-mapping portal visualised reported crime using choropleth maps. Although it could be argued that such maps offered a more appropriate representation of their underlying data, they are certainly not of street-level precision, nor do they offer the same resolution of display as some of the equivalent US services. As such, it is understandable why policymakers may have desired a website that offered a representation at a more disaggregated scale.

It has been suggested elsewhere (Chainey and Tompson 2012) that one alternative to the use of points on police.uk might be kernel density estimation (KDE), a technique commonly used in crime ‘hot spot’ analysis (Chainey 2005; Brunsdon 1995). In a practical sense this would be a little complex (although not impossible) to implement. For example, the Google Maps platform API 3.0 now offers hotspot rendering and on other platforms this technique could be possible with the pre-rendering of hotspot surfaces. In both cases, implementing the KDE technique would need access to the underlying ‘true’ point data, that is, those points relating to the addresses of the crimes rather than the generalised points representing centres of neighbourhoods. However, this would also have disclosure implications. The point data as supplied by police.uk have displacement from their true locations, and as such, distance between those general points as supplied by police.uk would not reflect the distances between actual crime locations in space – this would impact on the reliability of the KDE surface – particular when zoomed in.

Figure 2 An example of Thiessen polygons created from crime centroids for an area of Liverpool

Source: Authors’ own
at a neighbourhood level. Furthermore, and perhaps more importantly, when KDE techniques are implemented on the basis of points that are positioned along a network, there is a tendency towards producing biased estimates (Okabe et al. 2009). Standard KDE algorithms use ‘crow-fly’ distances, which do not account for those barriers lying between network geography (such as buildings) which all impact on how near points are to one another in street topology. Furthermore, street topology has an arguable influence on patterns of certain types of crime (Hillier 2004), but this would also not be accounted for in standard KDE. One alternative might be to implement network-based KDE (Okabe et al. 2009), where the KDE is carried out on points in a road network (and distances used in the calculations are computed in this way), rather than for at any point in geographical space. This can be implemented through the SANET software – although again, this would require more precise crime point locations than the generalised police.uk points.

Given the issues with KDE, we propose an alternative solution using street networks as the unit for display of crime events. We argue that streets represent the most appropriate cartographic unit of display given that the crimes are supplied at this resolution of geography; additionally, avoiding those issues of spurious precision associated with point mapping, or the lack of ‘street-level’ mapping that would be incurred through the use of choropleth or KDE approaches. However, the general public is not privy to raw crime data, and as such, in the creation of any demonstrator of alternative representations, there is reliance on those data provided through the police.uk data repository. These comprise downloadable CSV files with records of crime by category for each street network centroid location. The underlying Thiessen polygon geography utilised for the assignment of individual crimes into the centroid points was also not available; however, as illustrated previously in Figure 2, this geography can be estimated using an amalgamated set of points for all reported events since the publishing of the crime data began. An accompanying caveat to this method is that there may be centroids in the police.uk full database that have yet to appear in the public data because no crimes have occurred there since public reporting began, potentially overestimating the sizes of the true Thiessen polygons in areas of very low crime. Once a Thiessen polygon geography was derived, street networks were extracted from the Ordnance Survey Open Data product Meridian. This product was chosen because it is supplied under an open licence enabling reuse and with reasonably uniform structure; an alternative might have been to extract OpenStreetMap data, but to some extent the products are interchangeable. Streets were then split at the intersection with the Thiessen polygon zones to create a relationship between a collection of underlying streets, or parts of streets, and a zone that corresponds to one of the police.uk centroids. Through this relationship the crimes specified for the points were then attributed to the street network topology, and thus, a collection of streets could be styled to reflect different rates of crime. These analyses were conducted using a PostGIS spatial database given that this is both open source and an efficient method of conducting this type of processing. Alternative spatial databases or more traditional GIS could also have been used for this task.

(Re)presenting the police.uk data at a street scale

With the crime data attributed to the street network, it was possible to use these features as the unit of display for a revised cartographic style. Relative rates of crime were calculated as the frequency of crimes within a category divided by the total street segment length within each Thiessen Polygon. These ratios were multiplied by a thousand to convert the rates into crimes per kilometre. As part of this process, all individual streets within each Thiessen Polygon were combined into a single feature, otherwise rates would appear artificially high on those composite street segments with shorter lengths as denominators would be smaller. Furthermore, given that information about which streets crimes were actually located on was lacking, the re-appropriation of the point data back into the zonal geography should only be used to apply styles to the whole Thiessen Polygon zone, again to avoid those issues of spurious precision that are exhibited by the point data.

The crime attributed street network data and rates were stored and processed within the PostGIS database. For display, these data were coupled with the map-rendering engine Mapnik that enabled the generation of map tiles with custom cartography. Using OpenLayers as a map interface, new cartography was developed and displayed on top of a neutral feature background map.

Two cartographic options were enabled to reflect the rates of crime: the first scaled the widths of the street network (see Figure 3a), and the second altered the street network by colour intensity (see Figure 3b). With regard to colour selection, a ColorBrewer Yellow-Orange-Red sequential nine step colour ramp was chosen for its aesthetic appeal and accessibility (Harrower and Brewer 2003). The purpose of showing different visualisation was to enable these to be evaluated by stakeholders at a later stage, enabling different styling options to be easily demonstrated. The ability to adapt cartographic styles, including other advanced features such as the selection of a scaling factor to adjust line widths or colour intensity, was only available on the full map view. We argue that in both the line width and coloured street cartographic styles,
these representations hold greater utility for interpretation than the display of points, and convey the lack of spatial accuracy due to disclosure control in a more appropriate way. At present the policestreets.co.uk website excludes those crimes that were recorded at those non-street centroids that were more recently added to the police.uk source data. These could however be integrated into the representation by aggregation into a revised Thiessen polygon geography, or, more appropriately, visualised as either points, lines or polygons, depending on the nature of the recorded feature.

When a user visits policestreets.co.uk, search is enabled by input of a full postcode, and the initial screen shown incorporates a more limited map view, highlighting a mile radius around the searched postcode (Figure 4). In addition to the map, those crime points falling within a one-mile radius of the postcode are aggregated for a rolling six-month interval, and tabulated according to absolute crimes by type per month. A trend rate is calculated by comparing the first and latter three months to give an indication of change over the course of the six-month period. Pink to green colours are also used to indicate the...
directionality and intensity of the percentage change. The limited map view enables the display of different crime types and durations. Furthermore, when the ‘slippy map’ is moved, the change of focus is detected and the user is asked whether they wish to refresh the table of results. In addition to the crime data, a call is sent to the police.uk API13 with a request for the contact details of the neighbourhood policing team who are responsible for the searched area; with the purpose of stimulating greater public engagement. This feature could be expanded in the future to incorporate an emailing system that might extract some statistics from the website, enabling stakeholders to send these to the neighbourhood policing team alongside further commentary related to the observed patterns, thus providing a community-based contribution to Problem Oriented Policing (Goldstein 1979 1990), where policing attention is encouraged to focus on underlying causes of events rather than the servicing of individual callouts viewed in isolation.

Concluding comments

In a general sense the police.uk website should be applauded for provision of such large volumes of open

Figure 3  Continued

-This website was built by academics at the University of Liverpool to demonstrate alternative visualisation options for the police.uk website. If you have any problems viewing the website then please get in touch with Alex Singleton.
data in a useful and well-documented format; and indeed, the work presented here is testament to this availability, as without it, the research would not have been possible, and certainly not within the public domain. However, this paper has illustrated representational issues with the mapping component of the police.uk website framing these discussions within wider debates related to the development and technologies of the geoweb, and furthermore, those political influences on the specification of ‘street-level’ mapping. The principal cartographic issue with police.uk we argue relates to the use of points to map aggregations of crimes, giving the users of these maps the impression of a spurious level of precision. This is exacerbated as further aggregation that occurs across the multiple scales of zoom.

A unique contribution of this paper has been to illustrate how the aggregation strategy used to assemble raw crime data into street centroids can be reverse engineered. An estimated set of polygon zones was created as an approximate to the geography used by the Home Office for the aggregation of individual crimes into street centroids. We make the argument that streets are a more appropriate unit of cartographic display because they match the geography for which the data are supplied more closely than generalised points. To demonstrate the cartographic potential of such data using open source geoweb

Figure 4  Search results showing the limited map view alongside the change analysis table
Source: www.policestreets.co.uk/Authors’ own
infrastructure, street topology was used to display rates of crime as coloured or scaled lines. These two new cartographies illustrated the intensity of criminal activity within a set of streets, but also apply a geographic resolution no more precise than the source data. These maps were integrated into the www.policestreets.co.uk crime-mapping portal and enable users to search and interact with these new cartographic representations, alongside presentation of a tabular display of absolute and relative crime rate changes within a mile radius of the search location or map centre. Further contextual details were provided through the police.uk API, enabling website users to contact the neighbourhood policing teams allocated to the search area.

The purpose of this work is to stimulate debate on the appropriateness of current public crime mapping, and to illustrate how new open source cartographic tools can create representations that are more appropriate to the underlying data. There are clearly areas for substantive and methodological extension that have been highlighted by this work. One clear disadvantage of the created system is that for each website query, the present solution renders raster tiles for each website visitor, which is computationally intensive, and not especially well suited to high volumes of traffic. However, as new vector mapping technology becomes available, the scaling or colouring of street topology could be completed in a more computationally efficient manner on a user’s web browser, thus increasing the website scalability. Further methodological refinement could involve the use of the actual zonal boundaries employed by the Home Office, or a lookup to the streets that are aggregated to each point. There are other cartographic challenges surrounding how to incorporate the non-street points that were added by recent updates to the website, and which are currently excluded from the site. From a substantive perspective, the representations created for this project would fit within a wider evaluation concerning how the public perceives different crime mapping techniques, and indeed, how these representations impact on their fear of crime. Finally, offsetting appropriate disclosure constraints within the political confines of providing ‘street-level’ mapping will remain a challenge from a technical perspective about how these are handled statistically, but also for policymakers, because these demands are out of line with other open data dissemination procedures implemented more generally across the public sector.

Acknowledgements

We would like to thank the anonymous reviewers and the editors for their useful suggestions during review of this manuscript. We would also like to especially thank Oliver O’Brien, UCL (http://oobrien.com/), for technical assistance during this project.

Notes

1 The latest iteration of the police.uk website uses Leaflet as the display interface: http://leafletjs.com
2 A full list of the additional categories is provided on the police.uk website: http://web.archive.org/web/20121010233724/http://www.police.uk/anonymous-anonymisation-methodology
3 See http://www.trulia.com/crime/
4 See http://www.crimemapping.com
6 See http://maps.met.police.uk/
7 See http://sanet.csis.u-tokyo.ac.jp/
8 See http://data.police.uk
9 See https://www.ordnancesurvey.co.uk/opendatadownload/products.html
10 See http://postgis.net/
11 See http://mapnik.org/
12 See http://openlayers.org/
13 See http://data.police.uk/docs/
14 As an example of vector mapping see: http://www.mapbox.com/blog/vector-tiles/ and the new Google Maps: http://www.google.co.uk/maps/about/explore/

References

Chainey S 2005 GIS and crime mapping Wiley, Chichester
Chainey S and Tompson L 2012 Engagement, empowerment and transparency: publishing crime statistics using online crime mapping Policing 6 228–39
Cope N 2004 Intelligence led policing or policing led intelligence? Integrating volume crime analysis into policing British Journal of Criminology 44 188–203
Crampton J 2009 Cartography: maps 2.0 Progress in Human Geography 33 91–100
Dodge M and Perkins C 2008 Reclaiming the map: British geography and ambivalent cartographic practice Environment and Planning A 40 1271–6
Field K 2008 Maps, mashups and smashups The Cartographic Journal 45 241–5
Gaffuri J 2011 Improving web mapping with generalization Cartographica 46 83–91

Oobrien.com/
Gale G 2013 Push pins, dots, customisation, brands and services: the three waves of making digital maps The Cartographic Journal 50 155–60
Haklay M 2013 Neogeography and the delusion of democratisation Environment and Planning A 45 55–69
Haklay M, Singleton A and Parker C 2008 Web mapping 2.0: the neogeography of the GeoWeb Geography Compass 2 2011–39
Harrower M and Brewer C A 2003 ColorBrewer.org: an online tool for selecting colour schemes for maps The Cartographic Journal 40 27–37
Hillier B 2004 Can streets be made safe? Urban Design International 9 31–45
Home Office 2008 From neighbourhood to the national: policing our communities together Home Office, London
Lloyd E L 2005 Attention on maps Cartographic Perspectives 52 28–57
MacDonald Z 2002 Official crime statistics: their use and interpretation The Economic Journal 122 F85–F106
McGranaghan M 1993 A cartographic view of spatial data quality Cartographica 30 8–19
Miller C 2006 A beast in the field: the Google maps mashup as GIS/2 Cartographica 41 187–99
Millie A and Herrington V 2005 Bridging the gap: understanding reassurance policing The Howard Journal of Criminal Justice 44 41–56
Monmonier M 1996 How to lie with maps 2nd edn University of Chicago Press, Chicago IL
Quinton P and Tuffin R 2007 Neighbourhood change: the impact of the national reassurance policing programme Policing 1 149–60
Veregin H 2011 GIS and geoenabled cartography Cartography and Geographic Information Science 38 286–8
Virrantaus K, Fairbairn D and Kraak M-J 2009 ICA research agenda on cartography and GIScience Cartography and Geographic Information Science 36 209–22
Weisburd D, Bernasco W and Bruinsma G 2009 Putting crime in its place: units of analysis in geographic criminology Springer, New York

© 2014 The Authors. Area published by John Wiley & Sons Ltd on behalf of Royal Geographical Society (with the Institute of British Geographers)