UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL

WORKING PAPERS SERIES

Paper 190 - Apr 13

Visualising Spatial and Social Media

ISSN 1467-1298

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Visualising Spatial and Social Media[‡]

Michael Batty, Steven Gray, Andrew Hudson-Smith,
Richard Milton, Oliver O'Brien and Flora Roumpani

Centre for Advanced Spatial Analysis, University College London,
90 Tottenham Court Road, London W1N 6TR, UK

{m.batty}{steven.gray}{andrew.hudson-smith}{richard.milton}
{o.obrien}{flora.roumpani.11}@ucl.ac.uk

Thursday, 28 March 2013

Abstract

In this chapter, we begin by surveying the development of computer graphics as it has influenced the development of the spatial representation of social and economic data, charting the history of computer cartography and geographic information systems (GIS) which have broadened into a wide array of forms for scientific visualisation. With the advent of the World Wide Web and the widespread adoption of graphical user interfaces (GUIs) to most kinds of computer device, visualisation has become central to most sciences and to the dissemination of many kinds of data and information. We divide our treatment of this domain according to three themes. First we examine how the 2-dimensional map has become key to many kinds of spatial representation, showing how this software has moved from the desktop to the web as well as how 2-d has moved to 3-d in terms of the visualisation of maps. Second, we explore how social data is being augmented by space-time series generated in real time and show how such real-time streaming of data presents problems and opportunities in which visualisation is key. We illustrate these new data for basic feeds from cities but then move to examine data from transit systems, social media, and data that is pulled from the crowd – crowdsourcing. Finally we note the development of visual analytics showing how 2-d and 3-d spatial representations are essential to interpreting the outputs and the workings of more complex models and simulations. We conclude with the notion that much of what we develop in this chapter for the space-time domain is generic to the future representation of all kinds of social data.

‡ This is a draft chapter for the book: Procter, R. and Halfpenny, P. (Editors) *Innovations in Digital Research Methods*, Sage Publishing Company, London, forthcoming, 2014

Defining Visualisation

Visualisation as a distinct activity in computing emerged in the mid-1980s, largely because computer memories on large-scale computing devices, particularly minis and workstations became increasingly associated with visual images. At the other end of the spectrum, personal computers acquired visual interfaces through their memories almost as soon as they emerged in the late 1970s and this gave a massive boost to computer graphics which hitherto was largely generated 'offline', once computation had been done. The outputs generated then were in forms that could be plotted directly on line plotters or displayed on storage tubes, normally separate from the devices that had produced the outputs in the first place. To an extent, visualisation was first associated with large-scale systems which generated what at the time was 'big data' and which hitherto had rarely if ever been cast as imagery. In this sense, the initial excitement of visualisation was that it provided opportunities for entirely new insights into scientific phenomena and it is no surprise that the first attempts in the early 1980s were labelled 'scientific visualisation' and associated with massive physical systems, from galaxies to particles (Smarr, 1991).

Visualisation is now defined much more widely for it pertains to the entire array of imagery that is associated with computation and this of course includes our access to computers which is now dominated by visual interfaces, often called GUIs (graphical user interfaces). Here we will cast our net quite widely but visualisation is now so dominant a force in digital media that we need to be clear about the limits to the range of ideas we will introduce. Our focus will be very much on the spatial, meaning geographical and architectural representations of social and economic data, laced with physical representations of people and places, and drawing on the extensive interest in how people and places interact. In this context, our focus will be on networks as well as on ways in which we are able to make sense of social and spatial data using various kinds of visual analytics. This latter term we define as the application of analytical ideas from statistical analysis primarily to visual images and the patterns that they display. This will become clear as we introduce examples to show the reader what is possible.

Almost from the first digital computers, graphics was important to usage. The original computer memories based on vacuum tubes could display a primitive kind of graphics (Dyson, 2012) but the various ways in which the workings of computers were monitored using oscilloscopes lead immediately to programs that displayed graphical output. These storage tube devices were slowly developed during the 1950s and 1960s – an example being the Chicago Area Transportation Study's display of trip data on the 'Cartographatron' (Plummer, 2003) – and by the 1970s devices such as those built by Tectronics were quite widely available. However it was line plotters that first dominated graphics where vector drawing on drum and flatbeds were used to develop various engineering drawings and rudimentary computer cartography. Then to speed up these kinds of output, line printers were programmed to overprint characters to rapidly produce exploratory outputs. The Symbol/Synographic Mapping program (SYMAP) developed by Alan Schmidt in the late 1960s was widely used in landscape planning (Chrisman, 2006) while Mandelbrot reports similar usage at IBM for the production of fractal images of mountainscapes around the same time (Mandelbrot, 2012). However it was not until the development of machines with specific screen memories devoted to graphics that visualisation really began to take off. This had to await the development of microcomputers in the late 1970s although in parallel, as computers were scaling down to minis and workstations, interfaces became increasingly graphic in and of themselves as pioneered at places like Xerox Parc and thence embedded in early versions of Apple's Macintosh.

As the graphics revolution intensified, the first major developments came in real-time animation, in virtual reality technologies which initially enabled users to interact with graphic scenes either passively in VR theatres or more actively in closed environments such as CAVES. Much of this then moved into the domain of virtual worlds in which users interact with other users usually remotely across the internet but using software that was downloaded to the desktop. These worlds too are now less significant then they were in the early 2000s but developments continue to be driven by two key forces: the development of games of various sorts from entertainment games such as the shoot 'em up variety to educational games and thence to real gaming environments which are now called 'serious games'. The other force in all of this is, of course, the internet – the web – which when it was first developed became graphic very quickly largely due to visualisation technologies pioneered at the supercomputer centres which were the cradles of scientific visualisation. The interfaces to the web were essentially graphic from the first instance and it is possible that its massive proliferation is due to the fact that the graphical interface enables user-friendly access. And so of course it was with the development of the personal computer. What is currently happening is a force driving software from the desktop to the web, and much of what we will say in this chapter relates to the fact that virtually everything we see on the desktop is, in principle, portable to the web. The evolution, for example, of geographic information systems (GIS) technologies has followed this path: mainframe, mini, workstation, PC – the last three essentially desktops – to the passive web (Web 1) and thence to the active/interactive web (Web2) and on to the semantic web (Web 3 and beyond).

In the last decade, various other visualisations of non-spatial data have emerged, the most notable being methods for displaying networks and various statistical graphs. These emerged almost prior to the revolution in scientific visualisation, for example from exploratory data analysis following Tukey's (1977) pioneering work and Tufte's (1992) popularisation of information graphics. In fact although there were rapid developments in graphics such as parallel plots, linked maps and charts, and in flow graphics, the real advances have come more recently in network analysis where there are now some impressive online packages. Web-based systems such as IBM's Many Eyes where the focus is on many different and alternative visualisations (http://www-958.ibm.com/software/analytics/manyeyes/) such as treemaps., bubble diagrams, tag clouds, etc. dominate the field. We will show some of these here but the array of new graphics is now very large and although many developments centre on a tight set of techniques, there is considerable ingenuity currently being employed in extending this range (Lima, 2011).

To summarise these developments, the elaboration of hardware has been one of providing computable devices to display data and information, from raw to processed and to interact with these using graphics interfaces; the development of software has become increasingly graphic in terms of how we display data, predictions and how we interact with the hardware; software has moved very quickly off-shore so to speak – to servers where data is increasingly stored remotely and now no longer locally but globally in the Cloud; and data itself in our field has suddenly jumped by orders of magnitude in size largely because computers have reached the point where they are being embedded in cities to capture data that is streamed in real time. In short it might be argued that in the domain of the social sciences, there is a sea change occurring in the type, size, storage and mining of data from a world where data was cross-sectional, static, to one where time is of the essence. All these themes are resonant in the discussion that follows.

What is currently happening in spatial and social media is confusing because there are so many interacting trends and innovations that making sense of all this is difficult. The challenge we address in this chapter is primarily to look at social data in the spatial domain and this does help bound our scope a little. But virtually every aspect of society that can now be represented digitally, can be visualised in many different ways and to make sense of all this, we will divide our discussion into the following sections. First, we begin with traditional spatial data – with static maps drawn largely from sources which may be digital but also hand-crafted rather than real-time data – and we show how the software for handling such data is moving from desktop GIS to the web. This is part of our focus on a 2-dimensional world and thence we move to 3 dimensions and broach visualisation of the built environment.

Second, we move to real-time data – data which is streamed in real time but is usually visualised after the event. Such data is what we refer to as 'big data', one definition of which is data that will not fit into the current generation of an Excel spreadsheet, thus requiring very different mining, sorting and visualisation methods than those used previously (Mayer-Schonberger and Cukier, 2013). We then shift to retrieving real-time data using digital techniques, introducing crowd sourcing but still linked to maps. At this point, it will become clear that many techniques that enable us to visualise such data interact with one another. For example, our crowdsourcing merges with our online map work while our online map work is used to visualise real-time data. Our third and last section deals with analytics – with modelling and simulation where we illustrate how traditional land use transportation models are utilising online portals and open web-based software such as Google Earth. We then speculate on how big data and new visualisation capabilities are changing the very systems that we are studying, thus impressing a central theme of this book: digital social media and the way we use and develop it is changing the very way we articulate the system of interest, thus illustrating the age old adage that 'more is different' (Anderson, 1972).

Visualising Map Data: From 2-d to 3-d, from the Desktop to the Web

The Rise of Geographic Information Systems and Science

Geographic information systems (GIS) emerged as a synthesis of computer cartography, usually plot or vector-line based with pixel or raster-based gridded spatial data, which in turn were underpinned by rudimentary spatial database technologies and spatial statistical analysis (now called spatial analysis). These developments began in the 1960s and proceeded apace in the 1970s but only came to fruition as formalised software systems in the mid to late 1980s when graphics screen technologies replaced offline plotting capabilities that had dominated the field hitherto. By the mid 1990s and after the release of Windows 95 which really brought graphical user interfaces to the PC for the first time, desktop systems became widespread with many vendors competing for the market.

In parallel, mapping systems emerged as the web took hold and by the late 1990s, several online mapping systems generally produced for purposes of navigation associated with some advertising emerged. But it was not until 2005 that Google Maps was released which became the *de facto* online mapping system of choice. However the move from purely passive web-based applications which the original map systems were focussed upon, that is providing information for the user, to more interactive systems which could be hacked to take additional layers and customised for various applications, also took place during the early 2000s. 'Map Mashups' took hold when Web 2 – interactive web applications – began to take off (Batty, Hudson-Smith, Milton, and Crooks, 2010). As these developments continued, GIS began to move onto the web with internet map servers being common for large-scale professional applications but with the real focus in this area on adding spatial analysis to such systems through plug-ins and various add-ons which customise such systems for very large-scale applications.

Three other developments are significant; first the emergence of crowd-sourced maps, particularly Open Street Map (OSM) which is now competing with many mapping agencies and provides a serious alternative to Google Maps and other online systems; second the development of online and often free GIS systems such as Quantum GIS with many variants of these types of application at present; and third, the recent move of large-scale GIS systems to entire platforms that might be integrated with other web-based services such as the Cloud for remote storage, enterprise systems for large-scale corporate transactions processing, and real-time GIS for problems dominated by the online streaming of data sets.

Open Mapping and the Power of the Web

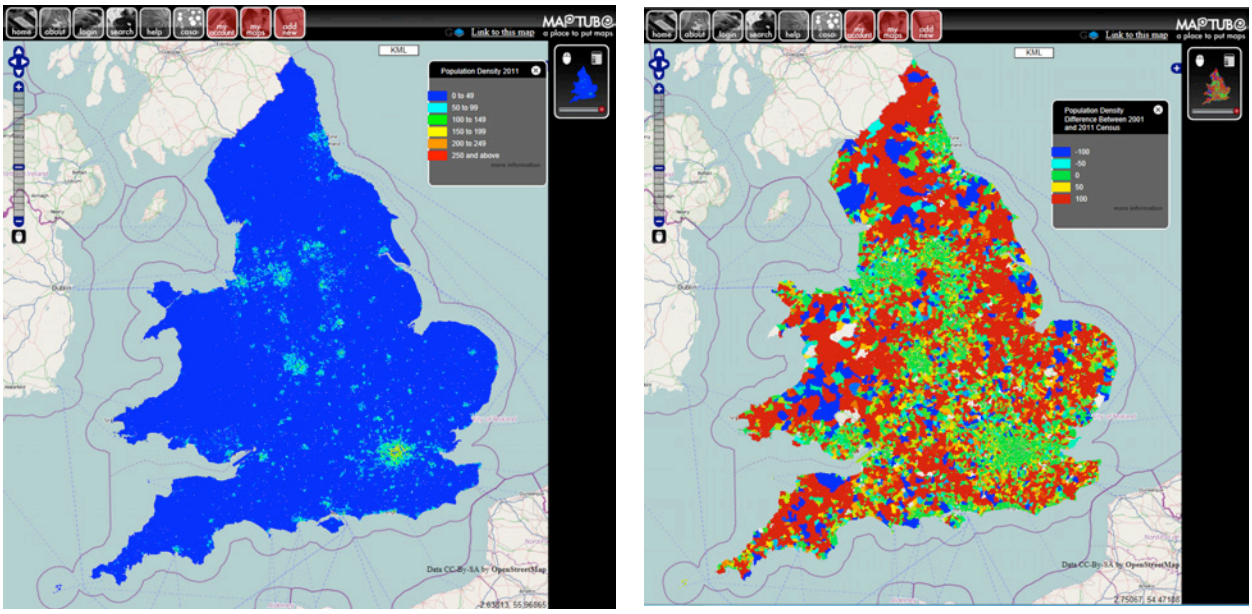
Desktop systems for GIS still tend to be more powerful than web-based systems but web-based systems enable much more integrated storage facilities. A good example of a web-based map analysis and storage system which is to an extent a one stop shop for simple spatial analysis and map display is in our own online resource MapTube (www.maptube.org). MapTube is a portal which contains access first to software that enables any user to convert their own map to a form that can be layered across a non-proprietary map base such as OSM or Google Maps. The software lets the user convert their data file to the map layer and then provides them with a web page that displays their map on top of one of these map bases. As the user may not own the copyright of the map data in question, MapTube only stores the web page – the URL (Universal Resource Locator) – where the user locates the map they have produced using the free MapTube software, and in this way using MapTube to access the map in question does not infringe any copyright. MapTube thus acts as an archive for maps as well and once the user creates the map, the user is asked to share the URL in the portal. If it is clear that copyright is not being broken, the map itself is stored on the MapTube web site. MapTube also contains a variety of simple spatial analytic tools. Obviously building up map layers which is key to GIS – that is displaying different map layers on top of one another, thus searching for spatial patterns – is a major function of the software. This is akin to exploratory data analysis as evolved from Tukey's (1977) work but the software is also able to produce various simple indices and queries related to the map data and also to output certain quantitative data on various map attributes that define the various layers in question.

In MapTube, the most common representation is the choropleth map, where a boundary file is used with polygon areas coloured according to a value. Alternative representations of data might use lines to represent flows of traffic on roads while point-based data offers great variation as both colour and size can be used to show the information, along with different icons or even pie-charts to show multivariate data at a point location. Finally, hybrid visualisations, like heat maps and gravitational interpolation, exist where point-based data is processed into a continuous space image. The aim in such contexts is to show how the data decays over distance from a source. Contour maps are closely related to heat maps and show lines of equal value or iso-surfaces as commonly seen in weather maps. Many of these features are invoked in MapTube but one must go to more comprehensive desktop GIS systems to enable the full range of such cartographic and display techniques to be invoked.

CASA's MapTube website, launched in 2008, provides geospatial data handling and visualisation functions as a free service, so the process of having to source boundary files, join with your own data and then build a visualisation from scratch is simplified. Using for example the UK's 2001 Census, there are only a limited number of geometries which make sense when visualising data and the spatial boundary files are generally an order of magnitude larger than the aspatial data. Due to the mismatch in sizes and re-use of boundary files, it makes sense to hold them on the server and so enable people to make maps by supplying only the aspatial attribute data. This is an approach that has been mirrored in Google's Fusion Tables and is a form of geocoding with spatial join functionality. Taking this a step further, by defining a dataset as a URI endpoint, maps can be built from raw data directly from the Internet, for example from data in various public (open) datastores. Using MapTube's automatic map generation functionality, 145 variables can be mapped from the 2001 Census and now 2,558 variables from the latest 2011 Census.

As an example, population density maps from the 2001 and 2011 Census can be mapped using MapTube's CSV (Comma-Separated Value) upload functionality. The maps were built at "Lower Super Output Area" level (LSOA) of which there are 34,753 in England and Wales. Raw data, for example, can be downloaded from the Census website as an Excel spreadsheet and a small amount of editing turns it into a table containing just a header row and rows of data values which can be uploaded to MapTube where the map column is automatically identified as one of the boundary datasets stored on the server, so the data can be plotted on the map immediately as we show in Figure 1(a) for the 2011 data.

Once data has been published on MapTube, visual comparisons between maps are possible by adding multiple layers and using transparency. In this way, maps can be compared even if they are not using the same geometry. Although the 2011 map in Figure 1 (a) is almost identical to the 2001 map at the country level but by zooming in towards the street level, significant differences can be detected, which is the main advantage of using a web-based map of this type. The scale doubles at each successive zoom level, so between zoom levels zero and 21, it is possible to represent detail that would be impossible on a paper map for an area this size. This is so obvious a point that it barely seems worth stressing but the real advantages of web-based systems (and desktop too) is that users can search for pattern at different levels as we show for population density in Figure 2.



a) b)

Figure 1: CASA's MapTube Website (http://www.maptube.org/)
Showing Population Density in 2011 (a) and Changes in Density 2011-2001 (b)

a) Plotted as 'Persons per Hectare', from the 2011 UK Census table QS102. The data picks out the population centres of London, Birmingham, Liverpool, Manchester, Leeds and Newcastle but b) the differences plotted clearly distinguish declines in density in the large cities and increases in the rural and small town locations.

The two datasets used to make this example are simple CSV files and there are a whole host of open-source and free tools available that can be used to work with this type of data. With the right tool chain, it is possible to achieve considerably advanced analysis without much effort, so we can use any two such datasets to build our derived maps of change. Tools like Microsoft's Excel, Google Docs or OpenOffice can be used to edit CSV files, Google's Refine (formerly Freebase) can be used to clean data, and free databases like MySQL and PostgreSQL can be used for more complex analysis. By creating a database table for the 2001 and the 2011 data using PostgreSQL 9.2, the CSV files can be loaded using the 'Import' function on the PGAdmin console. Once the files have populated the two tables, a result set can be made which joins the two sets of data on the primary geography and attribute fields. Once the two tables have been joined and the result is available, the 'Export' function is used to save the data into another CSV file. In this way, population change per hectare which is plotted in Figure 1(b) is computed. One of the issues with data based on boundary files is that the boundaries change over time. In the case of the 2001 and 2011 data, this difference is very small, but there are a few white holes visible in Wales where the base map shows through. A good way round this is to grid the data, but this adds a layer of complex geospatial processing.

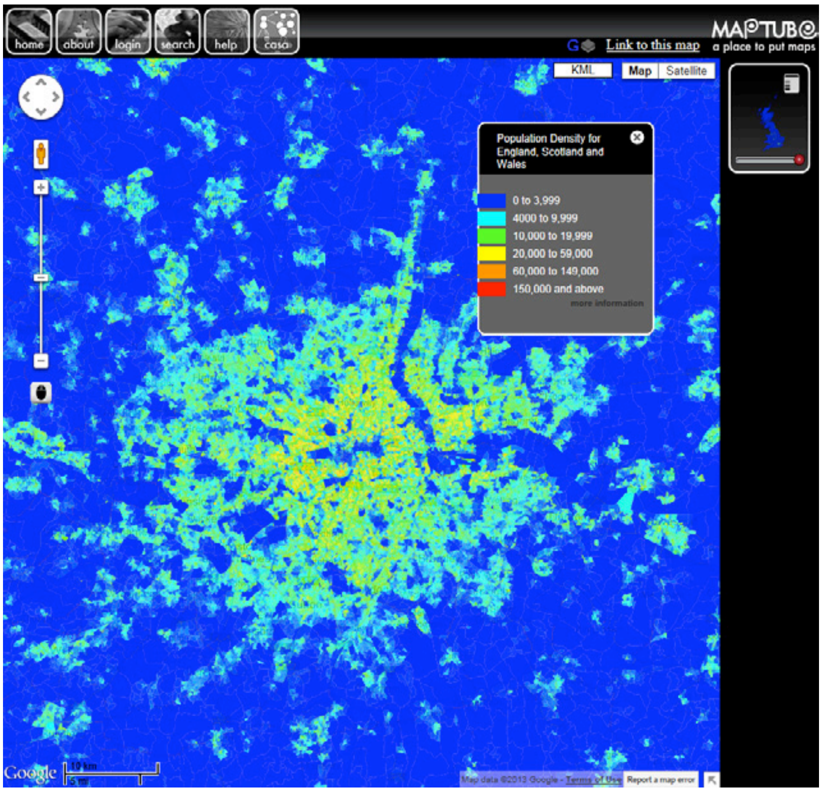


Figure 2: 2011 Population Density at the Metropolitan Scale in Greater London

The system also allows data to be displayed in any desktop GIS that implements the Open Geospatial Consortium's WMS standard (e.g. Quantum GIS or ArcGIS). It is this blurring of the boundaries between the traditional GIS methods and the new methods in geospatial analysis which are now emerging on the web that can potentially offer new insights for social research. With today's web-based technologies, many components of a traditional GIS system can be implemented as web services, with the advantage of being able to handle much larger datasets more easily. Libraries like Geotools are starting to move into the Cloud and run on distributed systems like Hadoop, offering GIS services with the power and scalability of Microsoft's Azure, Amazon's EC2 and Google's AppEngine. This type of infrastructure can still be implemented by organisations internally, using their own hardware together with open-source software, but the flexibility now exists to push big problems onto Cloud systems when local computing power is inadequate. Through virtualisation, computing power is now flexible enough that spare compute cycles can be harnessed to work on these big problems, perhaps running big geospatial workloads overnight on local computer systems and pushing really big workloads onto the Cloud as necessary.

In using computer mapping systems in general and GIS in particular, it is worth noting that there are recurrent problems of ensuring that the map base – how the map is projected into the plane – is understood. Following the release of Google Maps, all contemporary web mapping systems use the Universal Transverse Mercator projection, but with a spherical model of the Earth used for performance reasons. The benefits of using these systems are in the dissemination of research outputs to a wider audience via the Internet and the ease with which data can be placed into its spatial context. Base maps showing street layouts, cycle routes, satellite images or terrain maps are available using Google Maps, OSM or Bing Maps. The only requirement is the ability to draw the data over the base map in the correct location, which usually requires a coordinate re-projection. Most people are familiar with the WGS84 system of latitude and longitude expressed in degrees, which is a different coordinate system from the one used to make the base map. In the Google Maps API, points (as clickable pins), lines and polygons can be added as overlays using coordinates in WGS84, but only for a limited amount of data. The situation in OpenLayers is similar, but a conversion between the WGS84 data and Spherical Mercator map needs to be defined first. The main component of a web map is a web page containing a Javascript, Flash or Silverlight application to handle the drawing of the maps using image tiles or vector overlays, along with implementing the pan and zoom functions. There can be differing levels of complexity at this layer, with Google Maps API v3 targeting more lightweight applications that can run on smart phones, while OpenLayers is more of a Javascript client-side GIS containing functions to import data from KML, GML or JSON. It is also currently unique in its handling of different map projections and coordinate transformations, so it is correspondingly harder to use.

In order to make full use of web mapping systems for visualising social data, we need a way of drawing the data onto the map tiles and also an ability to compare maps from different sources of data. The base maps, which are required to give our data spatial context, are cartography, which is drawn using a set of rules for visualising features like roads and rivers. In contrast to this, the visualisations shown in Figures 1 and 2 uses a choropleth visualisation technique where areas in a boundary file are coloured according to their data value. In this case, the boundary file is the Census Output Areas file while the colour scale follows the 'person per hectare' field in the data. Here we make a distinction between a tile renderer that can handle cartography and a specialised tile renderer for handling data. While tile renderers designed for drawing the base maps can usually draw choropleths using a similar set of rules, there are often more complex visualisation techniques like heat maps and gravitational interpolation for visualising density which require specialised systems.

The Move from 2-d to 3-d: Populating the Built Environment with Social Data

The main way of displaying spatial data is the map, enshrined in the notion that the spatial variations of interest at all scales are best understood in the 2-dimensional thematic world of the point or choropleth map. This is based on the premise that the key feature of the spatial world is distance on the plane and that the third dimension contributes only a small fraction to the structuring and location of urban activity. In fact there is variation of spatial attributes in this dimension but it is nothing like as rich as the plane and it is only during the last one hundred years that this has become significant due to our ability to use new technologies to build upwards. Nevertheless, the third dimension offers a much richer sense of understanding the city in that non-experts identify more closely with pictures in this dimension than they do with maps. Simply in terms of communicating an understanding of some social variations in space, the third dimension is becoming more important. Moreover there are now substantial variations in space within tall buildings and building complexes which hitherto have almost gone uncharted. Here we will briefly introduce these developments which we expect to gather pace over the next decade.

The first models of cities (which we will present in the last section of this chapter) were symbolic – representing social and economic relationships in two dimensional space and explaining how different activities locate and interact using various social physics and econometric methods for articulating causal effects. It was only when graphics really took off that the notion of modelling the city as a set of geometric objects became possible. Very rapid strides were made by companies such as AutoCAD from the 1980s onwards and in the last decade, geometric models of buildings have been extended city-wide creating what have been called 'virtual city' models of which there are now many hundreds. In fact hard on the heels of Google Maps came Google Earth which enabled the 3-d geometry of buildings to be inserted into the software thus providing a ready made platform for building such models. A little before, GIS software had begun to extend itself into 3-d – first in the late 1990s with plug-ins like 3-d Analyst and ArcScene into ArcInfo and ArcGIS. Our own development of a 3-d block model for Greater London was operational by 2005 and was built using map parcel and street line data from the UK National Mapping Agency Ordnance Survey, supplemented by 3-d height data from LIDAR surveys made by InfoTerra, all synthesized within ArcGIS which provided the basic software for construction. Once models like this have been built, they can be exported into various other CAD (computer-aided design) software which enables more elaborate graphics to be generated.

To an extent, our interest here is in populating these geometric models with social data, that is, tagging key variables and attributes to buildings, noting for example things like working population, density and so on at the building or parcel level while also linking these to more material attributes such as energy, information use, traffic generated, pollution and so on. There are not many examples of this so far with most tagging smoothed map surfaces to 3-d representations. In Figure 3(a), we show our typical block model of London for a portion of the centre – the model contains some 3 million building blocks out to the M25 orbital road or beltway– but we have only succeeding in tagging a fraction of these blocks to socio-economic attributes. We have however begun to merge these 3-d geometries with 3-d surfaces as we show in Figure 3(b), for example, for air pollution (Batty and Hudson-Smith, 2005). In Figure 3(c), we show how we can utilise 3-d visualisations by building a crude geometry of the city using histograms. These are no longer buildings but levels of density or intensity – the 3-d equivalent (with respect to floor-space this time) of the maps in Figures 1 and 2 rooted to a grid interpolated from geometric boundaries such as LSOAs. We can fly around these maps and they do provide excellent ways not only to engage stakeholders and users of these data sets in policy issues but also to explore different possible patterns that come from looking at the city – at the data, that is – in 3-d city form from different perspectives, from different angles. In Figure 3(d), we show how land use can be tagged to this kind of media at the more local scale, thus bringing alive the traditional land use map which is coloured using the same shading as recommended in the 1947 Town and Country Planning Acts.

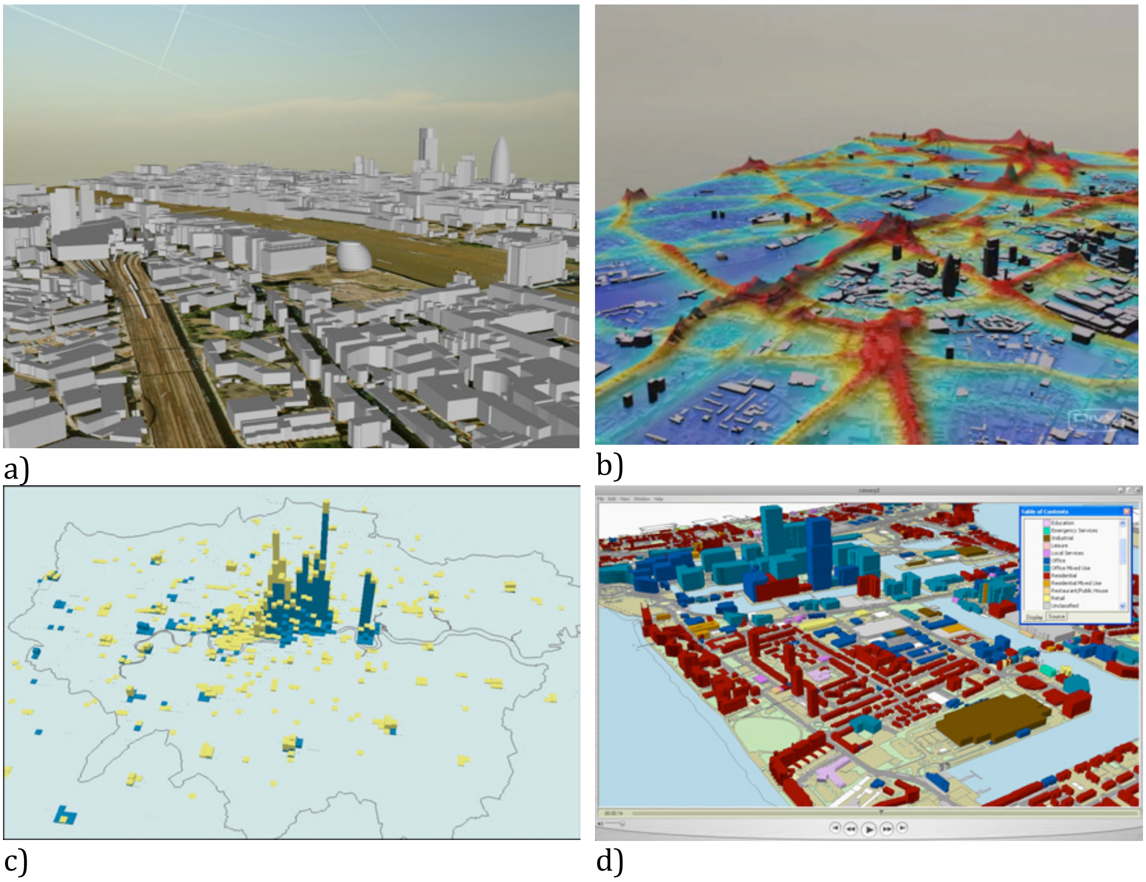


Figure 3: Moving to 3-d Visualisation and Navigating Through the Models

(a) The virtual London model built from vector land parcel, streetline and 3-d LIDAR data and visualised in ArcGIS (b) Layering a pollution surface (NO_2) across the model (c) Office (blue) and retail (yellow) floorspace visualised as 3-d histograms (d) Land use mapping in Canary Wharf, London's second central business district

There are many extensions of these kinds of 3-d visualisations. First we can embed these into virtual worlds, 3-d environments in which users or stakeholders can appear as avatars and actually walk in the space in which the 3-d city might be embedded. These worlds enable users to manipulate the scene. They are often considered as virtual exhibition spaces where users can engage in dialogue and debate but they might also be considered as virtual design studios for design over the net where the designers are remotely logged onto the world as we show in Figure 4(a). Here two avatars are positioning a high building from a menu of such buildings into the urban scene which is a portion of our Virtual London block model. These kinds of reality can be augmented on the spectrum from theatres where users engage passively in the scene to CAVES which are more interactive and separate from the real world, all the way to internalised head set types of device that embed the user completely in the scene. In fact many variants exist where there is a mix of immersion and external usage. For example, the notion of the 3-d world being accessible as a fly-through but in a theatre like context is one way of immersion in this media. In fact flying through 3-d models of the kind we have built for London actually within the software is extremely slow, still. Thus movies of such animations are often made and then shown off-line.

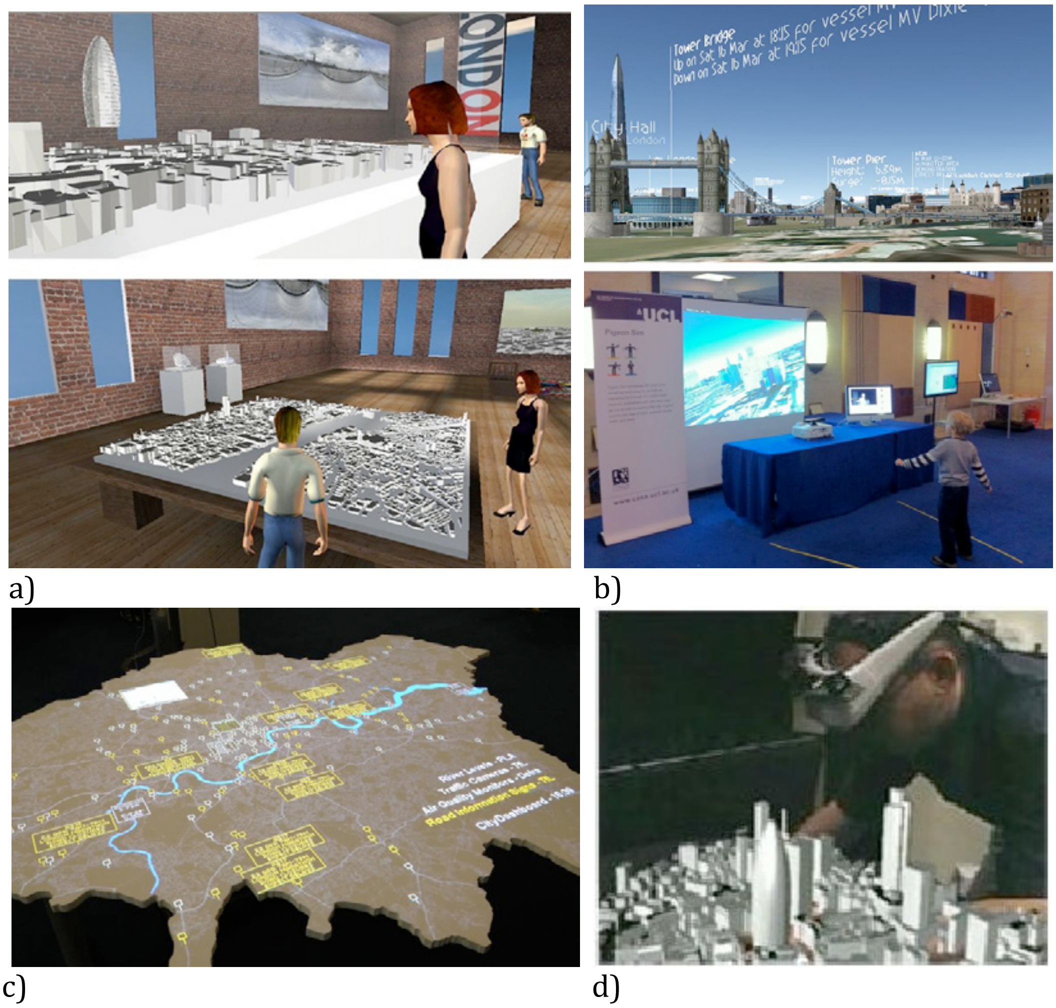


Figure 4: Augmenting 3-D Visualisation Merging the Virtual with the Real

a) Creating a virtual exhibition space/ design studio where users can interact in a virtual world b) Easier navigation of a virtual model using gaming media c) The London Data Table: projecting digital media back onto a material surface d) A holographic like projection as seen through the headset of one of the users interacting with another

In fact we have developed a fast method for such fly-through which we call PigeonSim illustrated in Figure 4(b) where a user is able to navigate through the model which is imported in Google Earth and then accessed using games technology, in this case through a Kinect motion sensing device that links an X-Box to a Windows application. Users can thus control their navigation in this way simply by standing in front of the box and controlling it with their body movements. We are also developing various other realisations of virtual media in more analogue like contexts where we consider that our digital media is best displayed in a more familiar way. Just as conventional material block models of the city are being augmented with digital inputs, we consider that augmenting our digital inputs with conventional material media is also a way forward to involve users and stakeholders. In Figure 4(c), we illustrate how we can gain real interaction between users and data by merging the virtual and the real by projecting our media for Greater London – animations of flows of traffic, GIS mapping and so on, onto the physical map that we call the London Data Table. Users can control the media from the computer but display it on the table. Last but not least, in Figure 4(d) we show how we can fashion a completely immersive environment where the virtual scene is projected directly into the glasses with the users interacting with one another through the common scene where users who are not physically connected to the media cannot gain access to it.

Real-time Data: Sensing and Mining Big Data in Space and Time

Much of the digital data that we continue to work with is collected using traditional sources and then transcribed to digital media with the best example being the Population Census. Some of this data is being captured digitally in the first instance through those whose data it belongs to, entering the data in direct digital form, while most of the secondary data that social scientists now use is already in digital form. However, as computers are being embedded into the built environment and as ever larger proportions of the population are interacting using hand-held devices such as smart phones and tablets, an increasing amount of data is being directly streamed into archives or into applications that involve real-time actions. In fact the data we will deal with here is mainly real-time data which is streamed into archives that as social scientists we access after the fact. There are multiple data streams that we now have access to and here we will focus on those associated with transport and with social media, transport which is associated with the use of smart cards (the Oyster card in London) and the location of vehicles used in public transit, and social media which is captured from short text messaging (Twitter data).

Before we examine these specific data, the easiest and most direct feeds of data from cities in real-time cluster around weather, transit, road traffic, pollution, Twitter trends, stock market prices, utilities and local news, some of which are national rather than local data. These feeds are fast becoming routine and we have organised some of these into what we call a City Dashboard. This is available for several UK cities, our exemplar being London which has more easily available data than any of the other cities (Birmingham, Brighton, Cardiff, Edinburgh, Glasgow, Leeds and Manchester) from which feeds are taken (http://citydashboard.org/). We show the dashboard for London in Figure 5 and we are also using different kinds of media to show the data besides a web site. Each element can be placed on a different screen and this has been installed in the Greater London Authority and a wall of iPads which is shown alongside the dashboard in the figure. This visualisation wall is built around the control room concept and displays the citywide data gathered via the City Dashboard which updates the wall on average every 2 seconds with feeds configurable from an array of options. The wall can also be controlled remotely via any computable device allowing a central user to not only change data but also enable a mode where a video can be shown across all 12 iPads simultaneously.

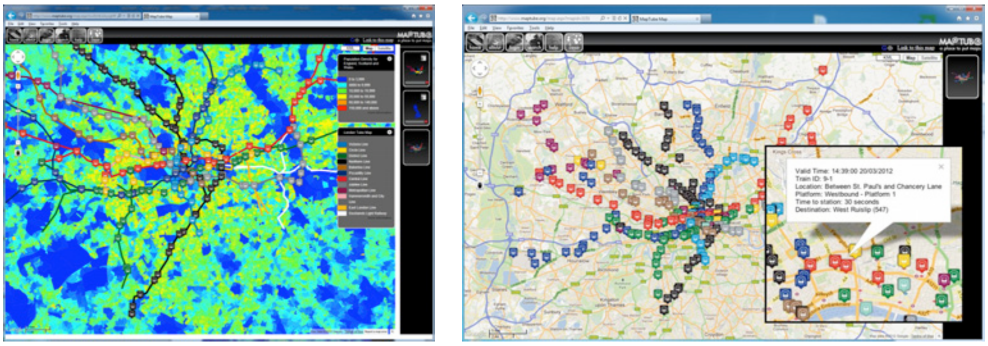
a) b)

Figure 5: The Dashboard (a)and its Display in a Visualisation Wall (b)

Retrieving and Visualising Real-time Transit Data

The simplest of such data is accessible using the various applications programming interfaces (APIs) that are available for identifying the stream of vehicles on transit systems which provide data on times, locations and delays. In fact, we have adapted MapTube to archive, analyse and display this data and this enables us to visualize the streaming data in near real time (with a latency of about 3 minutes from the time when the query is made on the server to whence the data arrives on our computers). Positions of all tube trains in London can be plotted on a map of the network, along with buses and overground trains for the whole country. During the rush hour, there can be up to 450 tubes running in London, 7,000 buses, 900 trains and 7,500 TfL Cycle Hire bikes, so we are capturing the daily cycle of people commuting to and from work. Population density, unsurprisingly, follows transport links (or vice versa) as we show in Figure 6(a) below, but by putting all this data together we are able to gain insights into how the physical fabric of the city is linked to social segregation. We also show in Figure 6(b) a clearer screen of the typical information available from the API for the tube with respect to one of the trains on the Central Line.

This data is useful for computing delays but it needs to be matched to demand for transit if it is to be useful for journey planning by individuals. In fact many transit systems are now using smart cards for payment and all public transit in London is now covered in this way with about 80 percent of all travellers of tubes, overground heavy rail and bus using the so-called stored value Oyster card. There are about 5-6 million tap in and tap outs on the tube system each working (week)day which is about 3 million travellers. This data records origin station (tap in entrance) and destination station (tap out exit) for a unique identifier (passenger) and the time at which the record is stamped. We can construct the complete flow matrix from station to station in any interval of time using shortest routes algorithms although the system is complicated and much work needs still to be done in piecing journeys together due to the fact that the underground has a complex geometry and many intricate corridors linking tube stations. Moreover the data is limited in parts due to the fact that there can be a mismatch between tap in and tap outs if barriers are open because those with fixed season cards or free passes (over 60s and young children) do not get fined if they fail to tap in or out. There can be a substantial mismatch between the data streams although the data does enable one to identify these kinds of problem quite easily.

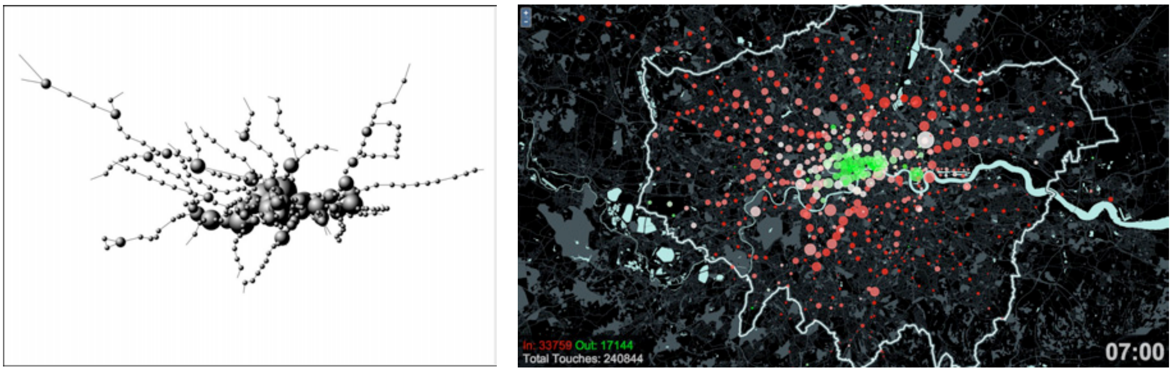


a) b)

Figure 6: Real-time Tube Train Locations

a) Population density from the 2001 Census overlaid on 11 March 2013 at 15:00 and
b) Tube delay data identified on 20 March 2012 at 14.39

Within the data set, we can easily explore routine behaviour over days and weeks and even months for the data set we have contains all journeys over a 6 months period. It is straightforward to identify deviations which might be due to disruptions, school holidays, sporting events and so on from such data and in principle, this provides important information for managing the system and ensuring that it is free-flowing. Moreover like all interaction or flow data, we are able to construct not only flows but locational activity at hubs (stations) and in Figure 7(a), we show the actual tube network with its links between stations using contemporary graph visualisation software and in b) the flow at the peak hour showing tap ins and tap outs at a typical morning peak. We have begun to explore questions of disruption using this data set by closing stations – a familiar enough occurrence at peak hours when some become overcrowded – and by closing segments of lines. We have done this for several key examples such as closing main line heavy rail stations where the tube is linked in and where volumes are substantial at the peak hour. What we do is to reroute travellers onto other lines and recompute the volumes that then enter or exit at all those stations unaffected by the disruption. In Figure 8, we show what happens when we close Liverpool Street which is one the busiest main line stations which has around 435,000 passengers moving through the tube stations that make up the complex each day with four lines intersecting. Figure 8 gives an indication of the increases in passenger volumes in the red bubbles compared to decreases in the blue bubbles. Without going into considerably more detail, we can say that the impact of closing this station does not spill out as much as might be anticipated. There are increases eastwards outside the circle line but not much around it for it seems that the redundancy of stations and lines in the centre of London is such as to mitigate such impacts. In one sense as soon as we begin to broach these kinds of problems, then we invoke a new style of analysis which is data-driven modelling. Visualisation of course is essential to portray this kind of complexity.



a) b)

Figure 7: The Geometry of the Tube Network and Real-time Volumes at Stations

a) The network visualised with stations shown according to the number of lines that intersect (indegrees/outdegrees) and b) Flow volumes at the morning peak where the maximum flow is shown at each station: green tap outs and red tap ins

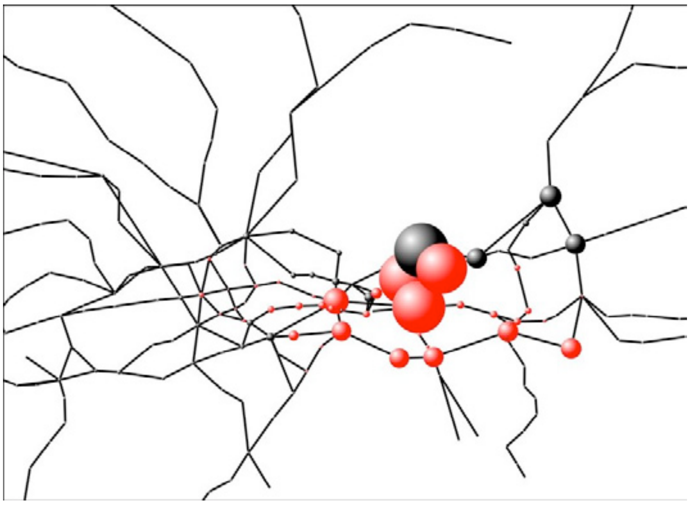
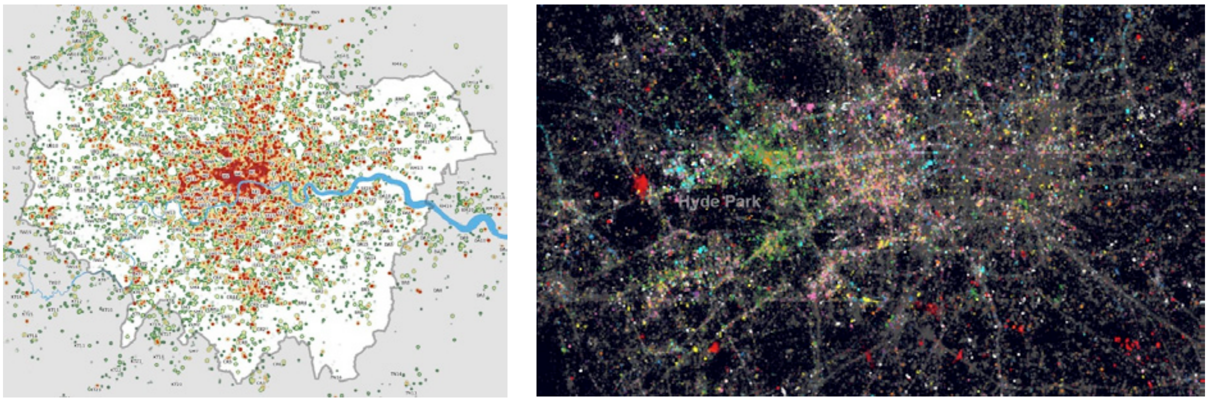


Figure 8: The Impact of Closing a Mainline Station (Liverpool Street) on the Flow of Travellers Passing through Related Stations

Visualising Social Media

The other sets of data that show spatial and social behaviours in real time are based on Twitter and similar data. Again we are able to record tweets in terms of the time they are sent, their location (if the GPS is switched on on the device), and their content using the Twitter API. At the time of writing however, there are changes to the streaming of data that are likely to limit what can be obtained freely in the future. We are able to produce locational data quite easily and we have recorded the density of tweets by location for a dozen or so large cities around the world (Neuhaus, 2013). Developed around the populist name 'Tweet-o-Meter', we have developed a system to mine data within a 30 km range of urban areas, focusing on

New York, London, Paris, Munich, Tokyo, Moscow, Sydney, Toronto, San Francisco, Barcelona and Oslo. Here in Figure 9(a) we show a map of the density of tweets in Greater London produced for 3 week period during the months of February and March in 2011 and this is clearly highly correlated with population density although it is possible by drilling into the map to associate tweets with particular activity locations such as schools, parks and entertainment centres. Analysis of the content and specifying location is much more problematic as content analysis of tweets is extremely difficult as there is so much ambiguity in text messages limited to 140 characters. One useful picture is given by extracting the language associated with those who use this medium. Using Google Translator software, Cheshire and Manley (2013) have produced a very detailed map of ten million tweets over 3 years for Greater London and this does excite interesting comparisons with ethnicity, multiple deprivation indices and related social area analyses. An earlier version of the map produced for the ten top languages is shown in Figure 9(b).



a) b)

Figure 9: The Spatial Density of Tweets in London

a) Recorded for 3 weeks in the months of February and March 2011 b) Tweets by language where the grey foundation of the map is formed from the majority of tweets in English. Other nationalities, in order of most to least prolific, are Spanish (white), French (red), Turkish (blue), Arabic (green), Portuguese (purple), German (orange), Italian (yellow), Malay (cyan) and Russian (violet).

There are other social media such as Facebook, Google itself, Foursquare and related sites which now enable their data to be mined and in our context here visualised in locational terms. Some of this data is particularly useful as it has much more substantive content than short text messages. A more focussed visualisation and analysis of Twitter data, however, relates to particularly significant events and the way in which people react to these spatially. The London riots in August 2011 were a case in point. Most of the regular media channels were suggesting that social media was being used to direct the riots, which may well have been true initially using encrypted instant messaging (as on the Blackberry device) which we do not have access to. Towards the end of the 4 days of riots, we collected a snapshot from Twitter to use for analysis which we illustrate in Figure 10 below.

Looking at the map of tweets which we show in MapTube, the majority do <u>not</u> contain tags such as "#londonriots", "#ukriots", "#riots", "#riotcleanup" or

“@riotcleanup”. However we can have access to the tweet content and in terms of our visualisation, by clicking on the individual points, we can explore the data and read the messages to find out what people are tweeting about. There are 34,314 points in this dataset and only 1,330 contained a riots hashtag (3.9%). The problem here is that while comparatively few people are using a recognised hashtag that can be easily identified computationally, they are still talking about the riots. The missing link at the moment in terms of this type of analysis is that natural language processing (NLP) has not advanced to the point where our analysis can be any more meaningful than this. Moreover Tweets also have their own language rather like text-speak, so an NLP system that can only read English is going to fail. We cannot be certain but many of the messages appear to imply that these tweets are simply from persons who are saying they have left work early and got home safely, no more than that.

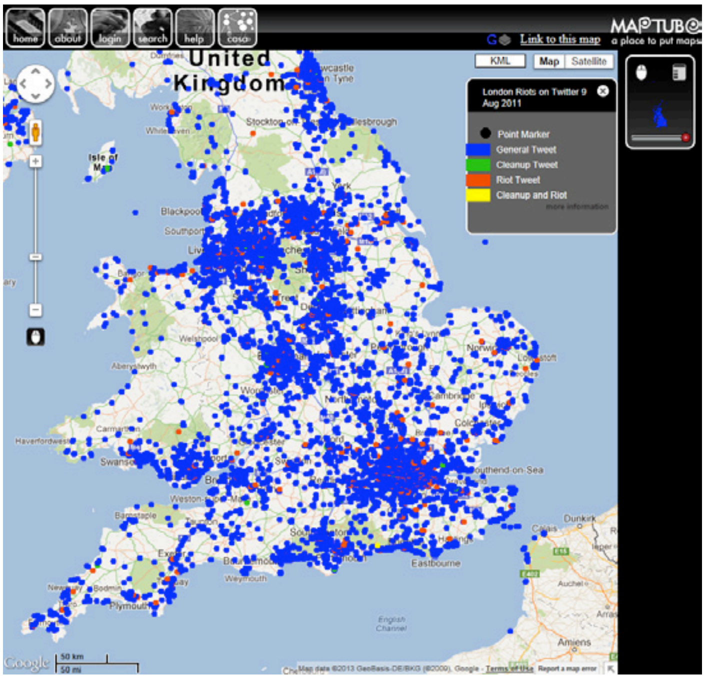


Figure 10: Geolocated Tweets Captured from Twitter Between 15:00 and 22:00 BST on Tuesday 9th August 2011.

Retrieving and Visualising Data from the Crowd

Crowdsourcing is the term used for methods of data creation, where large groups of users not organised centrally, generate content that is shared (Howe, 2008). From direct public involvement via citizen science initiatives, such as the well-documented Galaxy Zoo, through to simply tapping into online data feeds, this is marking out a new era of volunteered information and knowledge creation. The data is captured of course in real time but the time scale may not be grounded in any routine process largely because it depends on the ‘crowd’ responding. To gain substantial data, there needs to be some sort of broadcast medium to the crowd alerting them to the need to respond. The notion that there might be value in harvesting these kinds of response is based on the observation that, although a large number of individual estimates may be incorrect, imprecise, uncertain and ambiguous, their average can be a match for expert judgment (Surowiecki, 2004). Judiciously handled, randomly sampling the opinions or calculations of a large number of users might lead to data and information that are surprisingly accurate and that, in some cases, cannot be recorded in any other way. The potential of crowdsourcing methods applies across the sciences but it has had specific impact in spatial content through Volunteered Geographic Information (VGI) (Goodchild, 2007). OSM is perhaps the best known VGI output, started in UCL in 2004and it is based on over 400,000 volunteers creating a free editable vector map of the world. While many early volunteers were highly technically literate, they were not necessarily experts in geographical collection. Yet, through crowd-based quality control and refined workflows aimed at members beyond the traditional community, OSM has become the map of choice for many users.

A series of toolkits and workflows now exist to enable social scientists to collect and analyse data from the crowd as a means to include mass human input in spatial analysis. One such tool we have developed is a public website called SurveyMapper, which allows 'anyone' (with access to a web site) to ask the crowd 'anything' and returns a live map of results on a number of geospatial levels from the globe to the street (http://www.surveymapper.com/). SurveyMapper, whose backend is MapTube for visualisation, has been used by the BBC (Radio 4, Look East, BBC North and BBC South) and the Greater London Authority as well as the wider academic community to carry out rapid data collection. Tens of thousands of inputs can be collected quickly, providing a near real-time view of research questions. Data can be exported later for more rigorous analysis or integration with existing datasets. In Figure 11, we show the example of a map of broadband speeds of TV viewers in the Eastern region of the UK that have been collected from such responses where individuals were asked to respond to by simply clicking on a link that would measure the speed of their internet connection, and then key in a locational referent such as a UK post code once they had completed their response. In this way, one can build up a dynamic map over the period when the survey took place and assuming the response is representative (a big assumption at present), this will converge to a picture of the data that is useful and whose bias is known. This data of course is streamed in real time and in some circumstances might be organized to actually reflect rapid response although in the examples we have worked with, the crowd has been left to react over a matter of days or weeks.

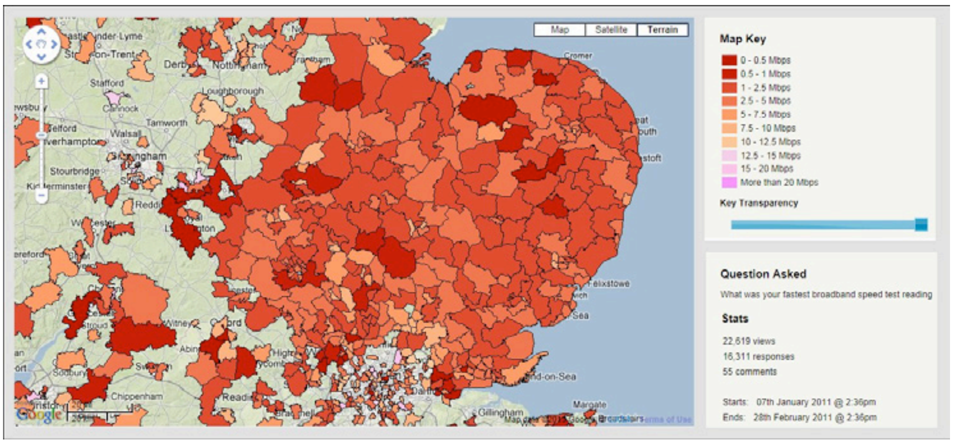


Figure 11: Spatial Crowdsourcing: Evolving Data in Real Time

All this kind of data raises enormous questions of privacy and security, as well as the degree to which it is representative. We are not able to explore these issues here but elsewhere in this book in the chapters dealing with ethics and confidentiality, they are touched upon. Nevertheless, they are important and are likely to become central to how we deal with the increasing deluge of digital social data that is upon us.

Predictive Analytics: Interpreting & Communicating Spatial Media

Our last foray moves away from data and representation to ways in which we can understand the mechanisms used to model or simulate that data through digital visualisation. There is at present a strong momentum to abandon traditional ways of theory building and testing through empirical verification for the focus on data is generating many new ideas about how the world works that can be elicited through crowdsourcing and the generation of big data (Anderson, 2008). In fact, quite the opposite is happening as many traditional models are being extended and disaggregated through simulating every agent in a population rather than sampling a population or modelling them in aggregate (Heppenstall, Crooks, See and Batty, 2012). This confronts big data head on in terms of inferring and deducing new theories from that data but the fact that the data volumes are large and the models often compute intensive means that visualisation is of the essence.

Visualisation is also needed for small data as well as big and models built on the basis of theory validation against traditional small data sets using extensive visualisation are as new as those for big data. Moreover the workings of the model can now be exposed in ways that are highly amenable to visualisation while the idea that observations and predictions can be exhaustively compared across many, many combinations of model calibrations elevates the problem of model visualisation to that akin to visualising real-time data. Moreover linking model processes to outcomes generates novel ways of visualising the relationships between processes and spatial outcomes. There are many such models that are now being visualised in these new ways and all we can provide here is a snapshot of our own experiences. Although we have worked with generative and agent-based models that tend to require visualisation because aggregate statistics do not give real meaning to their outcomes, we will show two examples here of essentially aggregate models that generate locational activities to small areas such as census tracts.

Visualising Empirically Calibrated Land Use Transport Models
The first model is an aggregative land use transport model that simulates the movement of people between work and home, work and shopping, home and shopping and other flows such as journeys to school and health care. These various sectors are simulated using spatial interaction (gravitational) models and are coupled together reflecting a sequence of competition between these activities but brought to equilibrium through iteration around the sequence of simulations. These types of model have a long heritage (Batty, 2009), they are aggregate and also static in that they simulate urban structure at a cross-section in time. They tend to be parsimonious with respect to their parameters – essentially the parameters control the effect of gravity and potential and the coupling, so current models tend to have numbers of parameters in the tens rather than hundreds. Only recently have such

models become statistically manageable in terms of the validation as often the process of estimating parameters is through trial and error. The model we show here is for Greater London and its wider metropolitan region divided into some 1767 zones (wards in UK Census terms). At every stage of the simulation, the model's outputs can be visualised in 2-d map terms but it is easy to extend this to 3-d using non-proprietary software which can be linked to the model outputs on the desktop and the web in the same manner we showed in Figure 3(c) where we examined floorspace as histograms arrayed across the space of Greater London.

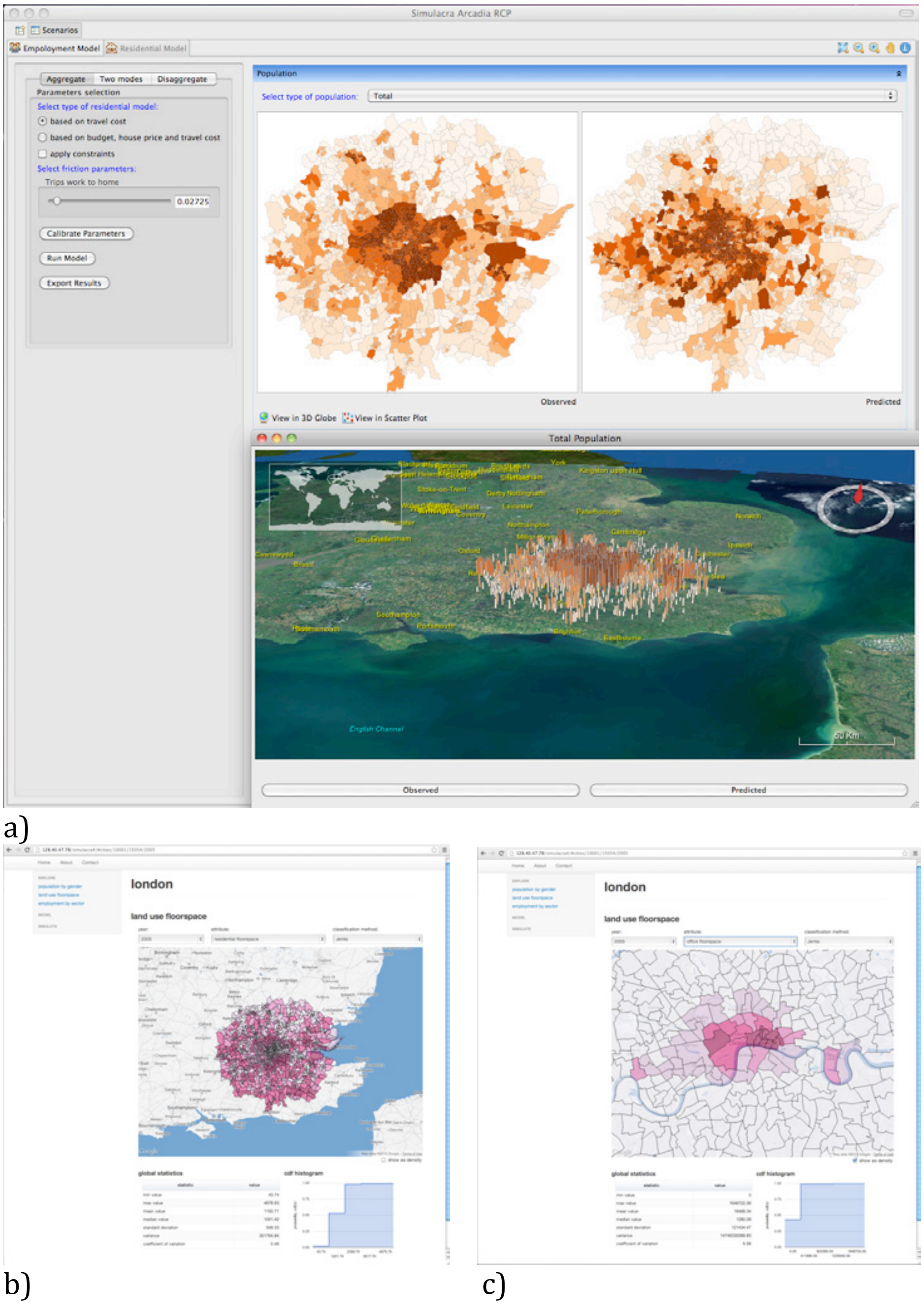


Figure 12: Visualising Model Outputs

a) In 2-d and 3-d with b) Data at the metro-region level and c) At the local level

In Figure 12(a), we show outputs from the model region and their visualisation in 3-d using the NASA World Wind software, and different levels of zoom in the model portal windows in 12(b) and (c) This enables one to visualise outcomes from the model as the simulation runs and data can be piped directly to the 3-d software while non-model data can also be compared against model outputs using these external software. These models tend to work at the level of thematic map layers and thus more geometric and raster data which is not required in the model simulation, can be compared with model outputs in these external software. In fact we are fast generalising these kinds of model outputs into web-based portals where the model type can be applied to many different examples. We show the model data used in London at two levels of zoom with some simple spatial analytics/statistics relating to the data in the portal window shown in Figure 12(b) and (c). The ideas behind these kinds of visualisation are explained in detail elsewhere (Batty, 2013).

Visualising Theoretically Inspired Location Models

Our second model implements the basic theory of location first proposed by von Thunen in 1826 (Hall, 1966) within a 3-d geometric CAD model framework. We have already seen how 3-d geometric models of cities which are essentially representations of urban activities tagged to the blocks associated with land parcels at the street level can be tagged to urban activities data. In fact such models can be tagged to data that is from model outcomes or predictions where such models simulate the structure of activities in cities at the parcel/block level. The software ArcGIS has been extended to include methods of procedural modelling based on City Engine originally developed at ETH in Zurich. Such modelling is rule-based and enables users to create and render very large scenes by formulating rules that can be applied systematically to extensive areas of the block geometry. It can be used for a variety of purposes from rendering the texture of large scenes to populating the geometric models with blocks that conform to certain rules of urban structure such as those developed using basic models of location such as those attributed to von Thunen through to models that reflect spatial interaction based on gravitation.

Von Thunen's model essentially simulates the competition for land uses around a market centre which is used to distribute products generated at different distances from the central market which imply different transport costs. The model determines how different land uses occupy different concentric rings around the centre according to their ability to compete for land in terms of the payment of rent which in essence depends on the importance of transport costs in the production process. Activities where transport cost is more important are likely to outbid activities where it is less important in such a way the land use in question locates nearer the centre and is able to pay more rent.

This bid rent model has been generalised by Wilson and Birkin (1987) and it can be extended to multiple market centres. What we are able to show using the procedural rules of City Engine is how a more complex dynamic than von Thunen's original model can be used to simulate complex radial structures of land use as we illustrate in Figure 13(a). The sliders in the interface in Figures 13(a) to (c) indicate how we can change the structure of land uses in the city and visualise the outcomes, in much the same manner Steadman introduced for the same model using much simpler 2-d software for small examples (Batty, Steadman and Xie, 2004). We also show in Figure 13(d) a generalisation of the model to retail gravitation where we introduce multiple centres and examine the intensity of spending around each centre. The application required a combination of Python and CGA rules (a scripting language used by the CityEngine interface) and the manual development of an information system where these two platforms automatically exchange information. This was a key feature, as it allowed the applications to perform the necessary calculations while maintaining an interactive nature. The result is the real-time visualization of dynamic models which can be directly affected by the user. These examples demonstrate real-time visualization of urban model structures and produce a 3-d real-time interactive animation of how an urban structure might evolve. The user in this case can participate in the simulation process by controlling or altering different attributes while the simulation is running and experiment with different scenarios. The outcome is both visual and analytical, as there is the option of providing 3-d diagrams and matrices of different statistics.

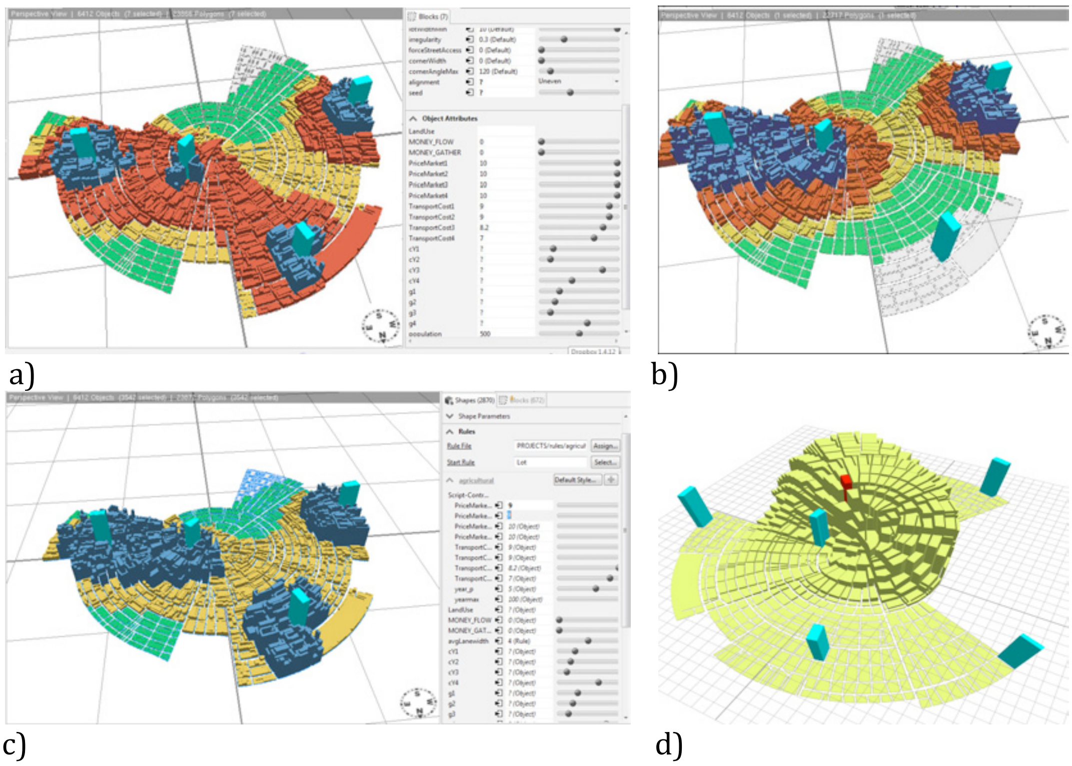


Figure 13: Using Procedural Modelling in City Engine to Visualise Radially Structured Land Use Activity Patterns Reflecting Density, Transport Cost, and Rent

a) the user selects a location for a new city centre b) the model calculates land uses in real time c) the user can use the generated sliders on the right hand side to change the parameters of the model d) multiple centres and their impact on residential activity in a model of retail gravitation.

The Future of Digital Visualisation

There are many developments in the visualisation of spatial and social data that we have not been able to catalogue here. In particular, the whole field of infographics which is essentially a development of exploratory data analysis is giving rise to new kinds of representation more abstract than the illustrations we have included here. New methods of representing networks, bipartite graphs and flow structures which present various kinds of correlations and interactions are being rapidly developed. Many of these new varieties of graphic are augmented by animations which we have not been able to show here although these are implicit in the construction of the pictures that illustrate these ideas.

The data we have mainly shown tends to be relatively large-scale in terms of capacity largely due to the fact that spatial and temporal extent can massively increase the size of the set but we have also noted that relatively small data sets which still dominate the social world especially through sampling are being radically improved through new methods of visualisation. Software for such visualisation is increasingly free, a good example being the IBM site Many Eyes and the sites that offer network software such as Pajek (http://pajek.imfm.si/doku.php?id=pajek/) and NodeXL (http://nodexl.codeplex.com/).

In fact, much social data which is often intrinsically non-spatial or where the spatial dimension is unimportant to the analysis is likely to make use of this more abstract software. Our focus here on 2-d and 3-d and on temporal streaming represent very new ways of thinking of social data in the spatial domain but we believe that the use of these new dimensions is more generic than space or time itself. All spatial data implies sequences which of course represent the organising principles of space and time – adjacency in space and continuity and irreversibility in time. There are some who proclaim that big data and the web are heralding the end of theory but in fact nothing could be further from the facts. New ideas about how social systems interact which break with the simple principles of sequence that have marked all the examples here are on the horizon. These are evolving through the very use of the same information technologies that we are using here to understand the system through the lens of globalisation which is destroying spatial and temporal simplicity as the world becomes more complex. These are trends that are driving social and spatial science and in the next decades, we are likely to see radically new kinds of visualisation that are driven as much by new ways of articulating the way social systems function as by the appearance of new forms of data and analytics.

References

Anderson, C. (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, *Wired*, June 23rd, 2008; accessed 28 March 2013, and available at http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Anderson. P. W. (1972) More Is Different. *Science*, *177*, (4047), 393-396.

Batty, M. (2013) Visually-Driven Urban Simulation: Exploring Fast and Slow Change in Residential Location, *Environment and Planning A*, in press.

Batty, M. (2009) Urban Modelling, in R. Kitchin and N. Thrift (Eds) *International Encyclopedia of Human Geography*, Volume 12, Elsevier, Oxford, 51–58.

Batty, M., Hudson-Smith, A., Milton, R. and Crooks, A. (2010) Map Mashups, Web 2.0 and the GIS Revolution, *Annals of GIS*, *16*(1), 1-13.

Batty, M., Steadman, P. and Xie, Y. (2006) Visualization in Spatial Modeling, in J. Portugali (Editor) *Complex Artificial Environments*, Springer, Berlin, 49-70.

Batty, M., and Hudson-Smith, A. (2005) Urban Simulacra: From Real to Virtual Cities, Back and Beyond, *Architectural Design*, *75* (6), 42-47.

Cheshire, J. and Manley, E. (2013) The Twitter Map of London, *Guardian Newpaper*, Tuesday 26-03-13, 28-29; The earlier map is available at http://spatialanalysis.co.uk/2012/10/londons-twitter-languages/

Chrisman, N. (2006) *Charting the Unknown: How Computer Mapping at Harvard Became GIS*, ESRI Press, Redlands, CA.

Dyson, G. (2012) *Turing's Cathedral: The Origins of the Digital Universe*, Pantheon Books, New York.

Goodchild, M. F. (2007) Citizens as Sensors: The World of Volunteered Geography, *GeoJournal*, *69*, 211-221.

Hall, P. (Editor)(1966) *Von Thünen's Isolated State*: *An English Edition of Der Isolierte Staat by Johann Heinrich Von Thünen*, Pergamon Press, Oxford.

Heppenstall, A. J., Crooks, A. T., See, L. M. and Batty, M. (Editors)(2012) *Agent-Based Models of Geographical Systems*, Springer, Berlin and New York.

Howe, J. (2008) *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*, Crown Business, New York.

Lima, M. (2011) *Visual Complexity: Mapping Patterns of Information*, Princeton Architectural Press, Princeton, NJ.

Mandelbrot, B. B. (2012) *The Fractalist: Memoir of a Scientific Maverick*, Pantheon Books, New York.

Mayer-Schonberger, V. and Cukier, K. (2013) *Big Data: A Revolution that Will Transform How We Live, Work and Think*, Houghton Mifflin, New York.

Neuhaus, F. (2013) *Urban Rhythms: Habitus and Emergent Spatio-Temporal Dimensions of the City*, unpublished PhD thesis, CASA, University College London, London.

Plummer, A. V. (2003) *The Chicago Area Transportation Study: Creating The First Plan (1955-1962): A Narrative*,
available at http://www.surveyarchive.org/Chicago/cats_1954-62.pdf

Smarr, L. L. (1991) Visualization Captures the Imagination of Physicists, *Computers in Physics*, 5, Nov/Dec, 564-567.

Surowiecki, J. (2004) *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, Little, Brown and Company, New York.

Tufte, E. R. (1992) *The Visual Display of Quantitative Information*, Graphics Press, Cheshire, CT.

Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley, Reading, MA.

Wilson, A. G., and Birkin, M. (1987) Dynamic Models of Agricultural Location in a Spatial Interaction Framework, *Geographical Analysis*, *19*, 31–56.