Geography and Computers: Past, present, and future

Abstract

The discipline of Geography has long been intertwined with the use of computers. This close interaction is likely to increase with the embeddedness of computers and concomitant growth of spatially-referenced data. To better understand the current situation, and to be able to better speculate about the future, this article provides two parallel perspectives: first, we offer an historical perspective on the relationship between Geography and computers; second, we document developments—in particular the nascent field of data science—that are currently taking place outside of Geography and to which we argue the discipline should be paying close attention. Combining both perspectives, we identify the benefits of tighter integration between Geography and Data Science, and argue for the establishment of a new space—that we term Geographic Data Science—in which cross-pollination could occur to the benefit of both Geography and the larger data community.

Introduction

> *It is not the use of computers that distinguishes the forthcoming revolution but the development of a new computationally intensive and totally computer-dependent paradigm in geography.*
>
> *Stan Openshaw (1994:500)*

The rise of 'big data' in academia and industry has triggered renewed debate about the role that quantitative and computational methods play in Geography in general, and in Human Geography in particular. The explosion of available data has brought back to the surface long-running debates about how Geography is taught, researched, and experienced as a set of sometimes fractious familial relationships (see Cresswell 2013, 2014; Johnston *et al*. 2014b; O'Sullivan 2014; Wyly 2014). To better understand the effect that the 'data revolution' (Kitchin, 2014) might have on Geography, we believe it is useful both to examine how our discipline has responded in the past to such pressures, and to consider how developments outside Geography might offer insight into the range of potential future directions.

We thus see this paper as complementary to perceptive histories of quantitative geography written elsewhere (*e.g.* Barnes 2004, 2013, 2014), but wish to draw attention to the seemingly overlooked connection between the subsequent evolution of geographical methods in connection to technological changes in computer hardware and software. We therefore have three distinct goals: first, to quickly review the history of computation *in* Geography so as to provide a context for contemporary debates; second, to document recent developments *outside* Geography that are reshaping our understanding of the world through data; and third, to *reflect* on how a putative Geographic Data Science might provide a foundation for further development.

A (very) brief history

> *In the past decade geography has undergone a radical transformation of spirit and purpose, best described as the 'quantitative revolution'... Although the future changes will far outrun the initial expectations of the revolutionaries, the revolution itself is now over. It has come largely as the result of the impact of work by non-geographers upon geography…*
>
> *Ian Burton (1963:151)*

Although a detailed history of the long, and sometimes combative relationship between (Human) Geography and computers is both beyond the scope of this article and has been done elsewhere before (*e.g.* Armstrong, 2000; Cresswell 2013; Torrens 2010; Haining 2013), it is useful to provide some historical perspective so as to better understand the current state of affairs and scope for future developments. Significantly, although Geography and the affiliated domain of Planning were amongst the earliest adopters of computers in the 1950s and 60s, they were also (in Britain at least) amongst the disciplines that turned most strongly *against* their use as a tool for thinking about and analysing space a few decades later. In Geography, the critical and cultural 'turns' of the 1970s and 80s were characterised by a range of cutting critiques (*e.g.* Ley and Samuels 1978 and see also Barnes 2004), including perhaps most notably the 'Damascene conversion' of David Harvey (1972).

There are reasons to believe the pendulum has recently begun to swing back, and a new appreciation of quantitative approaches in geography is taking shape. An important explanatory factor behind this shift resides in the coincidence of a set of methodological and technological advances over the past decade—though these build on accelerating changes over the *preceding* two or more decades—that is reshaping how we employ and understand computation and computers in almost every aspect of human life. The declining size and cost of processors, storage and geospatial technology has given rise to new sources of data about the world and to the possibility of using them to provide answers to long-standing geographical questions for which relevant data simply was not available. Consequently, to understand the (re)emergence of computing in Geography is to understand the effects that the embedding of computers in every facet of daily life is having on social science research as a whole (Lazer *et al.* 2009; Watts 2007).

The First Wave: a computer in every institution

As early as 1963, Burton was arguing that the *first* quantitative revolution was a theoretical one and *not* a methodological one (Burton 1963). The vanguard of this revolution saw statistics as a tool with which to uncover spatial structure, arguing that without 'observation and description of regularity' there would be nothing against which to measure—and judge—the unique and the exceptional. In his historical work on the discipline Barnes (Barnes 2013; Barnes 2014) echoes this view, suggesting that the work begun by, for instance, Brian Berry at the University of Washington encouraged a major shift towards the use of statistics as a tool for theory-validation. And, despite the subsequent critical and cultural 'turns', quantitative methods *did* spread from the select few journals and departments of the early years documented in Barnes (2004) and carved out a place of their own in the discipline.

The flagship journal *Progress in Human Geography* (*PiHG*) offers a good illustration of this process: launched in 1977, its early volumes included reports on advances in time series analysis (Cliff 1977), spatial diffusion (Cliff 1979), and modelling (Cliff 1980). But it is neither easy, nor particularly useful, to separate this theoretical shift from the technological changes that made it possible: although most of the computation done at that time could still, in principle, be carried out by human 'calculators', the punch card and magnetic tape made it possible to do matrix manipulation and other demanding tasks at seemingly breakneck speed (*e.g.* Goddard 1970; Demšar et al. 2018) and consequently had a significant effect on the adoption of quantitative methods in Geography.

Of course, computers at this time were very large. They were expensive and hard to operate too. These constraints meant that geographers were forced to share the few machines available on campus, and this made clock cycles and computation time a precious luxury not to be wasted. A good example of the consequences these limitations imposed can be found in the numerous shortcuts, simplifications, and assumptions that fill appendices in statistical papers from those years with the goal of obtaining "computational feasibility" (*e.g.* Cliff and Ord 1981). The computer in those days was an exciting new tool for statistical analysis at scale (see the personal recollections in Billinge *et al.* 1984), but it would not be unfair to characterise it as primarily substituting for the time and energy of users in the midst of a more theoretical project.

The Second Wave: a computer in every office

Without wishing to suggest that the next wave of innovation in computing *determined* the accompanying transformation of—and, ultimately, divisions within—quantitative geography, the growing availability of desktop computers in the 1980s inevitably had a profound effect on how we 'do geography with computers' (Harris *et al*. 2017). The desktop system is, of course, intimately bound up in the rise of Geographic Information Systems (Goodchild and Haining 2004) and, consequently, of Geographic Information Science (Goodchild 1991). But the dedicated personal computer also enabled the design and use of much more computationally demanding methods, notably the development of 'local statistics' in the 90s (Haining, 2014). Poon (2003) and, later, Johnston *et al*. (2014a), argued that such statistics should be seen as an empirical response to the critique of the cultural geographers because they explicitly incorporate variation over space.

With a computer on every geographer's desk, the discipline quickly began to imagine new ways to use them. The cumulative impact that the explosion of computing power was having on the discipline was summarised in the three-part series for *PiHG* that Stewart Fotheringham wrote exploring the local (1997), the computational (1998), and the visual (1999). Well before that, however, the *Progress* reports had already highlighted developments in discrete choice modelling (Wrigley 1982), longitudinal data analysis (Wrigley 1986), and input-output analysis (Thomas 1990). This is also the period where Agent-Based Models and Cellular Automata (O'Sullivan 2008, Torrens 2010) emerge as a distinct path in geographical model development, principally for exploring complexity (*e.g.* Batty 2005, 2013).

Of course, in many respects the 1990s are usually seen as the decade of GIS, with a new 'reports' series in *PiHG* focussed solely on this approach beginning in 1995. Chronicling the fast evolution of this nascent field, they explored issues in the representation, storage and analysis of spatial data. The first two reports by David Unwin covered uncertainty (1995) and the relation between GIS and spatial

statistics (Unwin 1996). The topics that featured most prominently during the latter part of this period were connected to challenges in data infrastructures (Longley 2003), representation (Longley 2004), time (O'Sullivan 2005), and geovisualization (Elwood 2009, 2010). The transition, from margins to mainstream, of GIS was cemented by the release of ArcMap in 1999 and the subsequent inclusion of 'GIS classes' in many undergraduate and graduate programmes.

In terms of academic research, the dominant discourse was one of the relationship between GI Systems and GI Science (e.g. Openshaw 1994; Openshaw and Abrahart 2000; Fotheringham 1998; Haining 2014). Beneath this train of thought, however, another line of investigation continued to bubble way, and in Couclelis (1998:19) the term 'geocomputation' is explicitly used to distinguish this other approach from GIS, which she defines as "a technological advance that would allow applied geographers and others to do faster, more comfortably, and better what they had always done." In this view, GIS should be seen as a ramping up of the process begun in the first wave: computers enable geographers to do quickly what was once done painfully by hand but is not, in and of itself, a form of (geo)computational *thinking* (Gahegan 1999). This distinction, noted by the geography community of the time and manifested in, for instance, the neglect of topics such as Artificial Intelligence (AI) is hardly coincidental (see, *e.g.*, Goodchild, 2010). We therefore feel that geocomputation should be seen as part of a separate tradition much more concerned with the interface between geography and computer science, with what computers make *possible* and not what they make *easier* (Gahegan 1999).

The Third Wave: a computer in every *thing*

By now it should be clear that the embedding of computation in everyday objects, not just dedicated computers, heralds another major shift for computational geographers. Part of the significance of this third wave lies in the vast amount of affordable computational power available to store, process, and analyse an ever growing amount of data. Much of geography's attention, however, has been focused on the byproduct of this embedding process (e.g. Reades 2007, Ratti 2010)—what Hartford (2014) termed the 'data exhaust'—and less attention given to more fundamental change that is afoot. Put simply, computers are no longer just desktop machines with which we 'ingest' and process observations; they have become 'autonomous data generators' in their own right, and machine-to-machine interactions spawn data in volumes that dwarf those of own (human) intentional activities. As an illustration, while the compressed file for all the US Census geographies takes 6.3GB (US Census Bureau, 2018), it is estimated that autonomous vehicles will each generate over 4TB of data every 90 minutes (Winter 2017), most of which will be inherently spatial data.

Whatever we may feel about the ultimate consequences of this process (*e.g.* Thatcher *et al*. 2016), the deluge of data is inseparable from a confluence of two trends: the declining size and cost of hardware, and the declining cost of software. It is now possible to make cellular network-enabled devices so small and so cheap that they are, literally, disposable in the name of research: Phithakkitnukoon et al. (2013) attached cell-enabled modules to track the movement of rubbish across America and even internationally! Sensors are now *everywhere*: in our phones and homes, in our bridges and tunnels, orbiting the Earth in the form of nano-satellites, and (implicitly) in the digital traces that we leave in the networks with which we interact. Thanks to the rise of affordable, low-power hardware platforms such as Arduino (www.arduino.cc), as well as cheap 'self-replicating' 3D printing systems (*e.g.*

reprap.org) that enable customised parts to be quickly manufactured on-site, a wealth of innovative applications in geographical data collection, particularly in the developing world, are now emerging.

The physical devices that sustain this revolution are not only cheaper, they are also more accessible: although the desktop era was largely dominated by proprietary software running on proprietary platforms, since the rise of the Linux operating system there has been a proliferation of 'cheap'—as in free—software. Free Open Source Software (FOSS) has been around since the early mainframe days but the use of open source has increased to the point where there is now an entire ecosystem of freely downloadable and (re)usable software. As the first quantitative *Progress* report in ten years notes (Brunsdon, 2016), the shift towards FOSS platforms such as Python, R, and QGIS to support open and reproducible workflows is becoming mainstream. Cumulatively, the ability of such systems capture *aspects* of the world in unprecedented detail—data in the form of text, imagery, and operational records—is seen as key to unlocking a wealth of insight into the social and physical environment.

It is important to note, in many cases, researchers can only access these in an "accidental" manner (Arribas-Bel, 2014), implying that several of the channels, formats and quality checks scientists use with traditional data do not necessarily apply in this context. To some (usually non-geography) proponents, the growth of 'big data' represents the 'end of theory' (Anderson, 2008), while to its detractors it represents a new kind of 'automated post-positivism' interested primarily in "selling you things that you don't actually need" (Wyly, 2014). The ultimate consequence of this reconfiguration of the data landscape is that the social sciences—and geography in particular—have gone from being data poor to being overwhelmed by a firehose of data sprayed towards us at high velocity, in high volumes, in a wide range of fast-changing formats, but much of it is of dubious provenance (Kitchin, 2013).

An emergent Data Science

> *I think statisticians are part of it, but it's just a part. You also want to be able to visualize the data, communicate the data, and utilize it effectively. But I do think those skills—of being able to access, understand, and communicate the insights you get from data analysis—are going to be extremely important.*
>
> *Hal Varian (2009:np)*

The accelerating co-evolution of hard- and soft-ware has led some prominent scholars to write of a 'data revolution' (Kitchin, 2014), while outside of Geography some have even argued for a re-thinking of the methods and practices that researchers and analysts use to make sense of data, proposing a 'computational social science' (Lazer *et al.*, 2009) or 'data science' (e.g. Donoho, 2017). Burton (1963:152) has suggested that geography is often a 'following discipline' whose "main currents of thought have had their origins in other fields." So as we turn towards the future of Geography we think it's worth looking to the emergent field of 'data science' and its use of algorithmic approaches to extract 'signal from noise' (Silver, 2012).

Data science is, at best, loosely defined (see Loukides, 2011 or Schutt and O'Neil, 2013 for illustrative attempts), and competing disciplines, from statistics (e.g. Wu, 1997) to computer science

(e.g. Naur, 1974), have sought to take ownership of a terrain already occupied in many cases by the corporate behemoths of the early 21st Century. Donoho (2017), however, traces the origins of contemporary data science back more than fifty years to John Tukey's *The Future of Data Analysis* (1962). Donoho's paper seems to be one of the few formal attempts to synthesise what Data Science *is* without falling into marketing propositions or hype. He points to the incorporation of six key ideas not traditionally taught as part of a 'statistics degree': data gathering, preparation, and exploration; data representation and transformation; computing with data; data visualization and presentation; data modeling; and a reflexive 'science of data science'.

Looking back, it's clear that these are issues with which our discipline has wrestled in the pages of *PiHG*, but data science provides a framework to not only better understand, but also to effectively leverage, the kind of broadly defined 'data' that is of interest to geographers. Directly or indirectly, many of data science's applications are inherently spatial and geographic in nature, although the degree of engagement by 'mainstream' data scientists with geographical methods and thinking has been fairly minimal to date. Within the discipline, however, there is a widespread appreciation—built on the advances and struggles outlined above—that the majority of the behavioural data generated by our 'networked society' is spatially embedded and that geographical traditions may have much to offer to 'big data' research. Everyone—from Google and Airbnb to mobile phone carriers—is in the geo-data business these days, and O'Sullivan and Manson (2015) have, tongue planted firmly in cheek, suggested that this is one reason why physicists (amongst others!) are now the ones with geography envy.

Conclusion: Towards a Geographic Data Science

> *The key question [...] is whether [this] is to be understood as a new perspective or paradigm in geography and related disciplines, or as a grab-bag of useful computer-based tools… The question whether or not we are witnessing the rise of a distinct intellectual approach to the study of geographical space through computation…*
>
> *Adapted from Helen Couclelis (1998:18)*

This paper has reviewed the relationship between computation and Geography since the invention of the modern computer in the 1950s. To recap, we argue that Quantitative Geography was born out of the possibility of using machines to replace human calculators for theorising with data, and that GIS extended this potential into the revolution wrought by (spatial) databases (Gahegan, 1999). We note that this is distinct from the geocomputational strand that flourished when researchers gained *personal* access to computers powerful and flexible enough to allow them to begin to think computationally. We then considered the emergence of a Data Science which, though it exists largely at the interface between computer science and statistics, incorporates ideas and practices that don't clearly fit into either discipline. And we have argued that this development must be seen as, in part, a response to the abundance of data on (human) activity in new forms from fundamentally new types of data sources.

These changes may have taken place largely at the periphery of Geography (as they have in some ways for the Humanities), but they point to a new phase in the ongoing evolution of our relationship with computers. We therefore argue there are substantial benefits from an explicit engagement by

computational Geographers with Data Science: as others (e.g. Kitchin, 2014) have already noted, and as our 'third wave' section seeks to make clear, the recent explosion of data is not only a quantitative change in the *amount* of digital, machine-readable information at the fingertips of researchers and industry, it is also a profoundly qualitative change in how we think about, and work with, data. In this respect, Data Science is an already existing effort to make sense of, and synthesise the most appropriate ways to cope with, this 'brave new world'.

And yet, these challenges are neither entirely foreign to, or entirely novel in, our discipline: we are not advocating the creation of a new domain *ab nihilo*. Even if we see this as a new phase in the relationship between geography and computers, it should build on the contributions of cognate areas of endeavour such as GIS, geovisualisation, spatial statistics and, of course, geocomputation. As seems common with the rush to develop a new discipline, however, data science has not been engaging with space but 'reinventing it'. But it would be desirable to avoid this duplication of effort, as over the last decades computational geographers have learned a number of hard lessons about dealing with geographic data to which data science would do well to listen.

We therefore see a need for an interface, and common ground for discussion, between Geography and Data Science. For pragmatic reasons we would call this a Geographic Data Science (GDS), but the ultimate objective this engagement remains the formation of new *knowledge* about the world as socio-spatial process (Taylor, 1990). Conceptually, a GDS would combine the tradition of 'spatial thinking', prevalent in computational Geography and GIS, with the modern approaches to data capture, transformation, processing, and analysis championed by Data Science. GDS would thus take both data and space as first-class citizens, and build a set of agendas, practices, and methods around them. There is also some urgency to this proposed project: a principled refusal to engage with data science on epistemological, methodological, or even political, grounds would leave parts of *our* disciplinary terrain and its (permeable) frontiers with other quantitative disciplines occupied by those with no appreciation of the history, techniques, and rationales underpinning spatially-aware quantitative analyses (see critical discussion in Brunsdon, 2014).

Geography has always been a bridging discipline between the natural and social sciences and the humanities, but in the last few years there has been a 'hollowing out' of the skills required to deal with spatial data (Singleton, 2014). The *value* of geography, however, has never been so pressing: our discipline has a long tradition of critical engagement with data *and* with their analysis and visualisation, and this is integral to the understanding(s) that result from sophisticated 'machine learning' and 'big data' research. Crucially, we see GDS employing modern computational techniques while incorporating the spatial, ethical[1], and conceptual training of the geographer as an integral element of their production and interpretation. The understandings at which we might arrive are ones at which neither group might arrive at on their own: a spatially-aware data science should be sensitive both to the substantive and insightful critiques of quantitative analyses mounted by critical and cultural geographers, and to the ways in which 'data-generating processes' (Lu and Henning, 2012) are spatially determined.

Ultimately, the challenges tackled by GDS may be new and driven by access to novel forms of geo-data, or they may be traditional questions tackled in new ways; but as a space for the co-production of

[1] For a recent illustration of the risks of ignoring ethical considerations when deploying data science in society, the reader is referred to O'Neil (2016).

new knowledge the proposed GDS 'space' benefits all parties: geographers are able to work with those who are pushing the boundaries of what it is possible to do with data and computers, while data scientists would benefit from a body of theory, practice, and expertise developed over decades that critically reflects on how space and location affects process and outcome, and that consequently seeks to explicitly account for such effects. To put it another way: geographical data scientists understand *both* that *x* and *y* are just two dimensions amongst many in a big data set *and* that these axes are 'special' because of the ways in which they both reflect and shape human behaviour and experience.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M. and Ghemawat, S. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv Preprint, *arXiv:1603.04467*.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7), 23 June 2008.

Armstrong, M.P. (2000). Geography and computational science. *Annals of the Association of American Geographers*, 90(1), 146–156.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49, 45–53.

Barnes, T.J. (2004). Placing ideas: genius loci, heterotopia and geography's quantitative revolution. *Progress in Human Geography*, 28(5), 565–595.

Barnes, T.J. (2013). Big data, little history. *Dialogues in Human Geography*, 3(3), 297–302.

Barnes, T.J. (2014). What's old is new, and new is old: History and geography's quantitative revolutions. *Dialogues in Human Geography*, 4(1), 50–53.

Batty, M. J. (2005). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. Cambridge USA: MIT Press.

Batty, M. J. (2013). *The new science of cities*. Cambridge USA: MIT Press.

Billinge, M., Gregory, D. and Martin, R. (eds. 1984). *Recollections of a Revolution: Geography as Spatial Science*. London: Macmillan Press.

Brunsdon, C. (2014). Spatial science–Looking outward. *Dialogues in Human Geography*, 4(1), 45–49.

Brunsdon, C. (2016). Quantitative methods I: Reproducible research and quantitative geography. *Progress in Human Geography*, 40(5), 687–696.

Burton, I. (1963). The quantitative revolution and theoretical geography. *The Canadian Geographer/Le Géographe canadien*, 7(4), 151–162.

Cliff, A. (1977). Quantitative methods: Time series methods for modelling and forecasting. *Progress in Human Geography*, 1(3), 492–502.

Cliff, A.D. (1979). Quantitative methods: spatial diffusion. *Progress in Geography*, 3(1), 143–152.

Cliff, A.D. (1980). Quantitative methods: Spatial modelling. *Progress in Geography*, 4(4), 568–576.

Cliff, A.D. and Ord, J.K. (1981). *Spatial processes: models & applications*. Taylor & Francis.

Couclelis, H. (1998), Geocomputation in context. In *Geocomputation: A primer*. Chichester: John Wiley, 17–29.

Cresswell, T. (2013). Chapter 5: Spatial Science and the Quantitative Revolution. *Geographic Thought: A Critical Introduction*. John Wiley & Sons, 79–102.

Cresswell, T. (2014). Déjà vu all over again: Spatial science, quantitative revolutions and the culture of numbers. *Dialogues in Human Geography*, 4(1):54–58.

Demšar D., Reades J., Manley E. and Batty M. (2018) Revisiting the Past: Replicating Fifty-Year-Old Flow Analysis Using Contemporary Taxi Flow Data. *Annals of the American Association of Geographers*, 108(3):811-828.

Donoho, D. (2015). 50 years of Data Science. *Tukey Centennial Workshop*. Princeton, NJ, presented 18 September 2015. Retrieved from: http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf *[Last checked 18 May 2018]*.

Elwood, S. (2009). Geographic Information Science: new geovisualization technologies—emerging questions and linkages with GIScience research. *Progress in Human Geography*, 33(2), 256–263.

Elwood, S. (2011). Geographic information science: Visualization, visual methods, and the geoweb. *Progress in Human Geography*, 35(3), 401–408.

Fotheringham, A.S. (1997). Trends in quantitative methods I: stressing the local. *Progress in Human Geography*, 21(1), 88–96.

Fotheringham, A.S. (1998). Trends in quantitative methods II: stressing the computational. *Progress in Human Geography*, 22(2), 283–292.

Fotheringham, A.S. (1999). Trends in quantitative methods III: stressing the visual. *Progress in Human Geography*, 23(4), 597–606.

Gahegan, M. (1999). Guest Editorial: What is Geocomputation?. *Transactions in GIS*, 3(3), 203–206. doi:10.1111/1467-9671.00017

Goddard, J.B. (1970). Functional regions within the city centre: A study by factor analysis of taxi flows in central London. *Transactions of the Institute of British Geographers*, 49, 161–182.

Goodchild, M.F. (1991). Geographic information systems. *Progress in Human geography*, 15(2), 194–200.

Goodchild, M.F. (2010). Twenty years of progress: GIScience in 2010. *Journal of spatial information science*, 2010(1), 3–20.

Goodchild, M.F. and Haining, R.P. (2004). GIS and spatial data analysis: Converging perspectives. *Papers in Regional Science*, 83(1), 363–385.

Haining, R. (2014). Spatial Data and Statistical Methods: A Chronological Overview. In *Handbook of Regional Science*. Berlin & Heidelberg: Springer, 1277–1294.

Harris, R., O'Sullivan, D., Gahegan, M., Charlton, M., Comber, L., Longley, P., Brunsdon, C., Malleson, N., Heppenstall, A., Singleton, A. and Arribas-Bel, D. (2017). More bark than bytes? Reflections on 21+ years of geocomputation. *Environment and Planning B: Urban Analytics and City Science*, 44(4), 598–617.

Harford, T. (2014). Big data: are we making a big mistake? *FT Magazine*, March 28, 2014. Retrieved from: http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0. html *[Last checked: 31 May 2018]*.

Harvey, D. (1972). Revolutionary and counter revolutionary theory in geography and the problem of ghetto formation. *Antipode*, 4(2), 1–13.

Johnston, R., Harris, R., Jones, K., Manley, D., Sabel, C.E. and Wang, W.W. (2014). Mutual misunderstanding and avoidance, misrepresentations and disciplinary politics: spatial science and quantitative analysis in (United Kingdom) geographical curricula. *Dialogues in Human Geography*, 4(1), 3–25.

Johnston, R., Harris, R., Jones, K., Manley, D., Sabel, C.E. and Wang, W.W. (2014). One step forward but two steps back to the proper appreciation of spatial science. *Dialogues in Human Geography*, 4(1), 59–69.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262–267.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. SAGE.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutman, M., Jebara, T., King, G., Macy, M. Roy, D. and van Alstyne, M. (2009). Life in the network: the coming of age of computational social science. *Science* 323(5915), 721–23. doi:10.1126/science.1167742.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436.

Ley, D. and Samuels, M.S. (eds.) (1978). *Humanistic geography*. London: Croom Helm, 337.

Longley, P.A. (2003). Geographical Information Systems: developments in socio-economic data infrastructures. *Progress in Human Geography*, 27(1), 114–121.

Longley, P.A. (2004). Geographical Information Systems: on modelling and representation. *Progress in Human Geography*, 28(1), 108–116.

Loukides, M. (2011). *What is data science?*. O'Reilly Media, Inc.

Lu, Y. and Henning, K.S. (2013). Are statisticians cold-blooded bosses? a new perspective on the 'old' concept of statistical population. *Teaching Statistics*, 35(1), 66–71.

Naur, P. (1974). *Concise survey of computer methods*. Lund: Studentlitteratur.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.

Openshaw, S. (1994). "Commentaries." *Environment and Planning A: Economy and Space*, 26(4), 499–508. doi:10.1068/a260499

Openshaw, S. and Abrahart, R.J. (2000). *GeoComputation*. London: Taylor & Francis.

O'Sullivan, D. (2005). Geographical information science: time changes everything. *Progress in Human Geography*, 29(6), 749–756.

O'Sullivan, D. (2008). Geographical information science: agent-based models. *Progress in Human Geography*, 32(4), 541–550.

O'Sullivan, D. (2014). Don't panic. The need for change and for curricular pluralism. *Dialogues in Human Geography*, 4(1), 39–44.

O'Sullivan, D. and Manson, S.M. (2015). Do physicists have geography envy? And what can geographers learn from it?. *Annals of the Association of American Geographers*, 105(4), 704–722.

Phithakkitnukoon, S., Wolf, M.I., Offenhuber, D., Lee, D., Biderman, A. and Ratti, C. (2013). Tracking trash. *IEEE Pervasive Computing*, 12(2), 38–48.

Poon, J.P. (2003). Quantitative methods: Producing quantitative methods narratives. *Progress in Human Geography*, 27(6), 753–762.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R. and Strogatz, S.H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PLoS One*, 5(12):1–6.

Reades, J., Calabrese, F., Sevtsuk, A. and Ratti, C. (2007). Cellular Census: Explorations in Urban Data Collection. *IEEE Pervasive Computing*, 6(3):30–38.

Singleton, A. (2014). Learning to code. *Geographical Magazine*, January 2014, 77.

Schutt, R. and O'Neil, C. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.

Silver, N. (2012). *The signal and the noise: the art and science of prediction.* London, UK: Penguin.

Taylor, P. J. (1990). Editorial comment GKS. *Political Geography Quarterly*, 9(3), 211-212.

Thatcher, J., O'Sullivan, D. and Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment & Planning D: Society and Space*, 34(6), 990–1006.

Thomas, R.W. (1990). Quantitative methods: input-output analysis as macroeconomic geography. *Progress in Human Geography*, 14(3), 404–419.

Torrens, P.M. (2010). Geography and computational social science. *GeoJournal*, 75(2), 133–148.

Tukey, J.W. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1), 1–67.

Unwin, D.J. (1995). Geographical information systems and the problem of 'error and uncertainty'. *Progress in Human Geography*, 19(4), 549–558.

Unwin, D.J. (1996). GIS, spatial analysis and spatial statistics. *Progress in Human Geography*, 20(4), 540–551.

US Census Bureau (2018). 2012-2016 ACS 5-year Estimates, Summary Files. Retrieved from: https://www2.census.gov/programs-surveys/acs/summary_file/2016/data/5_year_entire_sf/ *[Last checked 14 July 2018]*

Varian, H. (2009). Commentary: Hal Varian on how the Web challenges managers. *McKinsey Quarterly January 2009*, Retrieved from: https://www.mckinsey.com/industries/high-tech/our-insights/hal-varian-on-how-the-web-challenges-managers *[Last checked 31 May 2018]*

Watts, D.J. (2007). A twenty-first century science. *Nature*, 445(7127), 489. doi: 10.1038/445489a.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23.

Wickham, H. and Grolemund, G. (2016). *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media, Inc.

Winter, L. (2017). For Self-Driving Cars, There's Big Meaning Behind One Big Number: 4 Terabytes. *Intel Newsroom*, 14 April 2017. Retrieved from: https://newsroom.intel.com/editorials/self-driving-cars-big-meaning-behind-one-number-4-terabytes/ *[Last checked 1 June 2018]*.

Wrigley, N. (1982). Quantitative methods: developments in discrete choice modelling. *Progress in Geography*, 6(4), 547–562.

Wrigley, N. (1986). Quantitative methods: the era of longitudinal data analysis. *Progress in Human Geography*, 10(1), 84–102.

Wu, J. (1997). Statistics = Data Science? Retrieved from: https://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf *[Last checked 1 June 2018]*

Wyly, E. (2014). Automated (post) positivism. *Urban Geography*, 35(5), 669–690.

Wyly, E. (2014). The new quantitative revolution. *Dialogues in Human Geography*, 4(1), 26–38.