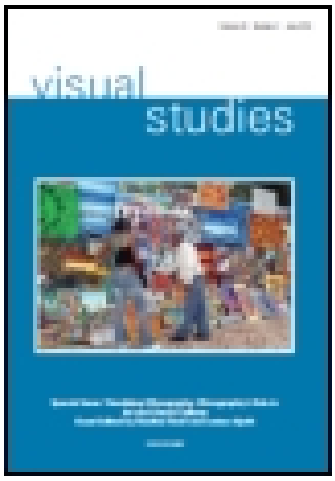

Visual Studies
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rvst20

Rethinking 'big data' as visual knowledge: the sublime and the diagrammatic in data visualisation

Anthony McCosker & Rowan Wilken

Published online: 07 Apr 2014.

To cite this article: Anthony McCosker & Rowan Wilken (2014) Rethinking 'big data' as visual knowledge: the sublime and the diagrammatic in data visualisation, Visual Studies, 29:2, 155-164, DOI: 10.1080/1472586X.2014.887268

To link to this article: http://dx.doi.org/10.1080/1472586X.2014.887268

Visual Studies, 2014
Vol. 29, No. 2, 155–164, http://dx.doi.org/10.1080/1472586X.2014.887268

Routledge
Taylor & Francis Group

Rethinking 'big data' as visual knowledge: the sublime and the diagrammatic in data visualisation

ANTHONY McCOSKER and ROWAN WILKEN

Informational data, we are told, are proliferating ever more rapidly and with increasing complexity. In an age of 'big data' we are seeing a broad reaching, and often uncritical fascination with data visualisation and its potential for knowledge generation. At its extreme this represents a fantasy of knowing, or total knowledge. Nonetheless, for those working in visual anthropology, big data and data visualisation offer significant extensions to our ways of knowing and our categories of knowledge. In this article we probe the fascination and potential of data visualisation and its relevance for understanding human experience, social relations and networks. First, we argue that the celebration of informational aesthetics can be understood as a version of the Kantian mathematical sublime. Extending this analysis, we argue that productive possibilities for thinking about data visualisation are to be found in Deleuze's engagement with the diagram. The diagram, for Deleuze, does not represent but rather operates both as expression and problem resolution. It is incomplete in the dual sense of never capturing the totality of the object and in its dynamism. This approach points to the merits of this investment in data visualisation (the way it works as expression and problem resolution), but highlights the need to be cautious about fetishising the sublimity of 'beautiful data'.

> Figures are 'the natural allies' of the sublime. (Longinus quoted in Guerlac 1985, 278)

> The operation of the diagram, its function, as Bacon says, is to 'suggest'. (Deleuze 1993, 194)

INTRODUCTION

Informational data, we are told, are proliferating ever more rapidly and with increasing complexity. The result, for many commentators (boyd and Crawford 2012), is that we have entered an age of 'big (social) data' which stages large, complex sets of information as a pressing problem in search of new methodological and analytical approaches. Across internet and media-related industries and a wide range of academic fields a fascination with data visualisation and its potential for knowledge generation has emerged. For instance, Dodge (2004) has explored the 'cybergeography' that developed in the early 2000s to capture the multitude of informational flows related to computer networks, culminating in Dodge and Kitchin's *Atlas of Cyberspace* (2001) and *Mapping Cyberspace* (2001). More recent publications, such as Lima's (2011) *Visual Complexity: Mapping Patterns of Information*, and Steele and Iliinsky's (2010) *Beautiful Visualisation: Looking at Data through the Eyes of Experts*, exemplify the new fields working to capture and present complex information about human experience and social relations and networks at a glance. Dominating many of these accounts is an emphasis on the explanatory power of beauty and the role of the 'dataviz' designer, as artist, mathematician and coder, in extracting information from immense data sets and generating knowledge in the form of innovative visual design.

Data visualisation has in this way become an attractive, but contentious tool for social analysis, challenging or extending the methodologies of fields as diverse as digital ethnography, data journalism, digital humanities and computer science. In particular, those working within anthropology recognise the importance of the visual as a way of introducing 'significant additions to how anthropologists define their ways of knowing' including rethinking categories of knowledge in relation to science and many technologies of the visual (Banks and Morphy 1997; Pink 2006). While there is recognition of the importance of information technology within anthropology, some anthropologists caution that in addition to approaching 'information technology as a

Anthony McCosker lectures in Media and Communications in the Faculty of Health, Arts and Design at Swinburne University, Melbourne, Australia. His research explores media affect, media technology, digital and visual cultures and social media practices and publics. He is author of *Intensive Media: Aversive Affect and Visual Culture* (2013) and has published a number of book chapters along with articles in journals such as *Information, Communication and Society, Convergence, Sexualities, Continuum* and *Transformations*.
Rowan Wilken is a Senior Lecturer in Media and Communications, Swinburne University of Technology, Melbourne, Australia, and holds an Australian Research Council funded research fellowship in the Swinburne Institute for Social Research. His present research interests include: mobile and locative media, digital technologies and culture, theories and practices of everyday life, and old and new media. He has published widely on mobile and location-based media. He is the co-editor (with Gerard Goggin) of *Locative Media* (forthcoming 2014) and *Mobile Technology and Place* (2012), and is author of *Teletechnologies, Place, and Community* (2011).

tool for anthropological analysis' (Taylor 1998, 537), it is crucial that we look ethnographically at how computer software is put into practice (Born 1997). This is because what is important is not simply the visual rendering of anthropological or ethnographic data, but 'a range of culturally inflected relationships enmeshed and encoded in the visual' (MacDougall 1997, 288).

This article begins by examining the problematic celebration of beauty in data visualisation before exploring its generative potential in its diagrammatic ability to pose problems and to suggest. As we argue in the first part of the article, the encounter offered by the visualisation of big data can be understood as a version of the Kantian mathematical sublime, a confrontation with 'something vast, elaborate, or complex enough to overwhelm its powers' expressed as an object of unity (Pillow 2000, 74). Certainly, the celebration of beauty in data visualisation indicates the prominence of sensation and intensity, an experience of both pleasure and anxiety in the visual encounter with big data. However, this sublime sensation is only partially revealing of the productive value of data visualisation. In the second part we argue that rather than the figure or end product, the potential of data visualisation can be located in the processes of 'diagramming' and data cleaning, those steps that take place in the gap between the extraction and abstraction of data. Our contention is that the diagram, as conceptualised by Deleuze, offers productive possibilities for understanding the work of data visualisation, if we can acknowledge the element of the sublime in the dataviz figure and move away from its assumed representational qualities. The diagram, for Deleuze, does not represent but rather operates both as expression and problem resolution. It is incomplete in the dual sense of never capturing the totality of the object, and in its dynamism. This approach points to the merits of this investment in data visualisation (the way it works as expression and problem resolution), but highlights the need to be cautious about fetishising the sublimity of 'beautiful data' (Segaran and Hammerbacher 2009).

VISUALISING 'BIG DATA'

While the origins of the phrase remain hazy (Lohr 2013), 'big data' has become a fashionable term applying to 'information that can't be processed or analysed' using traditional tools (Zikopoulos et al. 2012, 3). The emergence of 'big data' is generally seen as the result of developments across a range of technologies, from digital sensors, computer networks, data storage, cluster computer systems and cloud computing facilities (Bryant, Katz, and Lazowska 2008, 2–3), and modes and methods of analysis, or what Berry and others refer to as the 'computational turn' in the humanities and social sciences (Berry 2011; Lazer et al. 2009). Beyond shifts in volume or scale, big data has been defined by three characteristics: volume (the amount of data generated), variety (the multiplying of, and ability to cross-reference, various types and sources of data) and velocity (the speed at which data is flowing) (Zikopoulos et al. 2012, 5–9). These interconnected elements – the 'three v's', as Zikopoulos et al. refer to them – lead some commentators to suggest that 'a radically new kind of "knowledge infrastructure" is materializing' (Bollier 2010, 1) with far-reaching implications for 'all facets of commerce, science, society, and national security' (Bryant, Katz, and Lazowska 2008, 2).

For the digital humanities and social sciences there is an uneasy sense in which big data offers a new kind of access to human experiences, relations and social activities. Numerous pronouncements have been made regarding the 'promises' of 'big data' – 'Big Data is a tagline for a process that has the potential to transform everything' (Jon Kleinberg cited in Lohr 2012; see also Anderson 2008; Lazer et al. 2009; Bollier 2010; World Economic Forum 2012; Smolan and Erwitt 2012); but tempering these have been detailed discussion of the possible 'perils' or challenges facing big data and its analysis (boyd and Crawford 2012; Manovich 2011; Puschmann and Burgess 2013). Data visualisation lies at the heart of this ambivalence, offering on the one hand broad, untrained and immediate access to the knowledge assumed to be contained within big data, but also the ability to test its representational capacity. As Berry argues: 'At all levels of society, people will increasingly have to turn data and information into usable computational forms in order to understand it at all' (Berry 2011, 15). And this echoes an earlier piece in *Wired* magazine, which suggested that 'the biggest challenge of the Petabyte Age won't be storing all that data, it'll be figuring out how to make sense of it' (Horowitz 2008). This job, it is argued, lies in the domain of data visualisation, and in the hands of a 'new breed' of dataviz designers (Bollier 2010; Horowitz 2008).

The close relationship between big data and data visualisation is significant in that it positions visualisation techniques and tools – which, like computational analysis of data, have a long history (Tufte 1990, 2001, 2006) – as vital for the visual knowledge they are able to produce. Like the ethnographer in the field, or the 'artist as ethnographer' to use Hal Foster's (1996) phrase, the dataviz designer plays a pivotal, productive role in knowledge generation. As one proponent writes, data visualisation offers a way

'to find things that you had no theory about and no statistical models to identify, but with visualization it jumps right out at you and says, "This is bizarre"' (Stensrud quoted in Bollier 2010, 11). This accords with the third element in boyd and Crawford's tripartite definition of 'big data' as 'a cultural, technological, and scholarly phenomenon' at the intersection of technology, analysis and mythology (2012, 663). Mythology here indicates the widespread belief that big data and data visualisation 'offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy' (boyd and Crawford 2012, 663). Another way of accounting for this aspect of big data and its visualisation is through the Kantian notion of the mathematical sublime.

DATA VISUALISATION AND THE SUBLIME

A striking feature of existing scholarship on data visualisation is the strong emphasis placed on beauty, particularly in the many book-length studies. For instance, there is *Beautiful Visualization* (Steele and Iliinsky 2010), which contains the chapters 'On Beauty' (Iliinsky 2010), 'Finding Beautiful Insights' (Perer 2010) and 'Beautiful History' (Wattenberg and Viégas 2010), Manuel Lima's chapter 'Complex Beauty' in his book *Visual Complexity* (Lima 2011), as well as *Beautiful Data* (Segaran and Hammerbacher 2009) and *Beautiful Evidence* (Tufte 2006), and the data visualisation website, 'The Information is Beautiful Awards' (www.informationisbeautiful.com). We see in this emphasis on beauty a kind of aesthetic engagement with big data, a form of knowledge encounter that turns on the complexity and aura of an unimaginable object. It is an important aspect of what data visualisation is thought to do in managing big data. For Edward Tufte, data visualisation graphics 'reveal data' through beauty, which he equates with clarity, to communicate with more precision than 'conventional statistical computations' (Tufte 2001, 13). Colour, as one tool for achieving beauty and precision in this context, is thus described as the 'Cinderella of data visualization' (Driscoll 2010). This accords with Edmund Burke's eighteenth century distinction between 'clearness', which he associates with beauty (for instance in the form of 'colours clear and bright'), and 'obscurity', which he associates with the sublime (Burke 2008, 59–60, 117).

Picking up on Burke's (2008, 135–136) later claim that 'visual objects of great dimensions are Sublime', we would suggest that this emphasis on aesthetic beauty reveals a key aspect of the general encounter with data visualisation but fails to illuminate its generative potential. A more revealing account of data visualisation follows from Immanuel Kant's theories of the mathematical sublime, as developed in *Critique of Judgement* (2007, originally published 1790), and its extension through Deleuze to the diagram, to which we will return below.

Kant's Theory of the 'Mathematical Sublime'

The notion of the mathematical sublime pertains to encounters we have with extreme magnitude or vastness. It is a concept that owes a debt to Burke's (2008, 73) understanding that 'greatness of dimension is a powerful cause of the sublime'. As Burke (2008, 73) explains: 'some large objects are so continued to any indefinite number, that the imagination meets no check', and thus, by 'not being able to perceive the bounds of many things, they seem to be infinite, and they produce the same effects as if they were really so'. Kant is more precise in his notion of the 'mathematical sublime, which he defines in the *Critique of Judgement* as the 'estimation of magnitude by means of concepts of number' (2007, sec. 26, 251). Kant's core argument is that the apprehension of magnitude 'is indeed possible, but not its comprehension in an intuition of the imagination (i.e., it is not possible by means of a *comprehensio aesthetica*, though quite so by means of a *comprehensio logica* in a numerical concept)' (2007, sec. 26, 254). While Burke and Kant were writing in the eighteenth century, this concept has a renewed relevance for contemporary informational societies rich in the flow of big data through which we might attempt to comprehend human relations and activities at a large scale.

For example, there have been numerous approximations and attempts at calculation of the magnitude or vastness of computer connectivity and speed. These endeavours range from estimates of the number of networked computers and computer users worldwide, as well as the number of domain names, websites and hits to these websites, to calculations of data transfer speed, computational cycles per second (or fractions thereof), data storage capacity and so forth. However, such estimations sometimes do little to aid human comprehension of the magnitude of connectivity, the speed of transfer occurring across these networks, and the data processing capacities required to facilitate 'big social data' interactions. Big data evade such attempts at representation, just as they evade processing and analysis via 'traditional processes or tools' (Zikopoulos et al. 2012, 3). In Kantian terms, any such attempt 'must be the aesthetic estimation of magnitude in which we get at once a feeling of the effort towards a comprehension

that exceeds the faculty of imagination for mentally grasping the progressive apprehension in a whole of intuition' (Kant 2007, sec. 26, 256).

It is also important to remember that 'instead of the object, it is rather the cast of the mind in appreciating it that we have to estimate as sublime' (Kant 2007, sec. 26, 256). It is via this cast of the mind, Kant argues, that 'in its estimate of a thing as sublime [the imagination] refers that faculty to reason to bring out its subjective accord with ideas of reason' (sec. 26, 256) because this point of excess 'is like an abyss' in which the imagination 'fears to lose itself' (sec. 27, 258). Within such a frame, we can argue that the contemporary fascination with 'big data', and attempts to account for the scale, speed and effects of globalised networked computing and associated social interactions, is the product of a sublime 'cast of mind' appealing to techniques of reason and rationalisation in the face of the conflicting sense of fear and pleasure of big data.

According to Kant's formulation, imagination and reason are countervailing forces, simultaneously producing feelings of displeasure *and* awakened pleasure:

> The feeling of the sublime is … at once a feeling of displeasure, arising from the inadequacy of imagination in the aesthetic estimation of magnitude to attain to its estimation by reason, and a simultaneously awakened pleasure, arising from this very judgement of the inadequacy of the greatest faculty of sense, being in accord with ideas of reason, so far as the effort to attain these is for us a law. (Kant 2007, sec. 27, 258)

The key point, as David Nye explains, is that, 'in the presence of this apparent infinity' or magnitude, 'Kant's subject experiences weakness and insignificance, but then recuperates a sense of superior self-worth, because the mind is able to conceive something larger and more powerful than the senses can grasp' (Nye 1994, 7; see also Crowther 1989, 78–151; McMahon 2004, 19; Lap-Chuen 1998, 25).

Thus, Kant's argument is that the subject facing the sublime experience goes through a transformation from humiliation and awe to a heightened sense of the power of reason in comprehending the phenomenon just experienced (Nye 1994). Or, as John Baillie (1996, 89) puts it in an essay from 1747, 'vast objects occasion vast sensations, and vast sensations give the mind a higher idea of her own powers'. Elaborating on this point, Kirk Pillow suggests that, 'imagination runs into difficulty in trying to comprehend an object as a unity – the more parts, the more complex, the more difficult the task – whenever its faces something vast, elaborate, or complex enough to overwhelm its powers' (Pillow 2000, 74).

Kant discusses this 'maximum of comprehension' in relation to our inability to grasp the vastness of the scale of the great pyramids of Egypt, or the complexity of St. Peter's basilica. The observer of such structures, Kant argues, has 'a feeling' of 'the inadequacy of his imagination for presenting the idea of a whole' (Kant 2007, sec. 25, 252). This passage emphasises the idea of an unattainable demand for unity that is associated with sublime reflection, and is instructive for thinking about the ambitions of many data visualisation projects, and the encounter they occasion.

Data Visualisation and the 'Mathematical Sublime'

Contrary to Chris Anderson's (2008) view that big data 'calls for an entirely different approach, one that requires us to lose the tether of data as something that can be visualized in its totality', we suggest that many attempts at the visualisation of big data can be interpreted as perpetuating the sublime fantasy of 'mapping the unmappable' (King 2000), and of pursuing the desire for knowing, or total knowledge, which seeks to 'objectivize with no remainder' (Derrida 1996, 68). This desire is realised, in effect, in the many attempts at mapping complicated networked social media interactions, such as, for instance, in network traffic maps, like Stephen Coast's 2001 visualisation of global router connectivity, which tracked 32 000 Internet Protocol (IP) addresses (Lima 2011, 121), or data visualisation projects that seek to visualise national blogging behaviours and their associated network links (see, for example, Lima 2011, 98–101). When we look at these visualisations, to take and adapt a passage by Pillow (2000, 72) on the mathematical sublime and art, a judgement of the mathematical sublime, in response to the question of the vastness and complexity of the internet, will seek (and fail) to unify as a whole the complexity of the phenomena represented. It does this 'by means of an imagination that emulates the example of reason in reaching [for] a maximum' (Kant quoted in Pillow 2000, 72).

In a 2002 unpublished article, media theorist Lev Manovich argues against the position sketched above. Rather than being characterised as sublime, he argues, 'data visualization art is concerned with the anti-sublime' (Manovich 2002). As Warren Sack (2007) explains, by 'anti-sublime', Manovich is referencing various non-art contexts in which data visualisation

attempts to 'create "user friendly" interfaces to huge amounts of data', and in which 'the primary measure of a good visualization should be that which can assist a user to perform a task more quickly or more efficiently than the user could do without the visualization'. In this Manovich would appear to be echoing the sentiments of Edward Tufte (2006, 9), who describes 'evidence illustration' as a moral activity with clarity of visual representation one of its core concerns (see also Tufte 1990, 2001).

There are, however, at least two complications with Manovich's assessment. The first concerns how we read data visualisations as images. In his book *American Technological Sublime*, David Nye (1994, 7) summarises the notion of the mathematical sublime as 'the encounter with extreme magnitude or vastness, such as the view from a mountain'. By following this particular conceptualisation, data visualisations of national and international networked telecommunications traffic, such as those included in Andrew Marr's 2008 six-part television miniseries *Britain From Above*, as well as those of other complex phenomena as gathered together in book form by Manuel Lima (2011) and by Julie Steele and Noah Iliinsky (2010), can be interpreted as contemporary – albeit more abstract – equivalents of earlier representations of experiences of the mathematical sublime, such as Caspar David Friedrich's (1818) painting *Wanderer Above a Sea of Fog*, or Albert Bierstadt's later image *The Matterhorn* (c. 1875).

However, while many data visualisations may strive to make clarity of representation their ideal, in many instances this ideal is either not fulfilled, or the image is celebrated for its aesthetic dimensions and visual complexity over the information it seeks to convey. In contrast to Manovich, then, we want to suggest that a great many of the representations gathered in these celebratory anthologies can be interpreted not only as representations of experiences of the mathematical sublime, but also as the source of sublime responses in their own right. A case in point is a September 2011 visualisation produced by mobile social media company Foursquare to mark the passing of one-billion global check-ins using their service (see http://blog.foursquare.com/2011/09/20/billion/). The visualisation depicts worldwide Foursquare check-ins over the course of a week. Each check-in is represented by a single dot, and every dot is colour-coded to represent different types of check-in venues (food, shops, travel spots). Given that these dots are overlayed on a large-scale Mercator projection of the world, and the 24/7 cycle of global check-ins is compressed into a 55-second video, the effect the visualisation creates is of a fluid, pointillist depiction of waxing-and-waning dots that are aesthetically pleasing, but which are experienced at such a scale and speed that comprehension is difficult. Even with the two magnifications of check-ins that occur during the video, and which permit an expanded view of activity in São Paulo and Tokyo, precision is lacking. The overall result is impressionistic at best, creating a sense of 'beautiful complexity', but as a simple factor of global scale without specific analytic insight. Such is the informational density and visual complexity of this visualisation that its content 'requires the interpretative powers of a sublime reflection' (Pillow 2000, 87). Reflection on the aesthetic ideas that are carried by and embodied in this visualisation 'assumes the form of a judgment of mathematical sublimity' (Pillow 2000, 72). The consequence is that, rather than bring greater visual acuity or clarity of understanding to the vastness of global check-ins (in the case of the Foursquare video visualisation), this representation arguably performs another function: it further reinforces the sense of 'unknowability' that has become associated with sublime experiences of phenomena connected with 'big data'.

The second complication with Manovich's earlier assessment of data visualisation relates to the difficulties that follow if, unlike Burke, we take Kant's lead in not drawing a clear distinction between the sublime and the beautiful (Kant 2003, 47–48). While Kant leaves this issue undeveloped, this acceptance of blurred categories or multiple kinds of sublime experience is instructive in the present context in that it draws out the range of at times simultaneous and seemingly conflicting emotions and responses that we experience in the face of particular phenomena, including visualisations of big data projects. Most importantly here, Kant's conception of the sublime as 'in turns of different kinds' (2003, 47–48) highlights how clarity and obscurity (to use Burke's terms) are not always readily separable or distinct categories. To understand data visualisation as process, as a creative encounter with the potential held 'within' big data, is to move away from the sublime beauty of the visual figure itself towards an investment in the act of diagramming, and in the diagram that precedes and produces the figure. In other words, where the uncritical celebration of sublime beauty in the dataviz figure might bring about a new level of apprehension of vastness and scale in human experience and social relations, it is in the diagram that we can develop the means for generating understanding.

DATA VISUALISATION AND THE DIAGRAMMATIC

The promise of data visualisation as a form of visual knowledge has been closely connected to the sublime encounter. But, as we argue in this final section, its

potential lies with its diagrammatic capacities. What is interesting about data visualisation, starting with the spatial topologies and relationality of network graphs, and extending through the complex aesthetic character of dataviz design, can be found in its distributions of colour, shape, pattern and line in ways that introduce and highlight 'intensive' properties. These are properties such as tone, density and sparseness, speed and slowness, and unlike extensive properties are not easily divisible and measureable. While such intensive properties can obscure, or simply present the sublime sense of ungraspable scale, they can also be put to use in the posing of problems, in the process of diagrammatic thinking or experimentation. Big data is put to productive use as a way of 'seeing' and engaging with the distribution of forces and relations only when we are able to both read the extensive properties available within the image and to encounter and relate to its particular distributions of forces and relations. In this, we also have to acknowledge the reflexive practises of the dataviz designer as central to this form of visual knowledge.

Our point here is that the material process of creation (their 'complex materiality' – Born 1997, 140), the distribution of forces and the manipulation of sensation in data visualisation – the expression of intensive as well as extensive properties – point towards the diagram and diagrammatical thinking as the aspect of data visualisation with the most potential for generating effective visual modes of knowing human experience, social activity and relations. This is to say that the diagram is the starting point for the production of knowledge from big data, rather than the end point in the process. And it is the simple recourse to visual or sublime beauty that most clearly diminishes this productive process. In common sense terms, the diagram functions as the plan, map, drawing, illustration or figure that leads to the construction or understanding of some object or phenomenon in the world. However, theoretical work regarding diagrams, particularly associated with the writings of Gilles Deleuze, has made a significant intervention into these conventional notions of the diagram as a visual tool of resemblance.

The practices and 'language' of diagraming have been augmented by the design and visualisation capacities of computational technologies applicable to fields such as cognitive and engineering sciences, architecture, social sciences, digital art and journalism, and by new 'sources' of information made available through the generation of 'big (social) data'. In these settings emphasis is often placed on the applications available to emerging forms of 'visual knowledge' (De Landa 1998, 30). In architecture, for instance, value is placed on the use of the diagram 'to organize and maintain different types of information in a single graphic configuration or set of configurations' (Confurius 2000, 5). Moreover, 'the diagram represents an open, generative process of design that promises to mediate between the disordered complexity of the information society and the demands of architecture' (Confurius 2000, 5). For Somol, the diagram in architecture 'is a performative rather than a representational device (i.e., it is a tool of the virtual rather than the real)' (Somol 1999, 8). Somol explains that: 'Diagrammatic work is projective in that it opens new (or, more accurately, "virtual") territories for practice' (Somol 1999, 23). It enables a continued process of construction and reconstruction 'on the fly' to 'momentarily subvert the normative course of architectural practice and inhabitation' (Massumi 2011, 101–102).

In some areas of recent cognitive science, emphasis is placed on the use of the diagram as a tool for visualising thought, for rendering complex cognition 'available' through the visual – symbolic and iconic – language of diagrams (Hegarty 2011; Tversky 2011). And this resonates with classic accounts of information visualisation, where clarity of thought is equated with beauty in design which, in the face of the figures of big data culminates in the sublime encounter. By contrast, Deleuze's approach to the diagram is non-cognitive or 'non-representational' and departs from the Kantian concern with reason and rationality, particularly in relation to aesthetic judgement. The diagram enters Deleuze's work in two key contexts: in art as the distribution of forces and sensation (in the work of Francis Bacon) and in the social apparatuses, regimes and techniques of power (what he calls the 'abstract machine' of sovereign and disciplinary societies and 'societies of control') (Deleuze 1988, 2003). In the context of Bacon's art, Deleuze sees in the diagram 'the operative set of asignifying and nonrepresentative lines and zones, line-strokes and color-patches' (Deleuze 2003, 101). Its function is not to represent or visualise some essence, but to 'suggest', and in this role, to generate (Deleuze 2003, 101). Similarly, in his book on Foucault, the diagram is defined as:

> the presentation of the relations between forces unique to a particular formation; it is the distribution of power to affect and the power to be affected; it is the mixing of non-formalized pure functions and unformed pure matter. (Deleuze 1988, 72–73)

Diagrams indicate here the historical 'social fields' that, while highly unstable and continually changing, become

manifest as social forms, for example, in figures, statements, visual arrangements or forms of architecture (Deleuze 1988, 35). In a sense that has always applied to the Foucauldian analysis of discourse as knowledge and power, a diagram is also 'any ideal standard of comparison that might be applied to a figure or an objective form to overlay a value judgment on it' (Massumi 2011, 99–100).

We wish to emphasise two elements of the diagram in Deleuze's thought for what it can bring to our understanding of data visualisation: the non-representational, or more-than-representational, 'intensive' properties of the diagram; and, perhaps more significantly, the role of the diagram in formulating problems. As Manual De Landa puts it in his reflection on thought and the diagram in Deleuze's work: 'It is only through skilful problem-posing that we can begin to think diagrammatically' (1998, 34).

Under different guises – illustration, graph theory, information visualisation, information graphics – data visualisation has a long history that extends well beyond digital computational technologies, but has been enhanced by those technologies able to capture and manage large data sets and complex relations. With the development of contemporary data visualisation techniques and a range of increasingly popular applications, a quantifiable shift is commonly thought to have taken place, particularly over the past decade. Histograms, pie charts, scatter plots and other traditional forms of information presentation, mostly invented in the nineteenth century, share much with recent visualisation techniques and forms. For Manovich, these traditional techniques and contemporary data visualisation 'both use the same graphic language: points, lines, curves, simple shapes, and other graphic primitives' (Manovich 2011, 12). Beyond this, however, it is argued that new forms of data visualisation have been pioneered to respond to an information rich society consisting of new types of data sets. Contemporary data visualisations are hence considered denser, more complex and varied, Manovich suggests, because,

> contemporary designers, artists and computer scientists are trying to represent considerably more data than ever before and to represent relations between more dimensions of data than is possible with older graph types such as bar charts (one dimension) or scatter plots (two dimensions). (Manovich 2011, 12)

In Edward Tufte's (1997, 10) now classic text on information visualisation, *Visual Explanations*, his stated aim is to work and rework the visual dimension, extend 'the visual capacities of paper, video, and computer screen', so as to bring about forms able to represent relations, express quantities, and ultimately draw out inferences with the maximum of visual clarity. Clarity of visualisation is equated with clarity of (statistical) thinking. These aims are still apparent with data visualisation, but increasingly, when data visualisation artists talk about their practice they emphasise elements such as beauty, originality and novelty. For Manovich (2011, 12), as artistic and cultural artefacts, 'we expect [data visualisations] to be unique – just as we expect this in fashion, product design, architecture, music, and other cultural fields'. Hence, data visualisation seeks to make innovative use of 'axes and layout, shape, colors, lines, and typography' to achieve novelty, efficiency and informativity (Steele and Iliinsky 2010, 1). By entering into a process of visualising the dynamic properties – the volume, variety and velocity – of big data, dataviz design moves from the extensive or divisible properties of traditional forms of mathematical visual knowledge, towards the intensive. It is in this shift that the diagram in Deleuze's sense properly comes into play.

One influential example of diagrammatic thinking at work in data visualisation lies in the emergence of the network graph or diagram, a class of visualisations that take networked data points, or nodes, and illustrates the links and relationships between them. Ironically, and perhaps indicating the source of the problems faced by data visualisation as a field for exploration and research, network diagrams have become almost iconic, as if they resemble the network function of the internet itself. Rather, it should be understood in reverse: because we have network sciences we have the 'diagram of forces' on which the network graph or figure depends (Deleuze 1988, 74). The iconic 'sociograph' forms the basic diagram, and as early as the mid-twentieth century introduced the potential power of networks within graph theory for explaining relations and movements among objects or people. With the emergence of big data produced through ubiquitous, distributed forms of digital media and communicative exchange, the power and promise of the network diagram has flourished, and has driven a great deal of innovation and attention for data visualisation as a practice. This has been fuelled by the development of simple data visualisation tools such as NodeXL and more sophisticated, often open source, programs able to visualise in three dimensions and in duration as animation. Such tools enable social network analysis, but also affirm the network as diagram.

While it is often the sublime beauty of the figures produced that is celebrated, with their intricate clusters of nodes and webs or explosions of line and colour, it is

as diagram that the network form poses problems and offers a virtual space for exploration, opposing distributed over hierarchical structures. As Alexander Galloway points out, a distributed network is 'never complete, or integral to itself. The lines of a distributed network continue off the diagram' (Galloway 2004, 34). Distributed network diagrams become actualised both in the protocols established to shift packeted data – Transmission Control Protocol/Internet Protocol (TCP/IP) – and in the (dynamic and never finally fixed) activity that takes place on the basis of such protocols. In this way the network diagram as figure, as image, offers on the one hand the fantasy of knowing, or total knowledge in Derrida's (1996) sense. But, on the other hand, it functions as propositional, as the productive use of varied, dynamic and complex data sets. In Deleuze's discussion of the 'logic of sensation' in Francis Bacon's painting, the diagram operates at the threshold of painting in the form of a 'chaos-germ': 'More than an abstraction, the diagram is a field of resonances and virtualities, an abstract which explores an experimental thought of possibilities' (Buci-Glucksman 1998, 35). If Bacon creates resemblances, Deleuze says, it is through the manipulation of manual marks that are 'irrational, involuntary, accidental, free and random' (Deleuze 2003, 100). This is where the creative aligns with the virtual and the diagrammatic, and also indicates the diagram's renewed application in fields such as architecture, where it surpasses the drawing (as pictorial form) as the focus of its creative practice, as a mode of discovery.

The network diagram offers a set of potential lines and points, topological forms through which distributed networks might be conceived and then visualised, but which operates as fluid, dynamic and always as emergent. For instance, some of the many, and expanding, possibilities for thinking distributed networks are presented in Lima's (2011) *Visualising Complexity*, a publication that focuses specifically on visualised network diagrams. Here, new forms are collated under an expansive vocabulary, using a range of aspects of plane, line, colour, shape and dimension, under names such as 'arc diagram', 'centralized burst', 'elliptical implosion', 'segmented radial convergence', amongst many others. While as part of a book-length compendium these visualisations are proffered as figures of the mathematical sublime, in their multiplication of shapes, colour, form, animation and temporal manipulability in practice, they exemplify the experimentation inherent to the diagram. The work of Jer Thorp, Mark Hansen and Jake Porway at the New York Times R&D Lab, with the development of the Cascade Project in 2011, is similarly revealing (http://nytlabs.com/projects/cascade.html). The Cascade project visualises 'sharing events', or how links to news stories are shared across Twitter networks. It uses an animated three-dimensional image that scales in and out to highlight important nodes or edges. The element of time in this animated network diagram is by design 'exploratory', and manipulable, and introduces new problems or questions that might be posed for the data. For instance, it allows attention not only to relations among nodes connected by the act of sharing news story links, but also the influence of time of day or day of week, the speed, intensity and duration of sharing, and the points at which influential actors receive and share links. The project as a whole, in this way, opens up new ways of conceiving the flow of data through distributed processes, the (big) data they produce, and also allows new interventions into the functioning of those networks.

In the construction of the tools that make such visualisations of big data possible, and in the design of the final visualisation itself, diagrammatical thinking is at work. Interestingly, Jer Thorp likens the dataviz design process to writing software:

> You write something and test it, write it and test it, write it and test it. Any project that I work on is really iterative, it steps through different forms, most of which don't work. But I always think about what I would call a 'data first' approach, where I don't have an idea what I want this thing to look like when I start, and then the data sort of informs the design. Usually I start doing some sketching with the numbers and the data, and there are some forms that emerge or some forms that are convenient to make the data readable. (Thorp cited Drell 2012)

In data visualisation, the data clearly matters. But it is important to distinguish between matter, or that which enters the diagram, and its substantiation as the visualisation, figure or objective form. Anna Munster's (2011) discussion of the diagrammatical character of functional magnetic resonance imaging (fMRI) offers a useful parallel here. Like fMRIs, as functional images, many data visualisations, that seek to manage the dynamic characteristics of big data, their volume, variety and velocity (Zikopoulos et al. 2012, 5–9), remind us that 'we are in the middle of something'. That is, to look at such images diagrammatically is to acknowledge that 'There's a lot we aren't seeing, that we miss. And so we have to insert, conjecture, and "back fill"' (Munster 2011); the diagram does not 'demonstrate', but rather casts light on the creative acts through which concepts, constructions and knowledge might emerge.

CONCLUSION

As a tool for social analysis and knowledge generation from within the ever expanding human contexts captured and enmeshed with big data, data visualisation remains contentious. At many points, the visual itself has come to override the 'range of culturally inflected relationships' encoded within (MacDougall 1997, 288). We have framed this often uncritical fascination with data visualisation as a contemporary version of the Kantian mathematical sublime. The pairing of beauty with data visualisation, with the apprehension of the magnitude of human experience and relations, that is, with the sublime, does not necessarily service fields such as visual anthropology as a useful form of visual knowledge. Hence, in the second half of the article, we sought to reconceptualise present understanding of data visualisation by examining the diagram that precedes the visual figure, and the diagrammatic work of the dataviz designer.

As a way out of the bind that faces researchers, commercial and state organisations seeking to utilise big data, data visualisation certainly offers promise, but its potential should be located less in the sublime beauty of the pictorial figure than in the diagram that provides the (virtual) space for experimentation, differentiation and problem-posing. That is, data visualisation is often considered as the end product of a series of cycles moving from the sublime, raw quanta of big data through stages of mining, cleaning, processing and visualising as the final output. However, the visualisation, the figure, does not stand as the final stage in a process of problem-solving, but should be better understood as the actualisation of new ways of problem-posing. If there is an individual skill or artistry to data visualisation, it is in this ability to think diagrammatically, as a form of problem-posing (De Landa 1998, 34). And these processes in turn offer a way forward through data visualisation for areas of the social sciences attempting to generate knowledge of human experience and social relations from within the expanding contexts of big data.

REFERENCES

Anderson, C. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired*, July 16. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

Baillie, J. 1996. "An Essay on the Sublime (1747)." In *The Sublime: A Reader in British Eighteenth-Century Aesthetic Theory*, edited by A. Ashfield and P. De Bolla. Cambridge: Cambridge University Press.

Banks, M., and H. Morphy. 1997. *Rethinking Visual Anthropology*. New Haven, CT: Yale University Press.

Berry, D. M. 2011. "The Computational Turn: Thinking about the Digital Humanities." *Culture Machine* 12: 1–22.

Bollier, D. 2010. *The Promise and Peril of Big Data*. Washington, DC: Aspen Institute.

Born, G. 1997. "Computer Software as Medium: Textuality, Orality and Sociality in an Artificial Intelligence Research Culture." In *Rethinking Visual Anthropology*, edited by M. Banks and H. Morphy. New Haven, CT: Yale University Press.

boyd, d., and K. Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* 15 (5): 662–679. doi:10.1080/1369118X.2012.678878.

Bryant, R. E., R. H. Katz, and E. D. Lazowska. 2008. *Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science, and Society*. http://www.cra.org/ccc/docs/init/Big_Data.pdf.

Buci-Glucksman, C. 1998. "Of the Diagram in Art." *ANY: Architecture New York* 23 (June): 34–36.

Burke, E. 2008. *A Philosophical Enquiry into the Sublime and the Beautiful*. Hoboken, NJ: Routledge.

Confurius, G. 2000. "Editorial." *Daidalos: Architecture, Art, Culture* 74 (October): 4–5.

Crowther, P. 1989. *The Kantian Sublime: From Morality to Art*. Oxford: Clarendon Press.

De Landa, M. 1998. "Deleuze, Diagrams, and the Genesis of Form." *ANY: Architecture New York* 23 (June): 30–34.

Deleuze, G. 1988. *Foucault*. Translated by S. Hand. London: University of Minnesota Press.

Deleuze, G. 1993. "The Diagram." In *The Deleuze Reader*, Edited by C. V. Boundas. New York: Columbia University Press.

Deleuze, G. 2003. *Francis Bacon: The Logic of Sense*. Translated by D. W. Smith. London: Continuum.

Derrida, J. 1996. *Archive Fever: A Freudian Impression*. Translated by. E. Prenowitz. Chicago, IL: University of Chicago Press.

Drell, L. 2012. "This Man Makes Data Look Beautiful." *Mashable,* December 12. http://mashable.com/2012/12/11/data-visualization-jer-thorp/.

Driscoll, M. 2010. "Color: The Cinderella of Data Visualization." In *Beautiful Visualization: Looking at Data through the Eyes of Experts*, edited by J. Steele and N. Iliinsky. Sebastopol, CA: O'Reilly Media.

Foster, H. 1996. *The Return of the Real: The Avant-Garde at the End of the Century*. Cambridge, MA: MIT Press.

Galloway, A. R. 2004. *Protocol: How Control Exists after Decentralization*. Cambridge MA: MIT Press.

Guerlac, S. 1985. "Longinus and the Subject of the Sublime." *New Literary History* 16: 275–289. doi:10.2307/468747.

Hegarty, M. 2011. "The Cognitive Science of Visual-Spatial Displays: Implications for Design." *Topics in Cognitive Science* 3: 446–474. doi:10.1111/j.1756-8765.2011.01150.x.

Horowitz, M. 2008. "Visualizing Big Data: Bar Charts for Words." *Wired*, July 16. http://www.wired.com/science/discoveries/magazine/16-07/pb_visualizing.

Iliinsky, N. 2010. "On Beauty." In *Beautiful Visualization: Looking at Data Through the Eyes of Experts*, edited by J. Steele and N. Iliinsky, Sebastopol, CA: O'Reilly Media.

Kant, I. 2003. *Observations on the Feeling of the Beautiful and the Sublime*. Translated by J. T. Goldthwait. Berkeley, CA: University of California Press.

Kant, I. 2007. *Critique of Judgement*. Translated by J. C. Meredith and edited by N. Walker. Oxford: Oxford University Press.

King, A. B. 2000. *Mapping the Unmappable': Visual Representations of the Internet as Social Constructions*. CSI Working Paper No. WP 00-05. https://scholarworks.iu.edu/dspace/bitstream/handle/2022/171/wp00-05B.html.

Lap-Chuen, T. 1998. *The Sublime: Groundwork Towards a Theory*. Rochester, NY: University of Rochester Press.

Lazer, D., A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis. 2009. "Social Science: Computational Social Science." *Science* 323: 721–723. doi:10.1126/science.1167742.

Lima, M. 2011. *Visual Complexity: Mapping Patterns of Information*. New York: Princeton Architectural Press.

Lohr, S. 2012. "How Big Data Became So Big." *New York Times*, August 11. http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html?_r=0.

Lohr, S. 2013. "The Origins of 'Big Data': An Etymological Detective Story." *New York Times*, February 1. http://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/.

MacDougall, D. 1997. "The Visual in Anthropology." In *Rethinking Visual Anthropology*, edited by M. Banks and H. Morphy, New Haven, CT: Yale University Press.

Manovich, L. 2002. *The Anti-Sublime Ideal in Data Art*. http://www.manovich.net/DOCS/data_art.doc.

Manovich, L. 2011. "Foreword." In *Visual Complexity: Mapping Patterns of Information*, edited by M. Lima. New York: Princeton Architectural Press.

Massumi, B. 2011. *Semblance and Event: Activist Philosophy and the Occurrent Arts*. Cambridge, MA: MIT Press.

McMahon, C. 2004. *Reframing the Theory of the Sublime: Pillars and Modes*. Lewiston: Edwin Mellen Press.

Munster, A. 2011. "Nerves of Data: The Neurological Turn In/Against Networked Media." *Computational Culture* 1. http://computationalculture.net/article/nerves-of-data.

Nye, D. E. 1994. *American Technological Sublime*. Cambridge, MA: MIT Press.

Perer, A. 2010. "Finding Beautiful Insights in the Chaos of Social Network Visualizations." In *Beautiful Visualization: Looking at Data Through the Eyes of Experts*, edited by J. Steele and N. Iliinsky. Sebastopol, CA: O'Reilly Media.

Pillow, K. 2000. *Sublime Understanding: Aesthetic Reflection in Kant and Hegel*. Cambridge, MA: MIT Press.

Pink, S. 2006. *The Future of Visual Anthropology: Engaging the Senses*. New York: Routledge.

Puschmann, C., and J. Burgess. 2013. *The Politics of Twitter Data*. HIIG Discussion Paper Series, Discussion Paper No. 2013-01. Berlin: Alexander von Humboldt Institut für Internet und Gesellschaft (HIIG). http://ssrn.com/abstract+2206225.

Sack, W. 2007. *Aesthetics of Information Visualization*. http://hybrid.usc.edu/SocialComputingLab/Publications/wsack-infoasethetics-illustrated.doc.

Segaran, T., and J. Hammerbacher. 2009. *Beautiful Data: The Stories behind Elegant Data Solutions*. Sebastopol, CA: O'Reilly Media.

Smolan, R., and J. Erwitt. 2012. *The Human Face of Big Data*. Sausalito, CA: Against All Odds Productions.

Somol, R. E. 1999. "Introduction." In *Diagram Diaries*, edited by P. Eisenman. London: Universe Publishing.

Steele, J., and N. Iliinsky, eds. 2010. *Beautiful Visualization: Looking at Data Through the Eyes of Experts*. Sebastopol, CA: O'Reilly Media.

Taylor, L. 1998. "Visual Anthropology Is Dead, Long Live Visual Anthropology!" *American Anthropologist* 100 (2): 534–537. doi:10.1525/aa.1998.100.2.534.

Tufte, E. 1990. *Envisioning Information*. Cheshire, CT: Graphics Press.

Tufte, E. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphic Press.

Tufte, E. 2001. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Tufte, E. 2006. *Beautiful Evidence*. Cheshire, CT: Graphics Press.

Tversky, B. 2011. "Visualizing Thought". *Topics in Cognitive Science* 3: 499–535. doi:10.1111/j.1756-8765.2010.01113.x.

Wattenberg, M., and F. Viégas. 2010. "Beautiful History: Visualizing Wikipedia." In *Beautiful Visualization: Looking at Data Through the Eyes of Experts*, edited by J. Steele and N. Iliinsky. Sebastopol, CA: O'Reilly Media.

World Economic Forum. 2012. *Big Data, Big Impact*. Cologny/Geneva: World Economic Forum. http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf.

Zikopoulos, P. C., C. Eaton, D. deRoos, T. Deutsch, and G. Lapis. 2012. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New York: McGraw Hill.