Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2019)
182, *Part* 2, *pp.* 419–441

Discussion on the meeting on ‘Data visualization’

Peter J. Diggle (*Lancaster University*)
I very much welcome this collection of insightful and complementary papers, Well-designed visualizations can add value both to exploratory analysis by revealing structure in a data set, and to confirmatory analysis as effective communication tools.

On the first point, the papers by Castruccio, Genton and Sun, and by Bowman demonstrate imaginative use of animation and colour to capture variation over time, space and in distribution. A gentle warning about the use of sophisticated visualizations such as these is that their beauty can be beguiling. A consequent risk is that the inexperienced user may pay insufficient attention to the basics. At the meeting, I showed an example from Cleveland (1994) in which changing the aspect ratio of a simple line plot can hide or reveal the typically asymmetric shape of the quasi-cyclic variations in annual sunspot activity. The simple messages in Cleveland (1994), and in Cox (1978), are as relevant today as when they were written.

On the second point, an important question is: communication to whom? The paper by Gabry and his colleagues focused primarily on self-communication, i.e. on providing feedback to the statistician during the various stages of a Bayesian analysis. As someone who is unconvinced that Bayesian inference should be used routinely, I found their discussion and example of prior predictive distributions striking in two ways. On the one hand, the demonstration that priors conventionally accepted as vague can induce implausible priors for the data is salutary. On the other, if you massage your prior until it generates realizations that are concentrated (to a greater or lesser extent) around the data and then ‘turn the Bayesian handle’, in what sense is your inference Bayesian?

Other potential audiences for statistical visualizations include policy makers, public sector workers and the general public. In my own work with colleagues at Lancaster and elsewhere on prevalence mapping, we advocate the use of *predictive probability mapping* for communicating uncertainty to non-statisticians of whatever hue. This consists of plotting at each location of the map the predictive probability that local prevalence exceeds a user-specified threshold.

Fig. 1 shows an example of predictive probability mapping, taken from an on-going multinational programme for the control of onchocerciasis (river blindness) in sub-Saharan Africa (Zoure *et al*., 2014). The programme specifies that areas with prevalence greater than 20% should be prioritized for mass distribution of prophylactic medication. Accordingly, Fig. 1(b) maps the predictive probability that this criterion is met, based on a generalized linear geostatistical model (Diggle *et al*., 1998) fitted to empirical prevalence data obtained from the locations shown on the map. Figs 1(a) and 1(c) show the corresponding maps using a more stringent (10%) or a more relaxed (30%) prevalence threshold for prioritization. We draw two conclusions. Firstly, the different implications for how much of the country should be prioritized for treatment under different prevalence threshold criteria are inescapable. Secondly, the 20% exceedance map forces the reader to recognize that in some areas we simply do not know whether the prioritization criterion has been met.

At the meeting, I also showed an animation, by Dr Emanuele Giorgi, of predictive probability exceedance mapping for malaria prevalence in Chikwawa, southern Malawi, over a 3-year period. This can be viewed at `http://www.lancaster.ac.uk./staff/giorgi/malaria/`. It shows very clearly both the seasonal variation in malaria and the dramatic reduction resulting from a range of government and community-led interventions over the 3-year period. In our experience, rural health workers have appreciated seeing this striking confirmation that their work can have such a positive influence on the health of their communities.

Mark Baillie and Marc Vandemeulebroecke (*Novartis Pharma, Basel*)
We are pleased to second the vote of thanks for the three interesting papers on the important topic of visualizing data and uncertainty (see also Spiegelhalter *et al.* (2011)). We comment on two of the three papers.

Bowman promotes two techniques to display uncertainty: shading and animation. We agree that shading, as introduced by Jackson (2008), can be effective for communicating uncertainty. It is intuitive; it conveys an appropriate sense of ‘fuzziness’; and it discourages a focus on point estimates or threshold-based binary judgements (in or out; significant or not). The graphical representations of the concept of a hypothesis test

0964–1998/19/182419

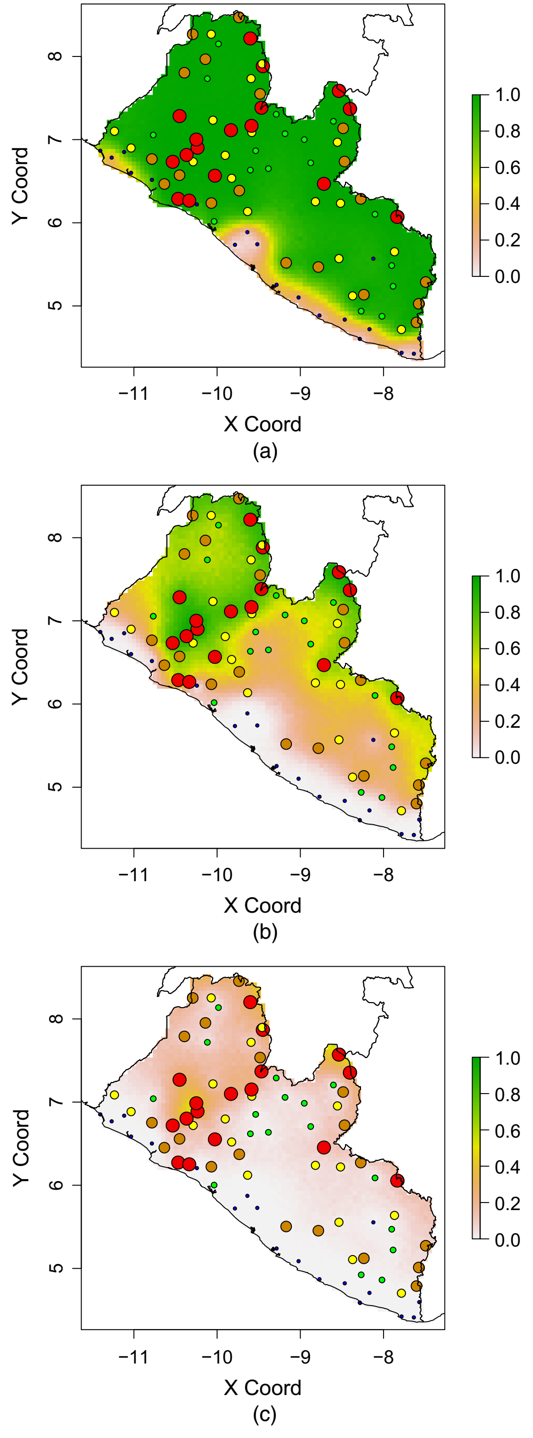


Fig. 1. Predictive probability maps for precontrol prevalence of onchocerciasis (river blindness) in Liberia, for exceedance thresholds (a) 10%, (b) 20% and (c) 30%: ●, observed prevalence, colour and size coded by quintiles of the empirical distribution of observed prevalence over all sampled locations

are interesting: they intuitively convey the idea of checking observed effects against some hypothesis or reference model, taking into account uncertainty but avoiding binary judgements. However, there are also drawbacks. Colour intensity performs poorly for encoding numerical values (Cleveland and McGill, 1985). One may argue that this is a 'feature not a bug' when we intend to convey *lack* of precision. Yet, visual perception of the same intensity may change, e.g. depending on the chosen colour hue. This illustrates that shading, although a powerful idea, may be more difficult to implement *well* than one may think—or than standard techniques. Nevertheless, we think that it can be interesting as 'another tool in the box' if implemented with care.

We are more sceptical of animation. Bowman quotes Tversky *et al*. (2002) who saw it as 'most effective in displaying changes over space and time'. We are tempted to add '... but not uncertainty'. We agree with Bowman that 'if animation is not carefully employed it may simply be a distraction'. Animation is such a powerful visual stimulus that the 'effect' easily trumps the 'message', like a motion picture with more special effects than story line. Conceptually, animation introduces time. Unless it encodes change, this extra 'dimension' appears misleading. By analogy with Wild *et al.* (2011) who (as quoted by Bowman) recommended that plots should 'stay in the same visual space' as the data, we feel that an undue introduction of time violates the 'conceptual space' of the problem. Finally, from a practical perspective, animation forces a pace on the observer: one can no longer scrutinize a display at one's own pace.

For both shading and animation, it would be interesting to see empirical evidence of their effectiveness compared with more standard techniques (Cleveland and McGill, 1985; Tufte, 2001; Heer and Bostock, 2010; Munzner, 2014).

Gabry and his colleagues demonstrate a workflow for the applied Bayesian with visualization at its core. This workflow resonates with our experience as statisticians working in drug development. The competency of visual communication is important to master; it is required at all stages from designing to analysing and reporting clinical trials, organizing information and supporting decision making. We often switch within or across workflows, from learning to confirming (Sheiner, 1997), or from the 'subjective' to the 'objective' (see also Gelman and Hennig (2017)). Effective visualization can help in all these modes. The authors demonstrate this competency at each phase of a Bayesian workflow from data exploration, setting up appropriately calibrated priors, diagnostic model checking to model comparison. This welcome contribution serves as a pedagogical example not only for applied Bayesian inference but also more generally for all quantitative disciplines.

The workflow implicitly demonstrates the important consideration of whom we are communicating to: ourselves as statisticians, domain experts or a more general audience. For example, a series of visualizations facilitate discussion around the theoretical and philosophical implications of 'using the data twice'. Here, we are speaking to ourselves as statisticians. The visualizations serve not only to understand and communicate the risks of adopting such a workflow, but also to communicate suitable approaches for amelioration. This discussion points to an important reminder: it is critical to acknowledge all sources of uncertainty from aleatoric (i.e. statistical variation or intrinsic randomness) to epistemic (i.e. due to lack of knowledge) through the entire workflow, not limiting ourselves to a narrow view of statistical variation alone (Meehl, 1990; Spiegelhalter, 2011; Gelman and Loken, 2014).

What is, however, missing is a demonstration of how to communicate visually the outcome of an analysis, especially for supporting scientific understanding or decision making. At this stage, we often speak to a general audience, where the task of an applied statistician is to translate results and uncertainty into simpler representation(s) that are easily understandable and communicable. As Bowman says,

> 'It is stating the obvious to say that graphical methods play a very important role in both the communication of statistical information and concepts in a manner which is largely free of technical language'.

How do we effectively communicate graphically to a wider audience? To do this with skill we still need to master and apply the 'simple' tools at our disposal (Gordon *et al*., 2015; Margolskee *et al*., 2017; Vandemeulebroecke *et al*., 2018).

Wayne Robert Jones (*Shell*)

The power of visualizing data is something that we statisticians very much take for granted. The ability to visualize data is a fundamental part of the statistician's toolkit and the first thing that we do when we are presented with new data. We already understand the interpretive power of plotting data rather than looking at numbers in a spreadsheet. When we gather together and discuss visualization we are, of course, preaching to the already enlightened. I would like to take the opportunity to step back and to consider the subject of data visualization from a broader perspective.

It is a saddening fact that data visualization is often very much underused and misinterpreted, in both industry and wider society. For example, despite many functions of the information technology industry being exclusively dedicated to the collection and management of data, little, or no, effort is given to visualization. To a large extent, this is unforgivable because the industry has the technological capability to do so. One must reluctantly surmise that this is due to an industrywide lack of understanding of the benefits of data visualization leading to its forming little or no part in the training of information technology professionals. Poor data visualization is, of course, not restricted to the information technology industry. My work in visualizing and interpreting environmental monitoring data (`www.api.org/GWSDAT`) has allowed me to work with environmental regulators from all over the world. They, repeatedly, state that their jobs would be made much easier if their stakeholders could plot the data rather than supplying data on a spreadsheet.

The Royal Statistical Society discussion on data visualization has been an extremely stimulating, thought-provoking discussion on the nuances of expert data visualization for the benefit of a highly educated audience. I would, however, ask my colleagues to consider the bigger picture also. It is my firm belief that we, as statisticians and Society members, have a responsibility to educate the wider public in the more basic aspects of data visualization. This would enable more people to 'see the light' and have a greater positive influence on society.

Alessandro Fassò and Francesco Finazzi (*University of Bergamo*)
The proposal of Castruccio and his colleagues for the assessment of spatiotemporal statistical models using a virtual reality environment has interesting potential. In fact, its implementation on cheap devices, easy to adapt to user data and models, is the key to the success of this approach. Moreover, they have introduced the use of the similarity index SSIM (Wang *et al.*, 2004) as a model diagnostic applied to a global climate model output on a three-dimensional regular grid.

This discussion first focuses on the use of the SSIM-index in statistics in general, suggesting a modified asymmetric version. Then, it motivates a virtual reality approach in modelling four-dimensional global climate observational data on a non-regular grid.

Structural similarity index in statistics
The structural similarity index SSIM is a standard for image and movie fidelity assessment, especially in its local, sliding window and multiscale versions (Wang and Bovik, 2009).

Inspired by Castruccio and his colleagues, the adoption of SSIM in statistics as a general model validation diagnostic is considered here. To do this, SSIM in their formula (1) is simplified assuming that $C_1 = C_2 = 0$, non-zero means and positive variances. Moreover, considering the case where $\mathbf{x} = \hat{\mathbf{y}}$ is a statistical model for $\mathbf{y}$, the following asymmetric version of SSIM is suggested:

$$\mathrm{SSIM}(\hat{\mathbf{y}}|\mathbf{y})^2 = \rho^2_{\hat{\mathbf{y}}\mathbf{y}} \left(1 - \frac{(\mu_{\hat{\mathbf{y}}} - \mu_{\mathbf{y}})^2}{\mu^2_{\mathbf{y}}}\right)_+ \left(1 - \frac{(\sigma_{\hat{\mathbf{y}}} - \sigma_{\mathbf{y}})^2}{\sigma^2_{\mathbf{y}}}\right)_+ \qquad (1)$$

where $\rho_{\hat{\mathbf{y}}\mathbf{y}} = \sigma_{\hat{\mathbf{y}}\mathbf{y}}/(\sigma_{\hat{\mathbf{y}}}\sigma_{\mathbf{y}})$ and $(x)_+$ is the positive part of x.

Virtual reality and the four-dimensional case
The information content of the radiosonde network depicted in Fig. 2 and known as RAOB is addressed here by using a four-dimensional statistical model for temperature profiles (Fassò *et al.*, 2018a, b; Finazzi *et al.*, 2018).

Because of the profile structure of data, a spatiotemporal model for functional data is used and estimated by using an expectation–maximization algorithm derived by Finazzi and Fassò (2014). Moreover, the information gaps (Fassò *et al.*, 2018a) are assessed by computing the kriging estimate, say $\hat{\mathbf{y}}$, on a fine grid. This gives big four-dimensional objects for $\hat{\mathbf{y}}$, its standard deviation and the local asymmetric SSIM of equation (1). Hence the challenge is to represent in a virtual reality environment such large four-dimensional objects.

Thomas King (*Newcastle upon Tyne*)
There is a genre of data visualizations which presents health inequalities by using local underground train network maps. The visualization is targeted therefore at the audience of local health commissioners, to make the case for local action on health inequalities. Social inequalities in health outcomes, and specifically healthy life expectancy, follow a steeper social gradient than life expectancy itself (Marmot, 2010). One might assume that this pattern is regional, with poorer areas seeing worse outcomes and needing greater investment in health services for that reason. In fact, healthy life expectancy varies enormously within municipalities; thence accessible communication could spur local action on health inequalities.

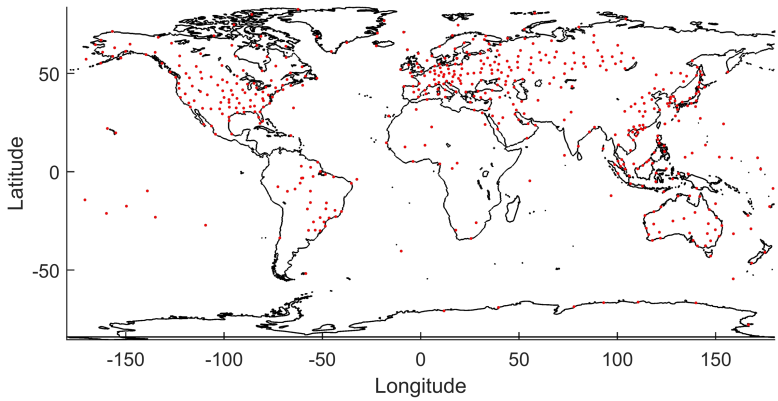


Fig. 2. Universal rawinsonde observation program data set for the year 2015: 12.5×10^6 observations come from 684 radiosonde stations, 25 observations per radio sounding and 720 time steps for twice-a-day measurement

The 'metro line map' is chosen for various reasons as part of dissemination (Institute for Ageing, 2014). The granularity of stations neatly matches that of the smallest area estimates of healthy life expectancy, with sufficient precision that, reporting figures rounded to one decimal place, one can dispense with sampling uncertainty. Yet the reader is not distracted by the geographical detail of most maps which can overwhelm sparse statistical estimates. This exemplar mapping is multivariate in nature; in contrast a univariate presentation tends to be judgemental in character—more than one dimension can present a pattern which leaves some interpretation to the audience. The second dimension is social deprivation, but not marked on the map—local commissioners are acutely aware; indeed they know it better than can be represented by areal measures of deprivation based on national indices.

There are other features which engage the imagination of the reader and serve to persuade. Tacit referrants are present in any statistic about age, even though life expectancy is not literally age: the poorest can expect to retire into ill health, as the figure for their area is below the state retirement age; in contrast the wealthiest may expect 10 years of healthy retirement. Within the range there is a steady gradient, not just a health penalty for severe deprivation. Finally, there is the rhetoric of the mapping, which serves as a visual metaphor. This highlights the proximity and ease of travelling between places with quite different outcomes; it forces the issue of the local inequity for which the audience is responsible. Commissioners retain agency over what to do, but the established inference is not presented with the objectivity of conventional statistical practice (Gelman and Hennig, 2017) and typical principles for visualization.

Jackie Carter (*University of Manchester*)
John Berger (art critic, photographer and painter) argued that we cannot assume that everyone sees a graphical object in the same way that we do, or as we would intend them to see it (Berger, 1972). People perceive maps and graphics differently, and we need to be aware of this when we assert claims about our visual messages. Berger also contended that photographs need language and a narrative to make sense. Data journalism is all about making sense of data and statistics through storytelling, using words and images. I would argue that, if we do not teach people the basic principles of graphics, the grammar of graphics or graphical literacy, we can hardly expect people to understand some of the more advanced visualizations that the papers in this session discuss.

My response to the papers is developed on two observations.

First, attention must be paid to the underlying principles of visual communication—both in developing these for research and associated communication purposes, but critically in testing different audiences' perceptions, to measure their understanding and interpretation of visualizations.

Second, ignore the narrative at your peril. I argue that to use visualization effectively a fundamental question is to address the matter of who exactly the audience is. Design for your primary audience (e.g. the Newcastle decision makers who would understand the geographic references to the example shown in the presentation) but be mindful that there will always be secondary audiences.

Finally—do not assume that the visualization—especially an infographic—can be readily understood on its own; a narrative is required to make sense of the figures and statistics. Berger's *Ways of Seeing* lends much to our understanding of the perception and interpretation of data visualizations.

Alasdair Noble (*AgResearch, Lincoln*)
Here we have seen some wonderful examples of graphics, to use in our work, but I have a question. I work in an agricultural research institute and help scientists to design experiments. A few months later some data arrive on my desk and I carry out some analysis, draw some graphs, that I believe illustrate the important features of the data, and send a report back to the scientist. In most cases we discuss the analysis and graphs and a report or paper is written. In a few cases a draft paper arrives on my desk a further few months later and it contains some bar graphs with standard error bars on each bar with a few *a*s, *b*s and *c*s etc. scattered around like fairy dust. The scientist correctly claims that that is what is typical in the intended journal and they say that my graph was too confusing.

My question is 'How do we persuade scientists to use some of these better graphical representations when they see a bar graph of the means as the gold standard?'.

Tanja Krone (*Netherlands Organisation for Applied Scientific Research, Zeist*)
I work as a statistician for a company where I must communicate daily my results and the related intricacies to colleagues and clients. A huge tool hereby is visualization and graphs, especially when the audience is not statistically knowledgeable. However, during my studies this was given very little attention. The graphs that we did use were created automatically by the programs used. There was no encouragement to be creative with graphs or visualization, or even an inkling that this is possible.

During my research Master's and doctoral degrees, both focused on statistics for psychologists, this was never a topic. Nobody enticed me to look further or to use new methods. When I asked my supervisor, halfway through my doctoral research, how to use `ggplot`, which is a well-known graphics package in R, I was told that it was too complicated and I would not need to use it anyway. I currently use `ggplot` almost daily.

If we want to revolutionize our use of graphics and visualizations, and if we want to find concise and clear ways to present data and results to the non-expert public, we should start at education. Teach students what is possible, teach them the basics and entice them to explore the world of data visualization themselves. Do not wait until they need it in the real world.

Zhou Fang (*Biomathematics and Statistics Scotland, Edinburgh*)
Castruccio and his colleagues make an interesting contribution. The use of virtual reality techniques adds another dimension to representation of data, taking advantage of what is enabled by recent advances in technology.

In a sense, the technique described is part of an existing literature, not just in terms of statistical visualization but also the use of aesthetic elements to indicate depth in art. There is a long history (Brooks, 2017) of various techniques, including

(a) linear perspective,
(b) atmospheric perspective,
(c) overlapping forms and
(d) motion parallax

(see for example Hixon (2018)). Although the present work uses stereoscopic virtual reality to achieve a depth representation and previous work (e.g. packages in R) have focused on the use of linear perspective to create the illusion of depth, these implementations may be viewed as part of a wider tool set that may be employed to produce representations that enhance understanding or widen public engagement. Some of these tools may produce a similar effect without requiring as great a technological component as virtual reality.

A *caveat* may also be suggested. Others (e.g. Gelman and Price (1999)) have previously suggested that map representations of estimated values may be misleading. It is possible that stereoscopic and other three-dimensional representations add to this—aspects of the visualization that are engaging and attractive may distract viewers from features of what is represented. An additional problem might arise if the depth value is an important part of visualization: artefacts in the representation may send contradictory messages to the viewer by signalling illusionary differences in depth. For these reasons, such techniques should be approached with care.

Peter Green (*University of Bristol and University of Technology, Sydney*)
The audience: most of the oral discussions at the meeting mentioned the audience for the novel graphical displays offered in these papers, whereas the presentations of the papers themselves rather neglected this issue. I want to re-emphasize it.

The sole purpose of any graphic is to convey information and to do so honestly, and the sole measure of a graphic's effectiveness is the success and integrity with which it is conveyed. The success depends on both the graphic and the audience (and its expectations and experience), and not necessarily very much on the graphic's beauty.

We can think about a whole spectrum of audiences—the creator alone (autographicism is perfectly healthy); the creator's team, or colleagues or students; the organization; the customer; the policy maker; the public. How often is no thought put into the intended audience, or a graphic designed with one audience in mind, and then shown to a different one? Even with rudimentary coding skills, it is just too easy to create a novel graphic, and these can be routinely reproduced in reports and articles with little objection. We would of course think it ridiculous to attempt to transmit written information in a made-up language!

And now, what about the honesty?: there are checks and balances on the integrity of written text; the sceptical reader can be on the lookout for hyperbole, can expect claims to be justified and referenced. Incompetent, careless, misleading or malicious communication can be called out. The near universality of written English as a medium for conveying scientific and technical information, and a limited and outdatedly formal version of English at that, does not make for great literature, but it does at least serve to make communication effective and help to make communicators accountable.

But graphical communication comes with more degrees of freedom, fewer rules and less history, and we need to be able to guard against accidental or purposeful misinformation. Tonight's authors are surely innocent of intending this, but whether suppressing inconvenient data, downplaying a relationship or exaggerating the performance of methodology, the lazy or unscrupulous can get away with anything in a novel graphic. Do we need more rules, less creativity or just trust to more stringent and critical refereeing?

The following contributions were received in writing after the meeting.

Anthony C. Atkinson (*London School of Economics and Political Science*) **and Aldo Corbellini and Gianluca Morelli** (*Università di Parma*)
We are sorry that we were not able to attend what looks to have been a most interesting meeting; some of the graphics in the papers are very elegant.

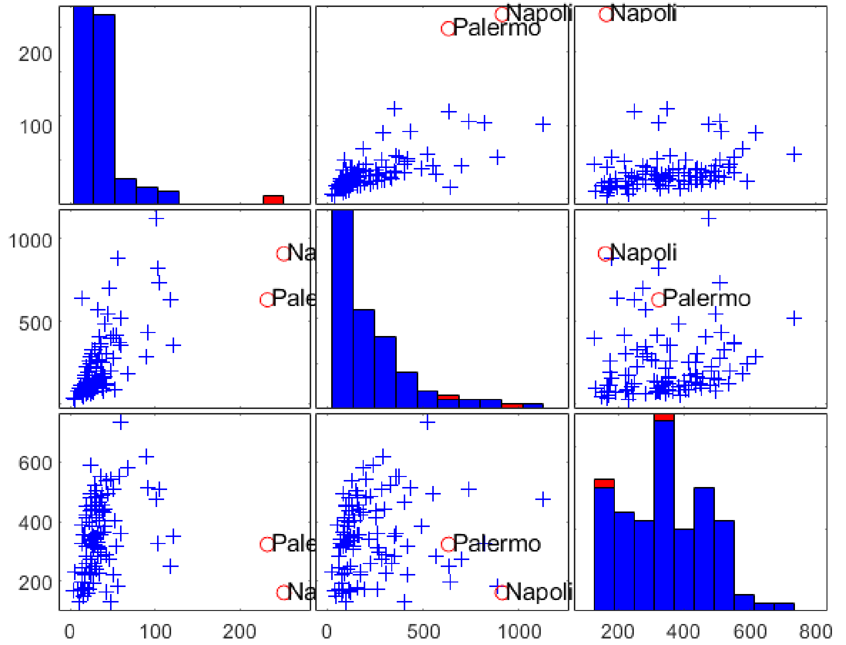
In his Section 5 Adrian Bowman mentions brushing and refers to Becker and Cleveland (1987). We report on how useful we have found brushing, not only of scatter plots but also of linked plots, in building and checking statistical models, the subject of the contribution by Gabry and his colleagues.

Our initial data are on the quality of life in 103 Italian provinces. We look at three variables: robbery, car theft and burglary. The scatter plot is in Fig. 3(a). In all cases large is worse.

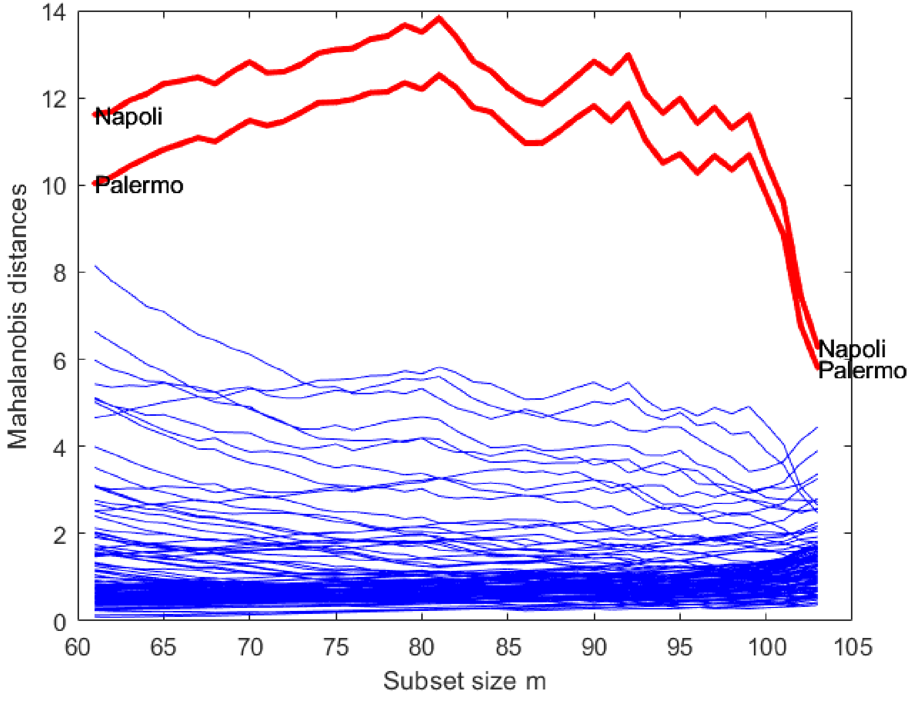
We analyse these multivariate data by using a robust technique, the forward search (Atkinson *et al*., 2010), which orders the data by estimation from subsets, the size m of which increases during the search. Fig. 3(b) shows a 'forward' plot of all 103 scaled Mahalanobis distances (Atkinson *et al*., 2004). Outliers and identifiable subsets enter towards the end of the search (Atkinson and Riani, 2004). The clearest feature in the plot is the two large distances. Brushing shows that these two trajectories belong to Naples and Palermo; they are most remote in robbery. We can brush in either direction or starting from other forward plots (Riani *et al*., 2012).

There is also a band of less large, but still appreciable, distances shown brushed in Fig. 4. These trajectories come from the large cities Rome, Turin and Milan, with Rome most outlying in car theft. The other smaller cities are all to the south of Rome.

Our second example shows the effect of linking plots which include tests of a fitted model. The data (Atkinson and Riani, 2006) are 509 observations on the behaviour of customers at a supermarket chain in northern Italy with response the amount, in euros, spent at the shop over 6 months. We use the Box–Cox transformation with the 'fan plot' (Riani and Atkinson, 2000) in which the approximate score test for the value of the transformation parameter (Atkinson, 1973) is calculated for the subsets produced by the forward search. Fig. 5(a) shows that, at the end of the search, the statistics for $\lambda_0 = \frac{1}{3}$ and nearby values all change rapidly. Brushing these 15 units for $\lambda_0 = 0.4$ and linking to the scatter plot of the data shows that there is a cluster of customers who are spending less than would be expected for their value of x_1, the number of visits to the shop.



(a)



(b)

Fig. 3. Quality-of-life data: (a) scatter plot of data (○, brushed units; +, unbrushed units); (b) forward plot of scaled Mahalanobis distances, brushing of the two most extreme observations

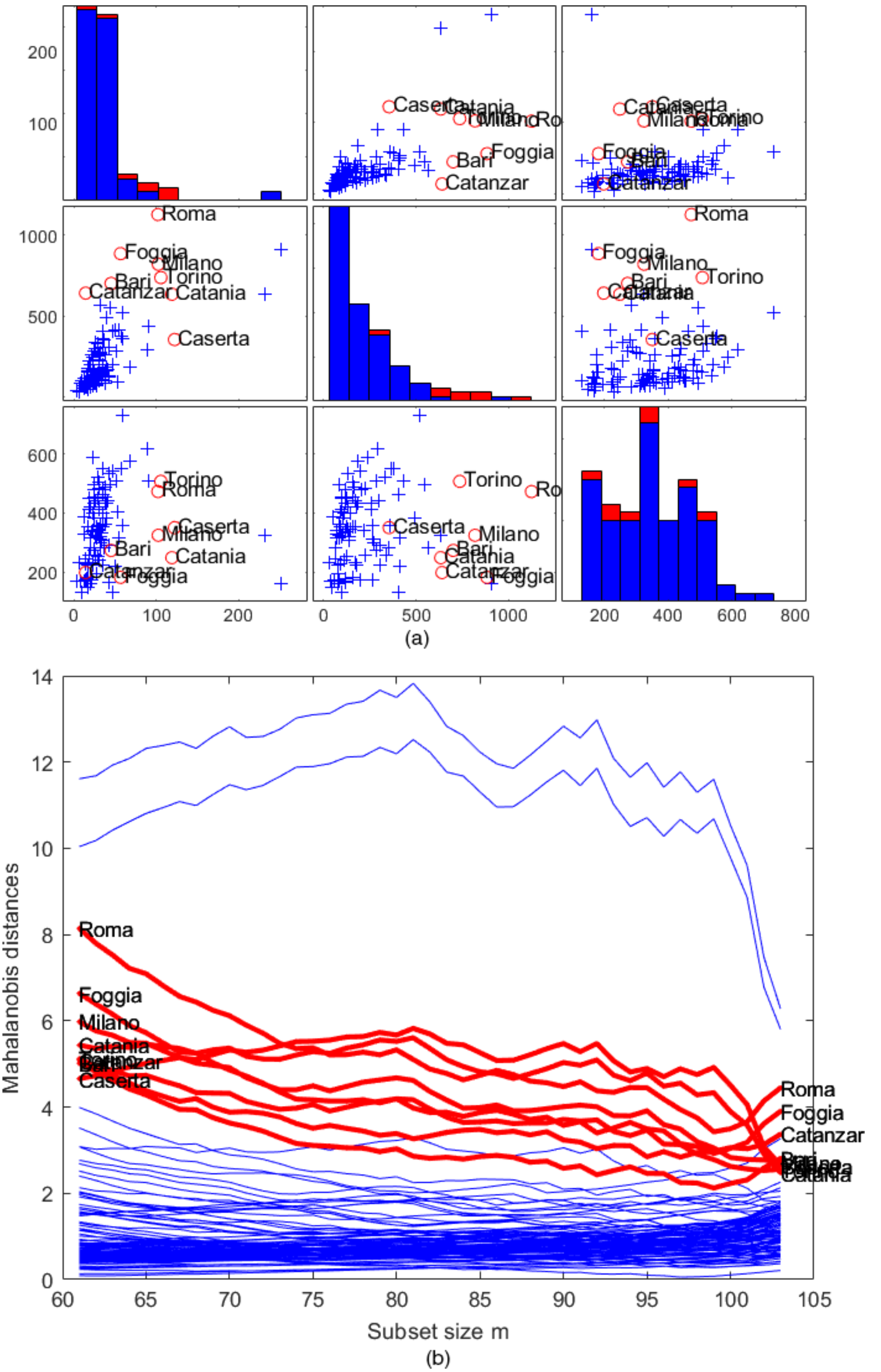
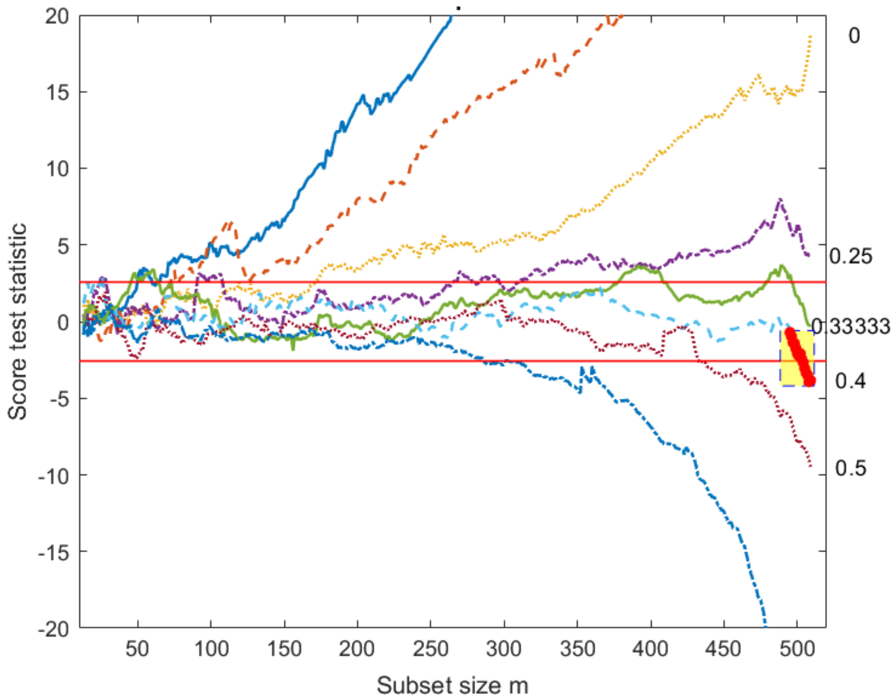
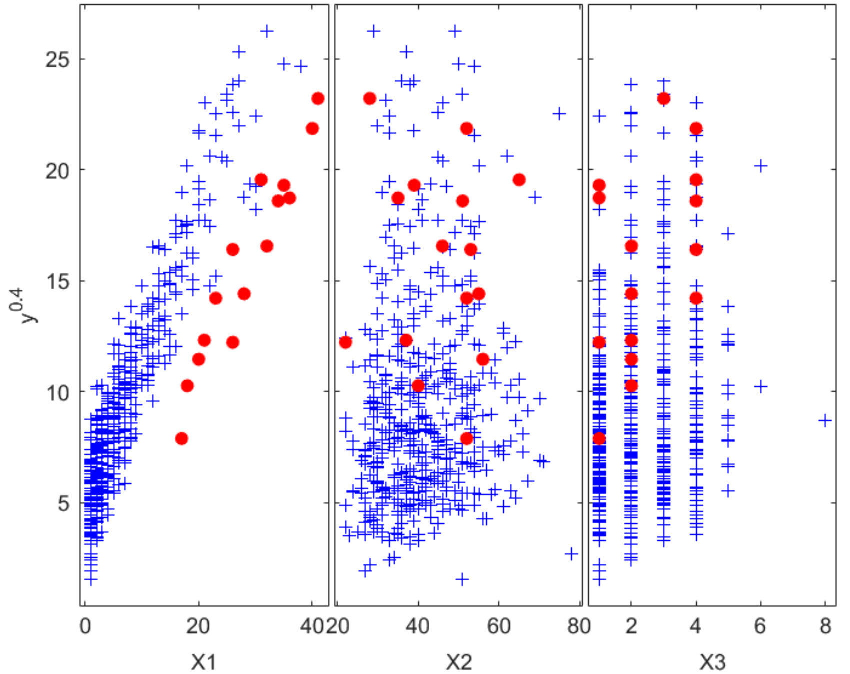


Fig. 4. Quality-of-life data: (a) scatter plot of data (○, brushed units; +, unbrushed units); (b) forward plot of scaled Mahalanobis distances, brushing of the next most extreme observations



(a)



(b)

Fig. 5. Supermarket data: (a) far plot; (b) scatter plot of sales$^{0.4}$ against number of visits, age and number of people in the family (●, brushed units; +, unbrushed units)

Noel Cressie (*University of Wollongong*)
Visuanimation is about understanding spatiotemporal data through visualization and animation (Genton *et al.*, 2015). Using what was new technology at the time, video recordings of animated visualizations of spatial and multivariate views of spatiotemporal data were developed at Iowa State University (e.g. Cook *et al.* (1996)). An immersive environment at Iowa State University called a 'cave' allowed us to stand in a room and to watch data points come from nowhere, spin around us, and even pass through us. Today's technology allows us to carry a cave in our pocket, but what extra scientific value is in the visuanimation of Castruccio and his colleagues? They answer this question by enabling data analysts to interact with their spatiotemporal data in this environment.

Looking at 'views' of the data in an animation, as time advances, is valuable for a first impression, but beware of temporal dependence! For illustration, let $\{X(\mathbf{s}_0, t) : t = 1, \ldots, T\}$ denote a slice of data at a fixed location $\mathbf{s}_0$ in the spatial region of interest. Assuming temporal stationarity, we look to estimate the central tendency $\mu(\mathbf{s}_0)$ and the native variability $\sigma^2(\mathbf{s}_0)$ of the time series at $\mathbf{s}_0$. If $\bar{X}(\mathbf{s}_0)$ is the sample mean and $S^2(\mathbf{s}_0)$ is the sample variance, then $E\{\bar{X}(\mathbf{s}_0, t)\} = \mu(\mathbf{s}_0)$, but

$$E\{S^2(\mathbf{s}_0)\} = \sigma^2(\mathbf{s}_0)\left\{1 - 2\sum_{1 \leqslant t < u \leqslant T} \frac{\rho(t-u; \mathbf{s}_0)}{T(T-1)}\right\},$$

where $\rho(h; \mathbf{s}_0) \equiv \mathrm{corr}\{X(\mathbf{s}_0, t), X(\mathbf{s}_0, t+h)\}$. Moreover, $\mathrm{var}\{\bar{X}(\mathbf{s}_0)\} \neq E\{S^2(\mathbf{s}_0)\}/T$, and so the usual basic inferences are no longer valid. The results above are given for one spatial location, but they generalize to whole maps changing through time.

The variability in the data is diminished by the dependence, sometimes drastically so. Zhang *et al.* (2017) show how satellite data collected during a period of several minutes, with a sample size of almost 3000, have an 'effective sample size' of just over 200. The message to neo-spelunkers is that, in the presence of dependence, perceived variability does not reflect the native variability.

Finally, the similarity measure SSIM lies between -1 and 1, but it achieves a value of approximately 1 when x is approximately y or approximately $-y$, which could make its interpretation challenging.

Michael Friendly (*York University, Toronto*)
A central goal of applied statistics can be simply stated: to tell a credible and insightful story about a real problem, amply documented with words, numbers and pictures. Words carry most of the burden, numbers can give evidence of credibility and graphs can provide visual overviews and details that contribute to insight in ways that words and numbers may fail.

The paper by Bowman represents an important contribution to this topic by considering how uncertainty can be conveyed as part of the story. He proposes the use of a density strip (Jackson, 2008) as a representation of distributions, representing variation (uncertainty) of data, model fits, parameter estimates, etc. An impressive feature of this paper is its range of application: a similar device can be applied to questions of group mean differences, regression or response surface models and spatiotemporal data.

But the crucial question is: how well does this device work to tell the story in an insightful way? A longer version of this discussion is available from `http://datavis.ca/papers/Friendly-Bowman-discussion.pdf`; the main points are these.

(a) Density strips have the advantage that they provide a background for uncertainty to which other graphic elements can be added in the foreground. However, they have the disadvantage that human perception of variation in density via shading is quite weak and depends fundamentally on the parameters that are used to define the density scale of shading. In most cases, we can more easily see variation in density with lines, points and other symbols, if care is taken in the graphic design.
(b) For contingency tables, fourfold plots (Friendly, 1994a) are better suited to the case of association in 2×2 tables, showing standardized frequencies by quarter-circles, with a confidence band for the odds ratio. Fig. 6 uses this for the Doll and Hill (1950) data. The visual evidence for the strength of association between smoking and lung cancer is direct. More generally, mosaic plots (Friendly, 1994b) show patterns of association more clearly in larger and multi-dimensional tables.
(c) The limitations of pure density displays appear most clearly in bivariate and multivariate problems, where the goal is to see both the overall relationships between variables and a reflection of uncertainty, both of the data and of the fitted model. For a collection of scatter plots, such as Bowman's Fig. 5(a), the uncertainty around the individual linear regressions of giving to the Church of

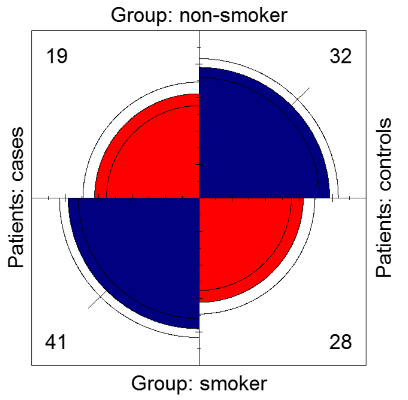


Fig. 6. Fourfold display for the Doll and Hill (1950) data: the 95% confidence intervals for the odds ratio do not overlap, showing a significant association

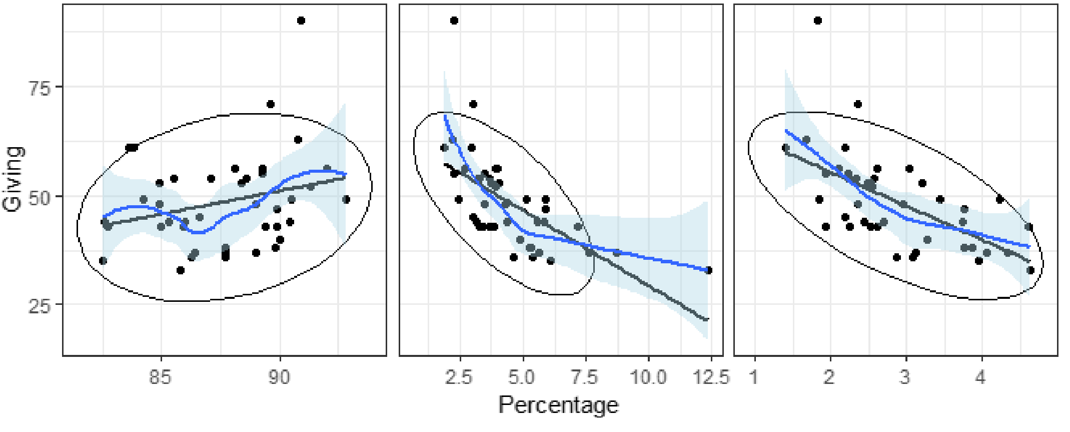


Fig. 7. Revised version of Bowman’s Fig. 5(a), showing uncertainty in the data and in a non-parametric smooth (——, linear regressions; ⬭, normal theory 95% confidence region; ——, non-parametric (LOESS) smooth; ——, pointwise 95% confidence envelope): (a) employ; (b) elect; (c) attend

England can be seen directly by adding data ellipses (Friendly *et al.*, 2013) to the scatter plots, as shown in Fig. 7.

Lasse Holmström (*University of Oulu*)
Since at least Tukey (1977), the statistics community has recognized the value of innovative visualization techniques in data analyses. However, despite many successful demonstrations of approaches based on virtual reality and other advanced technologies (e.g. Cru-Neira *et al.* (1992)), such techniques so far have failed to enter the mainstream and have therefore been of limited practical utility. Has the time finally come for a widespread adoption of these technologies in applied statistics? One would certainly hope so and the paper of Castruccio, Genton and Sun describes promising steps in this direction. A good way to promote wider adoption of such tools would be to make them a part of the education of statisticians and data scientists. This today is made easier by the emerging virtual reality and augmented reality capabilities of mobile devices and the availability of the relevant statistical applications that run on them.

I would additionally like to point out that statistical scale space methods also represent a family of techniques for exploratory data analysis that shares some goals with the techniques that are described in the papers discussed here. These include graphical summaries of inference and its uncertainty in a form that can be easily interpreted also by non-statisticians, as well as maps and animations that can be used to discover systematic features underlying the observed data (e.g. Holmström and Pasanen (2017) and Vuollo and Holmström (2018)).

Raphaël Huser (*King Abdullah University of Science and Technology, Thuwal*), **Miguel de Carvalho** (*University of Edinburgh*) **and Luigi Lombardo** (*King Abdullah University of Science and Technology, Thuwal*)

We congratulate Castruccio, Genton and Sun on their timely paper, which explores modern visualization tools for communicating statistical results and uncertainty more efficiently for spatiotemporal models defined on complex domains.

Although they chose to focus on classical space–time applications, the benefits of advanced visualization tools reach far beyond the standard Gaussian geostatistical setting. We highlight two different examples of risk assessment applications.

Statistics of extremes (Davison *et al.*, 2012, 2018; Davison and Huser, 2015) focuses on estimating low probability tail events and has obvious applications in Earth science and finance, where the notion of risk is often related to individual or simultaneous extreme events. Genton *et al.* (2015) used 'visuanimations' to explore dynamically the dependence characteristics of a (highly non-Gaussian) spatiotemporal model for precipitation extremes proposed by Huser and Davison (2014). In this context, interactive visualization tools could be very useful, especially when adding an extra 'tail' dimension, enabling people to explore dynamically the estimated joint tail behaviour and to visualize efficiently extreme spatiotemporal return levels for increasing return periods.

The most important object to estimate and visualize from point pattern data (Cressie, 1993), such as wildfire or landslide data, is their *intensity function*. Lombardo *et al.* (2018) used a log-Gaussian Cox process to model landslides and to create spatial predictive maps to identify regions at risk in Italy. The data and the fitted intensity function were visualized by using Google Earth™; see Fig. 8 and `https://vimeo.com/266344095`. The visualization of risk maps and how they evolve in time would highly benefit from immersive virtual reality tools, which more realistically render a three-dimensional perspective.

As explained by Castruccio and his colleagues and further motivated above, advanced three-dimensional, immersive, or portable virtual reality visualization techniques may be used to 'explore and assess the structure of the data and to improve resulting statistical models' dynamically and interactively. We fully concur, although we point out that using such modern facilities as the King Abdullah University of Science and Technology three-dimensional visualization cave and creating smartphone or virtual reality headset applications for scientific visualization often require a significant investment of time and the support from a specialized staff. Although this is undoubtedly worthwhile for the *communication of results* from statistical analyses, we think that, for *data exploration*, simpler approaches such as R Shiny applications,

Fig. 8. Visualization of landslide data and fitted log-Gaussian Cox process intensity function by using Google Earth™; see also https://vimeo.com/266344095

or popular geographic information system environments including Google Earth™, which are visually efficient, cheaper, accessible and significantly less time demanding to set up, are still useful.

Christopher Jackson (*University of Cambridge*)
I thank Professor Bowman for boldly going much further than I did in Jackson (2008)! Whereas I used density strips to show uncertainty about parameter estimates, mainly as Bayesian posterior distributions, he uses them in two new contexts.

The first is to illustrate estimates of the sampling density of observed data. In Jackson (2008) I cautioned against showing density estimates in place of plots of the raw data, since we typically want to examine data for interesting features such as outliers. So I was happy to see Figs 2 and 3, which use jittering to enable the data to be shown on top of the density strip, while retaining a compact one-dimensional form. Ideally we would want also to examine the data interactively—the author mentions `shiny`, and I recommend also the `plotly` R package (Sievert, 2018) for this purpose.

The second new application of density strips is to illustrate the sampling distributions of frequentist test statistics. With my Bayesian instincts, I find these more difficult to interpret than plots of posterior densities, but I welcome any way of communicating hypothesis testing which is more informative than *p*-values.

An interesting dilemma is whether exact numerical summaries, such as point and interval estimates, should be added to density strips. Shading deliberately obscures the exact densities being plotted, while conveying an impression of the extent of uncertainty. Previously I have included at least point estimates, but not always interval estimates, which might encourage drawing conclusions based on whether the interval includes a 'null' value. It is also important to note that such summaries are conditional on the exact data observed and any model used to analyse them. Thus they may not represent unmodelled measurement errors, selection biases or alternative plausible models. The density strip is also conditional on the model and data, but it should be sufficient to convey the uncertainty as long as the shading pattern does not look different under plausible alternative assumptions.

It is worth mentioning software. Density strips can be plotted in base R with the `denstrip` package (Jackson, 2008). In the `ggplot2` R package (Wickham, 2016), an existing estimate dens of the density at a set of points x can be illustrated as a strip using code such as

```
x <- seq(-1, 1, by=0.01)
dat <- data.frame(x=x, dens=dnorm(x, -0.5, 0.15))
ggplot(dat, aes(x=x, y=0)) + theme_bw() + ylim(-1, 1) +
  geom_tile(aes(fil=dens), height=0.2) +
  scale_fill_continuous(low="white", high="black")
```

However, as `ggplot2` is now user extensible, I would welcome the development of a new `geom_denstrip` function which estimates the density and creates the plot in one step, like the `denstrip` package does.

As an attractive alternative to density strips, the `ggridges` package (Wilke, 2018) implements 'ridge line plots' which use perspective and colouring to display a stack of several two-dimensional density estimates compactly.

Jorge Mateu (*University Jaume I, Castellón*)
I congratulate Castruccio, Genton and Sun on this interesting, timely and certainly attractive paper on the edge between computing and statistics, that touches the big field of data science and data analytics. How to visualize data effectively and to use them to drive decision making is probably a cornerstone in the field of data science that uses scientific methods and algorithms to extract knowledge and insights from data, both structured and unstructured. And undoubtedly statisticians who work with data indexed in space and time can much better explore and assess the structure of the data by visualizing them by using the current availability of modern technology. This paper reaches the point in a timely fashion. I would like to reinforce several aspects that are not touched on enough here, in my opinion.

One important point in visualization is that it should help in graphical-based statistical learning. If a graphic represents a snapshot of a moving phenomenon, by considering a movie, we can analyse the dynamics of such graphics and can better understand the statistical underlying model. Connected with this is the possibility of exploring the uncertainty through animation in several dimensions. Uncertainty is crucial in statistical reasoning, and any device (hardware and software) that can disentangle uncertainty in the data is a necessary and welcome contribution. And I am specifying several dimensions because visualizing spatiotemporal data with a geometric mark, so sort of four dimensions, is extremely useful for a refined statistical modelling exercise. This would be so for spatiotemporal marked point patterns, or functional spatial data.

Another point that is not discussed in the paper is visualization of functional spatial data. This would be the case of trajectory-based spatial data within a broader context of big data. Trajectory data are becoming increasingly more available and there is a need for modern devices for visualizing such enormous amounts of data which are available in a realtime fashion. Point change analysis or relative changes with respect to various quantities of interest can be much finer discussed under nice animations.

I end with a more general comment. Entering the era of big data analytics and data science, computer scientists, in particular visualization experts, are needed in a working group where the emphasis is statistical modelling and reasoning. And, because of this, I congratulate the authors for their collaboration with such a team.

R. Allan Reese (*Dorchester*)
l read Professor Bowman's preprint and then corresponded with him. The stated aims of

> 'making statistical concepts accessible ... to those who are unfamiliar with statistical methods ... and more informal evaluation of the statistical evidence ... greatly aided by graphical methods which clearly communicate the uncertainties ... particularly geared towards communicating with those who do not have technical statistical knowledge'

suggest that the examples are presentations. But, within the graphical interpretation of data framework of *analytic* and *presentation* graphs (Reese, 2017), these examples fail as presentations; they are better described as suggestions for stages in analysis and diagnostics for a researcher. It is suggested that a non-statistician applying statistical methods will better understand variation by seeing where cut points fall in the shading.

Analytic graphs generally rely on default layouts and labelling. In the preprint, Fig. 2(a) lacked group labels and Fig. 3 lacked units or indication that it was a log-scale; these points have been improved since my comments, reflecting the need to read and edit graphs as rigorously as text.

The original Fig. 2(a) showed three unlabelled bands of observations. The text mentions measurements at 3 and 6 months plus a control group, that I originally thought were the bands. It actually shows the same values (each patient's change over time) three times with added context. I did not find the grey shading helpful and was unclear what was meant by 'the distribution'. Boxplots seem to me a better option.

Representing the probability density function by shading raises problems of reproduction (device capabilities) and perception (individual vision), before considering understanding and interpretation. The vertical jitter that is used on these univariate displays may confuse; it is better to stack points within value classes, representing the probability density function with a histogram.

Plotting a *reduction* in facial asymmetry (presumably desirable but a negative move) as a *positive* value may also confuse viewers; the axis should be labelled for the direction of improvement. A similar example of non-intuitive orientation is discussed in a forthcoming graphical interpretation of data column (Reese, 2019).

The regression example smacks of Mathsworld, the student mindset that elevates 'doing the right arithmetic' over understanding the context and semantics (Reese, 2019). The immediate impression is one outlier at £90 and maybe a second at £70, which I would identify during analysis (and label in a presentation) when doing regression; that is why we do diagnostic plots.

I am dubious about animations; could stepping through a sequence of 'freeze-frames' offer more time for comprehension as well as being less computationally intensive?

I applaud Professor Bowman for bringing these ideas to debate, but I do not see the evidence that they are a move in the right direction.

Priyantha Wijayatunga (*Umeå University*)
For Bowman's interesting paper I would like to contribute to its discussion by concentrating on visualizing both conditional and marginal associations in contingency tables by using simple diagrams. One such diagram is used to visualize Simpson's paradox in Wijayatunga (2014). In contingency tables of data, it is important to look at how observed associations change when new conditioning variables are introduced. Let X, Y and Z be binary random variables such that, for example, $X=x$ denotes being admitted to graduate school, $Y=y$ denotes the gender status of female and $Z=z$ denotes applying for the subject of mathematics, in the context of students applying for graduate school admission. Here, for value α its

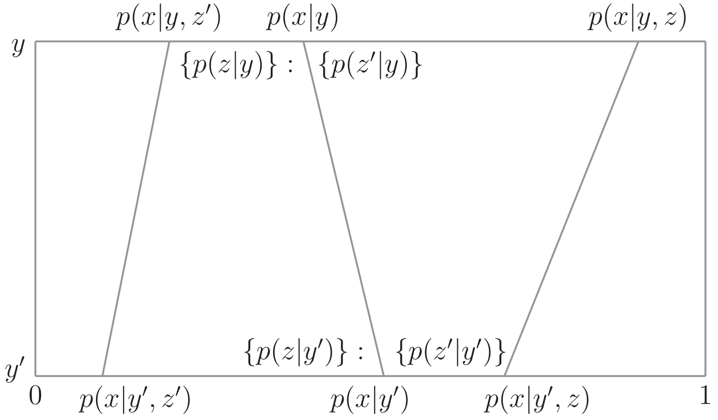


Fig. 9. Conditional probability values are marked on two length sides of the rectangle: for a probability value p, the notation $\{p\}:\{1-p\}$ means that the lengths of two line segments on which $\{p\}$ and $\{1-p\}$ appear are according to the ratio $p:1-p$

complement is denoted by α'. In Fig. 9, the length of the rectangle is a unit; thus it can be used to mark probability values and the height of it represents two values of Y, as shown. The diagram shows $X-Y$ and $Y-Z$ marginal associations, and $X-Y$ conditional association given each value of Z. Note that the slope of the line connecting probability values $p(x|y)$ and $p(x|y')$ shows the first and the line segment connecting $p(x|y,z')$ and $p(x|y,z)$ is dissected by $p(x|y)$ according to the ratio $p(z|y):p(z'|y)$ and similarly for the other line segment; thus it shows $Y-Z$ marginal association, and so on. Other associations can be shown with similar diagrams. According to the data (which are not explicitly given here), it shows that there is a gender bias towards males for the overall admission to the graduate school, but the opposite is true if it is looked at subjectwise (mathematics or not), i.e. then females are favoured. This is an instance of Simpson's paradox. This diagram can be used not only to explain the paradox to students of statistics but also to visualize the statistical associations between three binary variables. And in the case of multinary variables the diagram can be extended appropriately.

Andrew Zammit-Mangion (*University of Wollongong*)
Effective visualization is often an afterthought in spatiotemporal statistics, and the paper by Castruccio, Genton and Sun endeavours to bring the statistical community up to speed with the state of the art. They have focused on field visual comparison, but an advantage of the visuanimation and a virtual reality environment is that they can be used to explore validation statistics such as those given in Bradley and Haslett (1992), Carroll and Cressie (1996) and Kang *et al.* (2009). Some of these statistics could be time dependent (Wikle *et al.*, 2019). There is, however, an obvious tension between what is technologically achievable and what is practical and effective. First, some spatiotemporal visual summaries that can be viewed in two dimensions, like Hovmöller diagrams, can convey spatiotemporal trends that may not be apparent in a three-dimensional virtual reality environment. Indeed, it is not clear to me whether a three-dimensional immersive experience would give me a better understanding of the data than a collection of summary visualizations that show me different angles of the data and the predictions. Second, the hardware or software requirement is a notable stumbling-block that acts as a barrier to entry (largely because the benefits of the effort involved are not fully apparent). The technology requirement might also be a hindrance to certain groups of readers; for example, it is possible to view the relatively simple visuanimations on only some portable document format readers that are not available on all operating systems.

In my view, the next stage in spatiotemporal visualization is not necessarily the use of three-dimensional immersive environments, but the availability of an interactive on-line platform whereby the reader or data analyst can visualize different aspects of spatiotemporal predictions or spatiotemporal data (in particular, different spatial, temporal and spatiotemporal summary visualizations) with simple mouse clicks or finger presses on a standard desktop computer or tablet. The use of a Web portal interface means that hardware requirements and software requirements are kept to a minimum. Technologies that allow this are now widespread and stable (e.g. `D3.js` (JavaScript) and `ggvis` in conjunction with Shiny (R)) and such versatility would also provide an 'immersive' data experience, although of a different kind (see, for example, the interactive visualizations in Lamigueiro (2018)).

The **authors** replied later, in writing, as follows.

Stefano Castruccio, Marc G. Genton and Ying Sun

We thank all the contributors for both their oral contributions during the discussion meeting and their written comments after the meeting.

Fang, Huser, de Carvalho and Lombardo, and Zammit-Mangion have all raised the issue of hardware and software requirement for the three-dimensional virtual reality (VR) environment that is used in our work. Such requirements could effectively act as a barrier to entry for developing apps and more generally for any type of interactive visualization. Our work is intended to be a showcase of a long-term collaboration with visualization experts and is clearly at the high end in terms of hardware and software requirements for final product dissemination and communication. Alternative, more accessible technologies for intermediate visualization products are available. Rshiny is the most popular of such products, at least among the statistical community. Many contemporary software programs are indeed equipped with alternative solutions for depth perception, as pointed out by Fang, that would well serve the purpose of improving a user's understanding of three-dimensional spatial objects.

Closely related to this hardware–software topic is the question of the audience whom we are addressing in our work. Diggle and Green reiterated that the perception of visuals depends on the stakeholders with whom we interface. Our proposed set of visualization tools could be used not just by statisticians, but also by a more general audience. The educational value of engaging with multiple audiences is, in our opinion, among the greatest strengths of interactive, portable visualization tools. As we stated in the meeting, VR visualization would improve the narrative of a presentation of results to any stakeholders but, perhaps more importantly, it could be helpful to bridge the gap across generations and to address young students better, with the implicit hope of engaging new generations and to teach them the value of our profession through a new language.

On a more quantitative note, Cressie has pointed out that the interpretation of the structural similarity index SSIM (Wang *et al.*, 2004) is not simple and straightforward because 1 is not necessarily an indication of a perfect match. Fassò and Finazzi have also proposed an alternative asymmetric index that accounts for the different roles of original data and fitted data. As is apparent from the contributions, SSIM as presented does not adequately fit the need of the spatiotemporal community as a diagnostic metric. We expect to see and to encourage new developments in this direction from the community, but we believe that, however limited in its scope, SSIM proposes a new paradigm based on a very simple yet apparently underused principle: spatiotemporal models generate images and, as such, they could be evaluated with large-scale diagnostics. This would allow our community a bridge to the wealth of literature that has been produced in recent years by the signal and image processing communities, as well as the machine learning and pattern recognition communities, should the main goal of the analysis be understanding large-scale variability.

Mateu observed that visualization of functional spatial data was not discussed in the paper. We agree that this is a very important area for which we indeed have developed a functional boxplot (Sun and Genton, 2011) based on a very fast ranking of curves (Sun *et al.*, 2012) with respect to their depths in the data. As an illustration of this powerful visualization tool, we consider the simulation of a realization from a Gaussian process with mean 0, unit variance and isotropic exponential variogram $2\gamma(h) = 1 - \exp(-h/\theta)$ at 400 locations in the unit square. Here h denotes the distance between two locations and we set $\theta = 0.25$, which corresponds to a moderately strong level of spatial dependence. We compare the estimation of θ from the simulated data based on ordinary least squares and weighted least squares fitting, as well as by Gaussian maximum likelihood estimation. We repeat this experiment 1000 times. Hence, for each estimation method, we have 1000 estimates of θ that lead to 1000 variogram curves. Those are summarized and visualized in Fig. 10 with functional boxplots. This toy example shows that, although weighted least squares reduces the variability of the estimated variograms a little compared with ordinary least squares, the maximum likelihood estimation method has much less variability. Other experiments with more complex scenarios can be found in Yan and Genton (2018). The functional boxplot can be used for outlier detection. Similarly to the classical boxplot, the empirical rule is to inflate the central region (the envelope of the 50% deepest curves in the sample) by a prespecified factor to identify the fences. In the example in Fig. 10, we set the factor to a very large value without performing outlier detection for ease of visualization. The default value of the factor is 1.5 but adjustable by users. For example, if the curves are not replicates but are dependent—i.e., for instance, if the curves are observed at different spatial locations as mentioned by Mateu—then we can take this

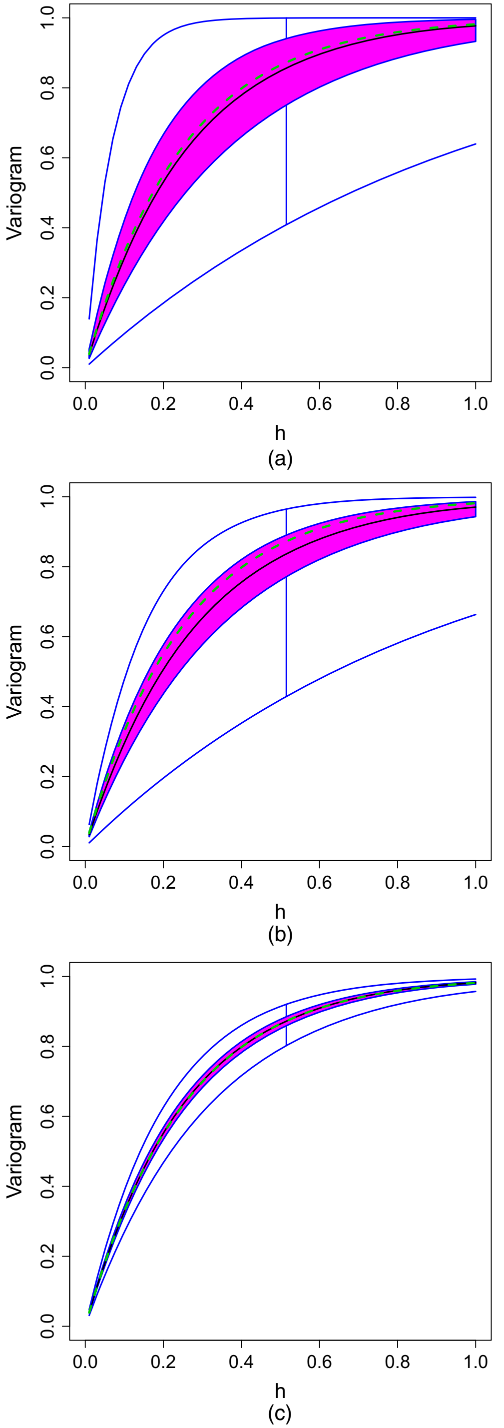


Fig. 10. Functional boxplots of 1000 estimated variogram functions by (a) ordinary least squares, (b) weighted least squares and (c) Gaussian maximum likelihood estimation: – – –, true underlying variogram ———, median; , envelope of the 50% deepest curves in the sample

spatiotemporal correlation into account and adjust the factor in the procedure to detect potential outlying curves (Sun and Genton, 2012). Extensions to multivariate functional curves can be found in Dai and Genton (2018a, b). The case of images that are indexed by spatial locations and surfaces leads to surface boxplots (Genton *et al.*, 2014). Further developments along these lines have already formed an active research area.

Cressie has advocated the use of videos in electronic papers, whereas Holmström has expressed a general hope that the works that were presented in the meeting would finally prompt the statistical community to a more widespread adoption of advanced visualization techniques. This is a point that all of us are particularly keen on reiterating. There is a pressing need for the statistical community to leverage the possibilities of electronic documents fully, given that the vast majority of scholars now rely on them instead of printed journals. To our knowledge, *Stat* is the first statistics journal that actively promoted the use of movies in the electronic document (visuanimation; Genton *et al.* (2015)). It is our hope that our discussion will be an additional prompt for embedding media in papers, possibly in an interactive fashion.

An important point that emerged in the discussion and also raised by Fassò and Finazzi is the possibility of using VR environments and apps for variables in three dimensions as well as in time. Although we have already proposed this solution with three-dimensional temperature data (`vimeo.com/109146573`; see full details for the model in Castruccio and Genton (2016, 2018)), it is apparent that a variable with only a limited number of non-zero values in space would present a more compelling case for the use of VR devices. We have performed a similar visualization exercise with data from Mount Pinatubo's 1991 eruption (`vimeo.com/103123075`) in which the sulphate aerosol concentration was plotted as a function of time. It is possible to observe and study how the plume of dust that was expelled from the volcano became dispersed across the equatorial belt throughout the year. The visual result is clearly more appealing than temperature data, although the application at present lacks a statistical model for comparisons.

Finally, a general concern expressed by many colleagues in the audience during the meeting, as well as by Carter and Jones, is how we can, as a community, promote the more widespread use of visualization in statistics and data science, both in the academic world and in industry. This is perhaps best summarized by Carter's questions at the end of her contribution; how can we persuade scientists to use better graphics if they are used to very simple and standard tools? The answer to this question is complicated, but the first step is clear: education. Statistics and data science *curricula* should give space to visualization and, more broadly, to communication. Given time, today's students will become tomorrow's leaders, and proper training will allow them to see the value of advanced visualization tools and to update practices in industry and academia. Education is a long-term solution, but with long-lasting positive effects for society.

J. Gabry, D. Simpson, A. Vehtari, M. Betancourt and A. Gelman

We thank everyone who has participated in this discussion.

In his proposal of the vote of thanks Diggle poses the following question in response to our paper: in what sense can inference be deemed Bayesian if

> '... you massage your prior until it generates realizations that are concentrated (to a greater or lesser extent) around the data ...' ?

This is a very reasonable question, but we were *not* suggesting that the prior should be manipulated until realizations from the prior data-generating process are concentrated around the data. Since most readers of our paper will not be experts in particulate matter air pollution, to make the comparison more accessible we compared the prior predictive simulations of $PM_{2.5}$ concentration with the actual measurements we have. But, when a researcher is actually conducting an analysis in their area of expertise, they should have enough familiarity with the subject matter to look at prior predictive simulations on their own, without needing to make direct comparisons with the data that will be used for model fitting. For example, a researcher studying $PM_{2.5}$ levels would know that the simulated data represent concentrations that would be fatal to life on Earth. So our point is really that a reasonable prior is a prior that yields a reasonable prior data-generating process, not that the researcher should tailor the prior to suit the particular observations in hand.

Baillie and Vandemeulebroecke say that our proposed workflow resonates with their experience as statisticians working on drug development. We are very happy to receive this confirmation that the workflow is being applied successfully in industry. Baillie and Vandemeulebroecke also point out one limitation of our paper: there are no recommendations for how to use visualization in the final step of communicating inferences to a more general audience. We used our limited space to focus on visualizations that are intended to help people doing statistical modelling to develop better models, but we strongly agree that translating the inferences from those models effectively for other audiences is a vital part of a statistician's job. We are also grateful that Carter and other commenters have brought attention to the challenges of communicating to different audiences by using visualizations. This was not something that we intended to address in our paper, but it is an important topic that deserves greater attention from the statistics community.

Adrian W. Bowman

One of the challenges of presenting a discussion paper on any subject is the breadth of the connections which can be made with other important areas. That is most certainly true of visualization, which is not only relevant to all areas of our own subject but which also plays a central role in communication, both within the statistical profession and with the wider public. The very limited time and space that were available to each presenter of the individual papers has required a focus on very particular aspects of the topic, so it has been very helpful for the agenda to be expanded considerably by the discussants. I am very grateful for all the contributions which have been made.

Several contributors rightly mentioned the central importance of the audience and identified the very broad continuum of stakeholders, and the correspondingly wide variation in statistical expertise, that applies in any graphical communication exercise. In leading this aspect of the discussion, Diggle provided timely reminders of the importance of the basics and the dangers of being beguiled by the medium rather than the message. He also gave several excellent examples of how the form of presentation can be powerfully adapted to the audience. Carter drew attention to the fundamental issue of perception and the considerable individual variation in ‘ways of seeing’ that will be at work when graphics are viewed, and the crucial importance of the enveloping narrative. In endorsing the importance of these issues, Green drew powerful comparisons with formalisms of written language and raised the issues of how we can ensure that information is communicated honestly.

I find the analogy with written and spoken language helpful and stimulating. As a profession, we often must contend with the uninformed view that ‘you can prove anything with statistics’. My own standard reply is to point out that one can say anything in English but that is not the fault of the language. One of the principal contributions of our subject is in providing a powerful and coherent language for talking about uncertainty and inference. The aim of my own contribution was to try to link that inferential language to graphical communication in as direct a manner as possible, so that the two languages, inferential and graphical, are as consistent as possible through appropriate displays of uncertainty.

A language requires words as basic building blocks and grammar as a means of forming words into coherent sentences. When there is agreement on how words and grammar should be used, then there is a platform for effective communication. The same principle applies to graphics so, as several contributors have commented, we need to make sure that the basic building blocks are well understood. From this perspective, effects such as shading and animation can be viewed simply as attempts to extend our vocabulary. However, any piece of text other than the very simplest aims to communicate a more complex meaning than that associated with individual words. The overall narrative is indeed crucial, as Carter emphasized. I would much preferred to have developed the stories around each of the examples, as all of those stories exist, but I regret that time did not permit that. To return to the issue of effective and honest communication, miscommunication—indeed deliberate misdirection—can occur at the narrative level even when the individual words are correct. This often arises not from what is said but from what is not said. From that perspective, I am not sure that graphics are more prone to abuse than text. In both cases we must depend on the honesty and integrity of the communicator.

Baillie and Vandemeulebroecke, Carter and others raised the issue of how we know we have succeeded in graphical communication. This is a very important and interesting topic which raises again the issue of multiple ‘ways of seeing’. I have myself had the sobering experience of realizing that a picture whose message to me seemed obvious was to others opaque. My own forays into the literature on graphical perception have made it clear that there is a body of research there which is certainly worth further investigation. As statisticians we should be ready to listen to the evidence on what works and what does not.

In terms of shading, Baillie and Vandemeulebroecke and Friendly in particular have raised questions about effectiveness in communication. I would like to emphasize that I agree entirely that this is not a means of showing details of distributional shape—other forms of display are much more effective in that. However, there are many occasions where a much broader interpretation is appropriate, indicating simply where we should expect to see values of interest and where we should not. In those situations I believe that the ‘smudge’ of a density strip can be very useful, precisely by avoiding the detail that other displays sometimes involve. However, I repeat that the aim of advocating this approach is to extend our graphical vocabulary, not to offer it as a preferred means of communication in all circumstances.

In the case of animation, Baillie and Vandemeulebroecke make the interesting observation that this introduces a time component which may not be appropriate and which may ‘force the pace’ for the viewer.

This is a very apt comment when the displays are viewed in prepared video form, or under the control of a presenter, as they were at the meeting. However, the principal use should be interactive, with the reader able to control the pace, and the R code that is available on line provides a mechanism which allows this. I believe that this mode of operation goes a long way to addressing the concerns which have been raised. Some issues remain when animations are embedded within the scientific papers which are the standard currency of research activity. I would like to endorse strongly Cressie's plea that animation should be available as an integral component of a paper in electronic form. The technology to allow this exists and will hopefully become a more routine part of scientific publication.

Reese has made a variety of criticisms of the details of some of the displays and examples. In partial defence I would plead that the graphics are intended to be viewed in the context of the surrounding text of the paper, and the commentary in live presentation, where some further explanation is provided. There will also inevitably be a variety of style preferences with, for example, Reese dubious of the merits of adding jittered data to a density strip whereas Jackson approves. The merits of 'freeze-frame' forms of animation is also covered, I believe, by the intended principal mode of delivery which gives the user control, as discussed above. More generally, the robust debate which attaches to any form of data analysis is always welcome.

The discussion contributions have also provided helpful reminders of the merits of a wide variety of other methods of graphical display. Jackson helpfully comments on different aspects of density strips and draws attention to the merits of 'ridge line plots'. Atkinson argues very persuasively for the merits of 'brushing' as a well-established technique, whereas King makes imaginative and very effective use of underground train network maps. Friendly has made a convincing case for the use of 'fourfold displays' in the Doll and Hill example, whereas Wijayatunga has discussed Simpson's paradox in the context of contingency tables and Holmström has strongly advocated scale space methods. This illustrates the wide array of graphical tools available to us.

Several discussants, most notably Jones, Noble and Krone, have laid particular emphasis on the need for education in data visualization to release its very considerable potential for effective communication within the wider community. I wholeheartedly agree. Data visualization is a considerably underrated topic. The ability to construct an appropriate and informative graphic depends on a good understanding of the question being asked and the structure of the data that are available. Data visualization can therefore often be an excellent *entrée* into the key aspects of any quantitative analysis.

References in the discussion

Atkinson, A. C. (1973) Testing transformations to normality. *J. R. Statist. Soc.* B, **35**, 473–479.

Atkinson, A. C. and Riani, M. (2004) The forward search and data visualisation. *Computnl Statist.*, **19**, 29–54.

Atkinson, A. C. and Riani, M. (2006) Distribution theory and simulations for tests of outliers in regression. *J. Computnl Graph. Statist.*, **15**, 460–476.

Atkinson, A. C., Riani, M. and Cerioli, A. (2004) *Exploring Multivariate Data with the Forward Search*. New York: Springer.

Atkinson, A. C., Riani, M. and Cerioli, A. (2010) The forward search: theory and data analysis (with discussion). *J. Kor. Statist. Soc.*, **39**, 117–134.

Becker, R. A. and Cleveland, W. S. (1987) Brushing scatterplots. *Technometrics*, **29**, 127–142.

Berger, J. (1972) *Ways of Seeing*. London: Penguin.

Bradley, R. and Haslett, J. (1992) High-interaction diagnostics for geostatistical models of spatially referenced data. *Statistician*, **41**, 371–380.

Brooks, K. R. (2017) Depth perception and the history of three-dimensional art: who produced the first stereoscopic images? *I-perception*, **8**, article 2041669516680114.

Carroll, S. S. and Cressie, N. (1996) A comparison of geostatistical methodologies used to estimate snow water equivalent. *J. Am. Wat. Resour. Ass.*, **32**, 267–278.

Castruccio, S. and Genton, M. G. (2016) Compressing an ensemble with statistical models: an algorithm for global 3D spatio-temporal temperature. *Technometrics*, **58**, 319–328.

Castruccio, S. and Genton, M. G. (2018) Principles for statistical inference on big spatio-temporal data from climate models. *Statist. Probab. Lett.*, **136**, 92–96.

Cleveland, W. S. (1994) *Elements of Graphing Data*, revised edn. Summit: Hobart Press.

Cleveland, W. S. and McGill, R. (1985) Graphical perception and graphical methods for analyzing scientific data. *Science*, **229**, 828–833.

Cook, D., Majure, J. J., Symanzik, J. and Cressie, N. (1996) Dynamic graphics in a GIS: exploring and analyzing multivariate spatial data using linked software. *Computnl Statist.*, **11**, 467–480.

Cox, D. R. (1978) Some remarks on the role in statistics of graphical methods. *Appl. Statist.*, **27**, 4–9.

Cressie, N. A. C. (1993) *Statistics for Spatial Data*. New York: Wiley.

Cruz-Neira, C., Sandin, D. J., DeFanti, T., Kenyon, R. and Hart, J. C. (1992) The CAVE audiovisual experience automatic virtual environment. *Communs ACM*, **35**, 36–72.
Dai, W. and Genton, M. G. (2018a) Functional boxplots for multivariate curves. *Stat*, **7**, article e190.
Dai, W. and Genton, M. G. (2018b) Multivariate functional data visualization and outlier detection. *J. Computnl Graph. Statist.*, to be published.
Davison, A. C. and Huser, R. (2015) Statistics of extremes. *A. Rev. Statist. Appl.*, **2**, 203–235.
Davison, A. C., Huser, R. and Thibaud, E. (2018) Spatial extremes. In *Handbook of Environmental and Ecological Statistics* (eds A. E. Gelfand, M. Fuentes and R. L. Smith). Boca Raton: CRC Press. To be published.
Davison, A. C., Padoan, S. and Ribatet, M. (2012) Statistical modelling of spatial extremes (with discussion). *Statist. Sci.*, **27**, 161–186.
Diggle, P. J., Moyeed, R. A. and Tawn, J. A. (1998) Model-based geostatistics (with discussion). *Appl. Statist.*, **47**, 299–350.
Doll, R. and Hill, A. B. (1950) Smoking and carcinoma of the lung. *Br. Med. J.*, no. 4682, 739–748.
Fassò, A., Finazzi, F. and Madonna, F. (2018) Statistical issues in radiosonde observation of atmospheric temperature and humidity profiles. *Statist. Probab. Lett.*, **136**, 97–100.
Fassò, A., Verhoelst, T. and Lambert, J.-C. (2018b) Report on the scientific assessment of gaps using a statistics approach. *Gap Analysis for Integrated Atmospheric Essential Climate Variables Climate Monitoring Deliverable D1.10*. (Available from `http://www.gaia-clim.eu/page/deliverables`.)
Finazzi, F. and Fassò, A. (2014) D-STEM: a software for the analysis and mapping of environmental space-time variables. *J. Statist. Softwr.*, **62**, no. 6, 1–29.
Finazzi, F., Fassò, A., Madonna, F., Negri, I., Sun, B. and Rosoldi, M. (2018) Statistical harmonization and uncertainty assessment in the comparison of satellite and radiosonde climate variable. *Environmetrics*, to be published, doi 10.1002/env.2528.
Friendly, M. (1994a) A fourfold display for 2 by *K* tables. *Report 217*. Department of Psychology, York University, Toronto. (Available from `http://datavis.ca/papers/4fold/`.)
Friendly, M. (1994b) Mosaic displays for multi-way contingency tables. *J. Am. Statist. Ass.*, **89**, 199–200.
Friendly, M., Monette, G. and Fox, J. (2013) Elliptical insights: understanding statistical methods through elliptical geometry. *Statist. Sci.*, **28**, 1–39.
Gelman, A. and Hennig, C. (2017) Beyond subjective and objective in statistics (with discussion). *J. R. Statist. Soc.* A, **180**, 967–1033.
Gelman, A. and Loken, E. (2014) The statistical crisis in science. *Am. Scient.*, **102**, 460.
Gelman, A. and Price, P. N. (1999) All maps of parameter estimates are misleading. *Statist. Med.*, **18**, 3221–3234.
Genton, M. G., Castruccio, S., Crippa, P., Dutta, S., Huser, R., Sun, Y. and Vettori, S. (2015) Visuanimation in statistics. *Stat*, **4**, 81–96.
Genton, M. G., Johnson, C., Potter, K., Stenchikov, G. and Sun, Y. (2014) Surface boxplots. *Stat*, **3**, 1–11.
Gordon I. and Finch, S. (2015) Statistician heal thyself: have we lost the plot? *J. Computnl Graph. Statist.*, **24**, 1210–1229.
Heer, J. and Bostock, M. (2010) Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pp. 203–212. New York: Association for Computing Machinery.
Hixon, D. (2018) Going deep—art techniques used to show depth. ECG Productions, Atlanta. (Available from `https://www.ecgprod.com/going-deep-art-techniques-show-depth/`.)
Holmström, L. and Pasanen, L. (2017) Statistical scale space methods. *Int. Statist. Rev.*, **85**, 1–30.
Huser, R. and Davison, A. C. (2014) Space–time modelling of extreme events. *J. R. Statist. Soc.* B, **76**, 439–461.
Institute for Ageing (2014) Healthy life simulation. *Report*. Institute for Ageing, Newcastle upon Tyne. (Available from `https://www.ncl.ac.uk/media/wwwnclacuk/instituteforageing/files/healthy-life-report.pdf`.)
Jackson, C. H. (2008) Displaying uncertainty with shading. *Am. Statistn*, **62**, 340–347.
Kang, E. L., Liu, D. and Cressie, N. (2009) Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models. *Computnl Statist. Data Anal.*, **53**, 3016–3032.
Lamigueiro, O. P. (2018) *Displaying Time Series Spatial, and Space–Time Data with R*, 2nd edn. Boca Raton: Chapman and Hall–CRC.
Lombardo, L., Opitz, T. and Huser, R. (2018) Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. *Stoch. Environ. Res. Risk Assessmnt*, **32**, 2179–2198.
Margolskee, A., Baillie, M., Magnusson, B., Jones, J. and Vandemeulebroecke, M. (2018) Graphics principles cheat sheet. Biostatistical Sciences and Pharmacometrics, Novartis Institutes for Biomedical Research, Cambridge. (Available from `https://graphicsprinciples.github.io/`.)
Marmot, M. (2010) Fair society, healthy lives: strategic review of health inequalities in England post-2010. *The Marmot Review*. University College London, London. (Available from `www.ucl.ac.uk/marmotreview`.)
Meehl, P. E. (1990) Why summaries of research on psychological theories are often uninterpretable. *Psychol. Rep.*, **66**, 195–244.
Munzner, T. (2014) *Visualization Analysis and Design*. Boca Raton: CRC Press.
Reese, R. A. (2017) Graphical interpretations of data: an introduction. *Significance*, **14**, no. 4, 42–43.

Reese, R. A. (2019) Graphical interpretations of data: east is east. *Significance*, **16**, no. 1, in the press.
Riani, M. and Atkinson, A. C. (2000) Robust diagnostic data analysis: transformations in regression (with discussion). *Technometrics*, **42**, 384–398.
Riani, M., Perrotta, D. and Torti, F. (2012) FSDA: a matlab toolbox for robust analysis and interactive data exploration. *Chemometr. Intell. Lab. Syst.*, **116**, 17–32.
Sheiner, L. B. (1997) Learning versus confirming in clinical drug development. *Clin. Pharmacol. Ther.*, **61**, 275–291.
Sievert, C. (2018) plotly for R. Sievert Consulting (Available from `https://plotly-book.cpsievert.me`.)
Spiegelhalter, D., Pearson, M. and Short, I. (2011) Visualizing uncertainty about the future. *Science*, **333**, 1393–1400.
Sun, Y. and Genton, M. G. (2011) Functional boxplots. *J. Computnl Graph. Statist.*, **20**, 316–334.
Sun, Y. and Genton, M. G. (2012) Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics*, **23**, 54–64.
Sun, Y., Genton, M. G. and Nychka, D. (2012) Exact fast computation of band depth for large functional datasets: how quickly can one million curves be ranked? *Stat*, **1**, 68–74.
Tufte, E. R. (1983) *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.
Tukey, J. W. (1977) *Exploratory Data Anaysis*. Reading: Addison-Wesley.
Tversky, B., Morrison, J. B. and Betancourt, M. (2002) Animation: can it facilitate? *Int. J. Hum. Comput. Stud.*, **57**, 247–262.
Vandemeulebroucke, M., Baillie, M., Carr, D., Kanitra, L., Margolskee, A., Wright, A. and Magnusson, B. (2018) How can we make better graphs?: An initiative to increase the graphical expertise and productivity of quantitative scientists. *Pharm. Statist.*, to be published.
Vuollo, V. and Holmström, L. (2018) A scale space approach for exploring structure in spherical data. *Computnl Statist. Data Anal.*, **125**, 57–69.
Wang, Z. and Bovik, A. C. (2009) Mean squared error: love it or leave it?—a new look at signal fidelity measures. *IEEE Signl Process. Mag.*, **26**, 98–117.
Wang, Z., Bovik, A., Sheikh, H. and Simoncelli, E. P. (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Im. Process.*, **13**, 600–612.
Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
Wijayatunga, P. (2014) Viewing Simpson's paradox. *Statist. Appl.*, **12**, 225–235.
Wild, C. J., Pfannkuch, M., Regan, M. and Horton, N. J. (2011) Towards more accessible conceptions of statistical inference (with discussion). *J. R. Statist. Soc.* A, **174**, 247–295.
Wikle, C. K., Zammit-Mangion, A. and Cressie, N. (2019) *Spatio-temporal Statistics with R*. Boca Raton: Chapman and Hall–CRC.
Wilke, C. O. (2018) ggridges: ridgeline plots in 'ggplot2'. *R Package Version 0.5.0*. University of Texas at Austin, Austin. (Available from `https://CRAN.R-project.org/packages=ggridges`.)
Yan, Y. and Genton, M. G. (2018) Gaussian likelihood inference on data from trans-Gaussian random fields with Matérn covariance function. *Environmetrics*, **29**, article e2458.
Zhang, B., Cressie, N. and Wunch, D. (2017) Statistical properties of atmospheric greenhouse gas measurements looking down from space and looking up from the ground. *Chemometr. Intell. Lab. Syst.*, **162**, 214–222.
Zoure, H. G. M., Noma, M., Tekle, A. H., Amazigo, U. V., Diggle, P. J., Giorgi, E. and Remme, J. H. F. (2014) The geographic distribution of onchocerciasis in the 20 participating countries of the African Programme for Onchocerciasis Control: 2, Pre-control endemicity levels and estimated number infected. *Parasit. Vect.*, **7**, 326.