Map LineUps: implications for spatial analysis

R. Beecham*[1] and J. Dykes[2]

[1,2]giCentre, City, University of London

*Email: roger.beecham@city.ac.uk

Abstract

Fundamental to the use of statistical graphics is the assumption that any statistical effect implied by a graphic can be reliably perceived. This is a particular concern for geovisualization since unanticipated artefacts are often introduced when presenting data within their spatial context. Techniques for supporting inferences made with graphics, specifically graphical line-up tests, have recently been proposed but not widely used, whilst within Information Visualization and Human Computer Interaction, a new class of experimental research aims to observe and model human perception of statistical properties implied by graphics. One such study found that the precision with which spatial autocorrelation structure is perceived in Choropleth maps varies as a function of the intensity of autocorrelation structure under investigation and the geometric irregularity of spatial units that comprise regions. These experimental findings were used to make claims around the circumstances under which Choropleth maps can be reliably used to study spatial structure. Such studies are prescient at a moment when alternative mechanisms for communicating statistical effect, for example using graphics, are being proposed as a means of broadening scientific communication and statistical literacy. We hope that exposing this work to a community expert in spatial analysis will lead to a lively discussion at Geocomp 2017.

Keywords: Spatial autocorrelation, graphical perception, graphical inference, geovisualization.

1 Graphical inference and the challenge of (geo)visualization

If statistical graphics are to be used in data analysis and reporting, then there needs to be some reassurance that the statistical effect implied by a graphic can be reliably perceived. Any possible mismatch between statistical effect and its graphical perception is a particular challenge to geovisualization. Whilst maps convey information around the location and extent of phenomena that may be difficult to imagine using non-visual techniques, they may also lead to artefacts that are incidental to the statistical structure under investigation and that may even induce interpretation of false structure.

Techniques for supporting *graphical inference* (Wickham et al., 2010) represent a means of enabling more consistent interpretation. One such technique is the line-up protocol. The graphical line-up can be understood as a visual equivalent to a test statistic. A plot of real data is hidden amongst a set of decoys generated under a *null* hypothesis. If an impartial observer, an individual who has not previously seen the plot, is able to correctly identify the real from the decoys, then this lends

*

confidence to the initial finding – or rather, following *null hypothesis significance testing*, that the observed data are not consistent with the specified *null*.

The example 'map' line-up in Figure 1 is constructed under a *null* hypothesis of spatial independence, or *complete spatial randomness* (CSR). Such a configuration is consistent with a long-standing concern in geography with spatial autocorrelation: the extent to which 'near things are more related than distant things' (Tobler, 1970), or casually, *Tobler's First Law of Geography*. When testing for spatial autocorrelation, geographers generate statistics that describe how probable the spatial structure in an observed dataset would be if CSR were operating – following the example in Figure 1, the probability of the observed distribution in crime rates were it drawn from a hypothesised population in which crime were distributed independently of location. One means of creating such a test statistic is to calculate an autocorrelation summary measure, such as *Moran's I*, described in detail elsewhere (e.g. O'Sullivan and Unwin, 2010), but which can be understood as the distance-weighted co-variation in attribute values over space. When testing for spatial autocorrelation, *Moran's I* can be compared against a theoretical *null* distribution. Since the statistic is partly a function of the geometric properties of the region under investigation – the map is also a parameter in the analysis (O'Sullivan and Unwin, 2010) – a superior approach is to generate this distribution empirically, by randomly permuting attribute values between spatial units many times and on each run calculating a value for *Moran's I*. The map line-up in Figure 1 can be understood, at least procedurally, as a patial analogue of this latter approach.

The assumption of CSR has received some criticism since if *Tobler's first law* is accepted, then CSR should never exist (O'Sullivan and Unwin, 2010). Within the more informal framework of graphical inference, one proposition is to construct decoys in map line-up tests that assume autocorrelation based on some sensible prior knowledge (Beecham et al., 2017): for the example in Figure 1, decoys that contain an autocorrelation structure typically observed in crime datasets for areas of similar population dynamics.

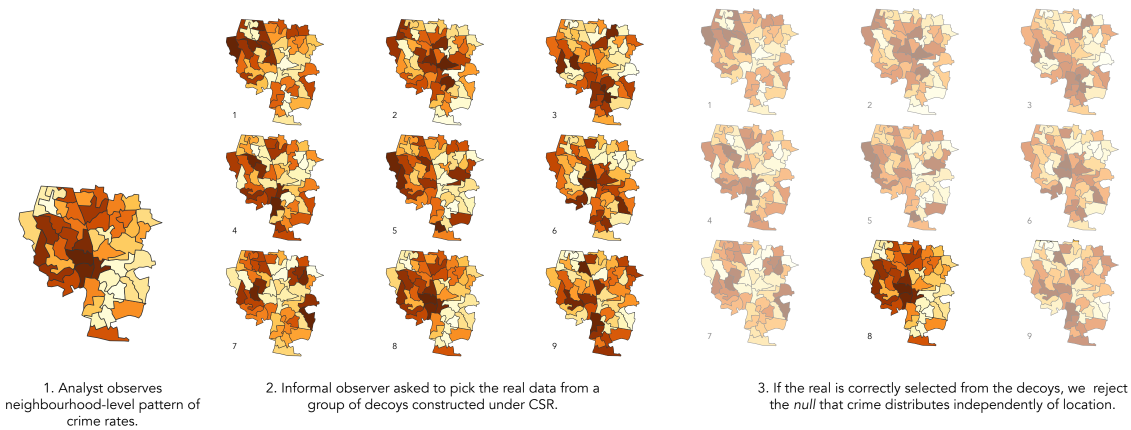

Figure 1: Line-up protocol as described in Wickham et al. (2010) – a line-up is a visual equivalent to a test statistic.

2 Attending to perception: map line-up $\neq_n$ test statistic

Although procedurally straight-forward and conceptually appealing, line-ups is a technique that has yet to be widely adopted. There are few case study examples of graphical line-up tests being regularly used to guard against false interpretation. If line-up tests are to represent a genuine technique to the extent that they might be presented in research papers, technical reports or press releases in place of more formal statistics, then there again needs to be evidence that the statistical effect encoded in line-ups can be reliably perceived.

Three notable studies (Rensink and Baldridge, 2010; Harrison et al., 2014; Kay and Heer, 2016) have recently carried out large-scale experiments measuring the precision with which non-spatial correlation (*Pearson's correlation coefficient*) can be perceived in statistical graphics. Using an established methodology in cognitive psychology, these studies found that the precision with which two plots displaying varying correlation effect can be correctly judged as different varies systematically with the intensity of effect (Rensink and Baldridge, 2010; Harrison et al., 2014) and with visualization design (Harrison et al., 2014). Importantly, the studies demonstrate how these these differences can be modelled: Harrison et al. (2014) and Kay and Heer (2016) estimate the precision with which correlation structure can be distinguished in different visualization types and at different intensities of correlation coefficient.

Beecham et al. (2017) used the same experimental procedure as these earlier studies to measure the precision with which *spatial* autocorrelation, as measured by *Moran's I*, can be distinguished in Choropleth maps. Following Harrison et al. (2014), Beecham et al. (2017) varied systematically the intensity of autocorrelation extent in visualization pairs (in this case pairs of choropleths), but rather than testing different visualization types they varied the geometric properties of the maps (as in Figure 2). The assumption was that with greater geometric irregularity (as measured by variance in spatial unit area), the greater the likelihood of visual artefacts interfering with perception. Consistent with the studies of non-spatial correlation, Beecham et al. (2017) found that ability to distinguish two maps of different spatial autocorrelation structure varies with the *amount* of baseline autocorrelation. Introducing greater irregularity into the geometry of choropleths also impacts upon perception and in the direction anticipated – the difference in autocorrelation required to discriminate maps is larger with greater irregularity. An important observation is that both these factors – geometric irregularity and intensity of autocorrelation effect – influence variability in perception.

The findings described in Beecham et al. (2017) have implications for the design of map line-up tests. If a relationship between visual perception and intensity of statistical effect exists and can be modelled, then the case for line-up tests that assume spatial autocorrelation rather than CSR becomes more compelling. If no adjustment is made for this relationship then map line-up designs become unduly sensitive where the phenomena being studied are very spatially autocorrelated. Equally, if the difference in statistical effect required to perceive differences in autocorrelation increases with geometric irregularity then line-up designs become less sensitive, and more prone to vary, in cases of irregular regions. These two properties – intensity of statistical effect and geometric irregularity – therefore relate to the likely *power* of a line-up test. In statistics *power* is the probability of rejecting the *null* if there is a true effect of a particular stated size in the population; in a line-up, it could be interpreted as the probability of correctly identifying the real from the decoys. Conventionally, *power* depends on experimental design – on sample size, confidence level and target effect. When constructing visual line-up tests with maps, however, Beecham et al. (2017) argue that *power* is likely to vary not only with sample size, or factors specific to lineups such as the number of decoys presented, but with intensity of autocorrelation and with geometric irregularity.

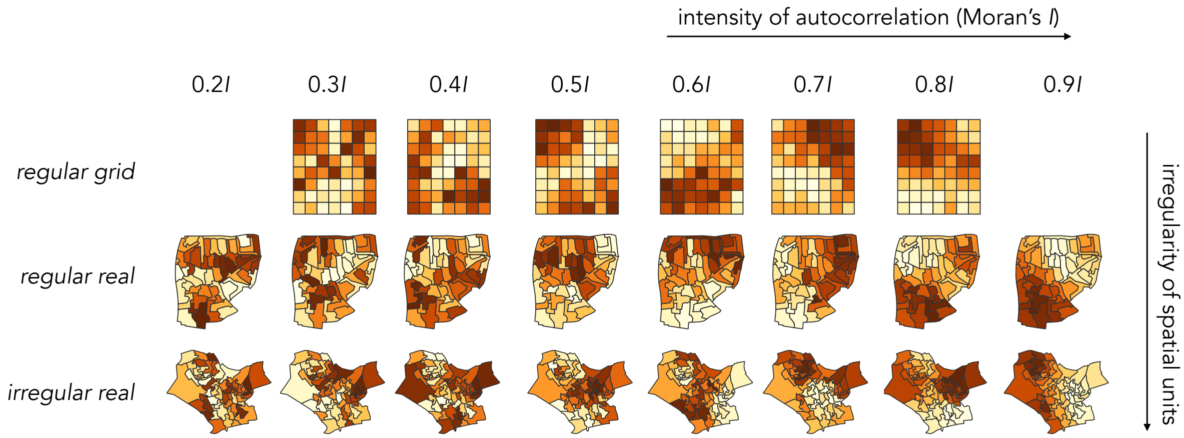

Figure 2: Conditions tested in Beecham et al. (2017).

3 Speculating on the sources of perceptual error

Whilst the model and results described in Beecham et al. (2017) may be used to make decisions on the configuration and design of map line-up tests, an important observation is of substantial between-participant variability in perception. One explanation is that data were collected via a large-scale crowdsourcing platform, Amazon Mechanical Turk (AMT), rather than from a sample known to, and experimental conditions controlled by, researchers. A counter argument to this is that Harrison et al.'s (2014) study was conducted via the AMT platform, but the authors were able to replicate the same findings as those published in an earlier study conducted using a more traditional, lab-based environment (Rensink and Baldridge, 2010). The variation observed in the spatial perception tests might instead relate to artefacts introduced into the more irregular geographies that could not be controlled in a systematic way: for example, biases introduced through a dominant colour effect (Klippel et al., 2011), the influence of geometric shapes or lineations that only become salient when attribute data are added to a map, or instances of extreme proximity in spatial unit centroids, which may have an inordinate effect on the spatial autocorrelation statistic but which are hidden visually. Modelling these types of artefacts is problematic and remains a challenge for experimental research.

4 Towards empirically-supported (geo)visualization

The approaches and studies discussed here represent a burgeoning class of research from within Information Visualization and Human Computer Interaction around the interpretation of statistical structure in graphics (Correll and Gleicher, 2014; Rensink and Baldridge, 2010; Harrison et al., 2014; Kay and Heer, 2016; Beecham et al., 2017; Correll and Heer, 2017). This work is highly prescient at a moment when the scientific community is demanding a New Statistics (Cumming, 2012) and new mechanisms for communicating scientific results (McInerny et al., 2014). With the exception of Beecham et al. (2017) and previously Klippel et al. (2011), comparatively few empirical studies have been published on the perception of spatial statistical structure. We hope that exposing this work to a community expert in spatial analysis will effect a lively discussion at Geocomp 2017.

5 References

Beecham, R., J. Dykes, W. Meulemans, A. Slingsby, C. Turkay, and J. Wood
2017. Map line-ups: effects of spatial structure on graphical inference. *IEEE Transactions on Visualization & Computer Graphics*, 23(1):391–400.

Correll, M. and M. Gleicher
2014. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization & Computer Graphics*, 20(12):2141–2151.

Correll, M. and J. Heer
2017. Regression by eye: Estimating trends in bivariate visualizations. In *ACM Human Factors in Computing Systems (CHI)*.

Cumming, G.
2012. *Understanding The New Statistics: Effect sizes, confidence intervals, and meta-analysis*. London, UK: Routledge.

Harrison, L., F. Yang, S. Franconeri, and R. Chang
2014. Ranking visualizations of correlation using Weber's Law. *IEEE Conference on Information Visualization (InfoVis)*, 20:1943–1952.

Kay, M. and J. Heer
2016. Beyond Weber's Law: A second look at ranking visualizations of correlation. *IEEE Trans. Visualization & Comp. Graphics (InfoVis)*, 22:469–478.

Klippel, A., F. Hardisty, and R. Li
2011. Interpreting spatial patterns: An inquiry into formal and cognitive aspects of Tobler's First Law of Geography. *Annals of the Association of American Geographers*, 101(5):1011–1031.

McInerny, G. J., M. Chen, R. Freeman, D. Gavaghan, M. Meyer, F. Rowland, D. J. Spiegelhalter, M. Stefaner, G. Tessarolo, and J. Hortal
2014. Information Visualisation for science and policy: engaging users and avoiding bias. *Trends in Ecology & Evolution*, 29(3):148–157.

O'Sullivan, D. and D. Unwin
2010. *Geographic Information Analysis*, 2 edition. New Jersey, USA: John Wiley & Sons.

Rensink, R. and G. Baldridge
2010. The perception of correlation in scatterplots. *Computer Graphics Forum*, 29:1203–1210.

Tobler, W.
1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240.

Wickham, H., D. Cook, H. Hofmann, and A. Buja
2010. Graphical Inference for Infovis. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):973–979.