

# Statistification or Mystification? The Need for Statistical Thought in Visual Data Mining

Antony Unwin

Augsburg University

**Abstract.** Many graphics are used for decoration rather than for conveying information. Some purport to display information, but provide insufficient supporting evidence. Others are so laden with information that it is hard to see either the wood or the trees. Analysing large data sets is difficult and requires technically efficient procedures and statistically sound methods to generate informative visualisations. Results from big data sets are statistics and they should be statistically justified. Graphics on their own are indicative, but not substantive. They should inform and neither confuse nor mystify.

This paper will NOT introduce any new innovative graphics, but will discuss the statistification of graphics – why and how statistical content should be added to graphic displays of large data sets. (There will, however, be illustrations of the Ugly, the Bad and the possibly Good.)