



Enabling young learners to distinguish mere association from cause-and-effect.

Tim Dunne, *

University of Cape Town, South Africa – tim.dunne@uct.ac.za

The Big Data challenge in the 21st century classroom is to support post hoc discovery and explanations of relationships between variables in convenient and relevant data sets. The data sets are likely to be Open Data resources. The construction of meaning within the data exploration process may begin with the provision of leading questions. Availability of data tabulation and visualisation techniques through suitably targeted software programs will permit data queries that reflect the nascent enquiry processes of young learners. The emergence of robust patterns of retrospective relationships in a data set will invite an inference of the prospective continuation of the pattern. This extrapolation is usually founded upon the assumption or the belief that the patterns of the past will be replicated into the future. The replication is assumed to apply at least for a period of time that might be explicitly stated. The data of the past is assigned the authority of being representative into the near future at least. Even when this inferential process is mistaken, the appeal to representativity of current data for all similar data is in some sense a natural part of naïve sense-making from the world of observation. What is required in the classroom is the introduction of critical stances that allow the assumptions and inferences to be more explicitly described. In this context it is important to distinguish between the language of comparison and contrast (switches of focus across subgroups), and notions of change within observable units themselves. Single-factor causation is a simplest form of cause and effect. Again the invocation of this simplest structure is a natural start for inferences amongst young data explorers in a classroom setting. The cause-effect relationship is a frequent device in the sense-making of everyday life, whenever we hazard a motivated prospective outlook. This currency makes the form a particularly tempting kind of inference. However in Big Data or large Open Data contexts, this type of inference is inherently more risky in terms both of being erroneous, and of having gravely consequential loss outcomes. This paper will argue for the importance of an explicit vocabulary of causality to be developed alongside quantitative methods and visualisation tools as early as possible in the experiences of Big Data or Open Data that are brought into the 21st century classroom.

Keywords: retrospective data; prospective change-change association