Statistical literacy and multivariate thinking

James Nicholson*, Jim Ridgway and Sean McCusker

SMART Centre, Durham University, England j.r.nicholson@durham.ac.uk

When politicians talk about 'old people' we intuitively know that they are not talking about an homogenous group of people. The characteristics of different subgroups of 'old people' are important in understanding the likely impact of different political decisions on their well-being. Relevant characteristics include: health status, economic situation, family support and community support networks and the local infrastructure – these may have important influences on how any new government policy for 'old people' impacts an individual old person. Ethnicity and cultural background are also likely to be factors in the level of family and community support the person has. The ability to disaggregate a broad population into subgroups and to anticipate the likely support needs of these different subgroups is an important aspect of statistical literacy in modern society.

In considering risk factors there is a natural tendency to consider individual risk profiles and to adopt an additive mental model of how the factors combine. However, the reality across a wide range of contexts such as epidemiology, security systems, crowd movement at events, migration, financial markets and weather patterns is that there are important interaction effects. While it may not be important for people to know the detailed characteristics of each of these situations unless they are working in that specific area, it would be a good thing if they were aware of the key aspects to attend to when considering complex data.

The multivariate thinking in the title refers to some broad principles of working with multivariate data and includes an awareness of interactions, confounding variables, non-linear and piecewise functions and the use of composite measurement instruments. This paper will illustrate an approach to broadening statistical literacy through the use of data visualisation in familiar data contexts that can be used to draw out some of these characteristics. Trying to teach statistical literacy formally is not an option in most situations, but where statistical bulletins contain 3, 4 or 5-way tables we believe that data visualisations embedded in the bulletin will allow the reader to explore the relationships in the data for themselves, rather than being reliant on headlines which are often over-simplistic.

The paper will report on progress made in making visualisations of data releases from the 2011 UK census available to the general population as well as embedding them in core subject content within school and college courses.

Key Words: data visualisation, interaction, curriculum, disaggregate.