Reply to: Review by Jan de Leeuw of *Correspondence Analysis and Data Coding with Java and R*, F. Murtagh, Chapman and Hall/CRC, 2005. Journal of Statistical Software, Vol. 14.

Jan's review of my book is challenging and invites a reply. As a review it is written well, except for being unfortunately wrong in every point.

While Jean-Paul Benzécri did dominate the data analysis scene in France for a considerable period, this was through formidable productivity, and through his large battalions of grad students. When I did my PhD in his lab at the end of the 1970s, I recall a "guesstimate" at the time of 75 students doing the first year of course-work (leading to the DEA diploma), and 75 students writing their theses. Benzécri was not the only data analyst though. In seeking a PhD topic, at that time I made a number of visits to Simon Régnier at the Maison des Sciences de l'Homme, on the Boulevard Raspail. I wanted to pursue work on a probability model of data sets with power law characteristics, as one finds in information retrieval. (This had been the topic of a previous MSc dissertation of mine, and a short account was published in "Une question de classifiabilité en classification automatique dans le cas particulier des rassemblements documentaires", Statistique et Analyse des Données, 5, 77–889, 1980.) On one occasion, not very long in France, and to my everlasting embarrassment, I was wrongly using the word "graphe" as often used in English to mean graphic. Régnier, with his pencil, said that graph meant one thing and one thing only, furiously drawing vertices and edges on paper, with such force that the pencil snapped in two. Some time after, I also went to talk to Edwin Diday about a thesis orientation, and Edwin told that if I was so interested in the work of R.F. Ling on random graph models, well then, I should have gone to the U.S. instead to do my thesis. Touché! For me, Diday was not close to Benzécri in publication space, and Régnier was even more distant.

Benzécri's work was never based on an "esoteric philosophy of science and data analysis". It was as broad as could be, ranging freely over analysis of high energy physics detectors, to Plato's "The One". My own interests are mostly not now in the social and behavioral sciences, which Jan indicates as being well served already with a solid correspondence analysis tradition, a fact with which I am well pleased. Making the somewhat different point that there is little that is modern in the book, Jan mentions that I have "however briefly, mixed in some references to ... work in neural networks and Kohonen maps." My reason for having such material is fully missed on Jan. Quite simply, my view of the likely reader of this book is one who will more likely know what neural nets are all about, and will have little idea of statistical modeling, and very little interest in applications in the social and behavioral sciences.

I am condemned as "insular" because I did not pay attention to other packages or programs for correspondence analysis and clustering. Yet a hierarchical clustering algorithm of mine is in R. In this book, I use a version of this algorithm which assumes weights on the entities being clustered. For such work, I personally rarely ever use other programs simply because they do not support the nearest neighbor chain algorithms that have been state of the art (for their particular hierarchical clustering aims) since around 1980. As regards other correspondence analysis programs, I would have cited them for such functionality as cross-linkage between the factor and the cluster analyses (VACOR, FACOR) if they were available. I have to admit it: I have written my own code and I use it because I have not found anything else elsewhere which does what I want, and what I use all the time in my research. That's life. My book is weak, to put it mildly, on graphical user interfaces which just shows my biases: I am and remain interested in the algorithms. This is not a matter of being old school in such matters, since the software engineer of today (I have educated enough of them...) more often than not needs classes (in the OO sense) quickly, and is quite adept at user interface development. So the latter is in safe hands, and was not my ambition in this book.

While I said that Jan's review was mistaken left, right and center, there is one problem of such major proportions that if I do hope I don't go to my grave believing that the fault was one of exposition on my side. This is when Jan says: "To be sure, the book is unique and interesting ... It pays more attention than most English language books on correspondence analysis to coding, updating Pascal code published by Benzécri in 1998." Oh no... The coding of the title, and of the book, has absolutely *nothing* to do with computer codes. I thought any reviewer of this book would at least have glanced through, and would have noted that this was so. No, and this means misery on my side (book by Stephen King, screenplay William Goldman...). Data coding in the correspondence analysis tradition has nothing whatsoever to do with computer programming. It has all to do with data analysis and interpretation though. In fact it probably has a lot to do with epistemology. Data analysts have far too often just assumed the potential for extracting meaning from the given data, *telles quelles*. The statistician's way to address the problem works well sometimes but has its limits: some one or more of a finite number of stochastic models (often handled with the verve and adroitness of a maestro) form the basis of the analysis. The statistician's toolbox (or surgical equipment, if you wish) can be enormously useful in practice. But the statistician plays second fiddle to the observational scientist or theoretician who really makes his or her mark on the discovery. This is not fair.

Without exploring the encoding that makes up primary data we know very, very little. (As examples, we have the DNA codes of the human or any animal; discreteness at Planck scales and in one vista of the quantum universe; and we still have to find the proper encoding to understand consciousness.) This book of mine is a short but I hope reasonable initial entrée into Benzécri's work. For me the continuing enthralling aspect of Benzécri's work is the possibility opened up for the data analyst, through the data encoding question, to be a partner, hand in hand, in the process of primary discovery.

Fionn Murtagh (Department of Computer Science, Royal Holloway, University of London)