Chapter 3

GRAPHICAL EDITING

FOREWORD

By Ron James, United Kingdom

For the purposes of this chapter, the graphical editing is the exploitation of human visual perception ability to identify an anomaly, a pattern or an implied relationship in data that might take much more time and effort to find by analytic non-graphical means. Graphical data editing methods proved to be in many National Statistical Offices an efficient tool for the reduction of statistical survey costs.

Furthermore, the experiences show that traditional editing on a case-by-case basis does not often allow to see clearly the impact of an individual data point on the aggregate estimate. The obvious examples are the identification of an outlier for specific data and a simple relationship between different data in a set of cases. The outlier may represent an error or an exceptional movement. The simple relationship, perhaps represented as a straight line with suitably transformed data, may provide an easier visual method for finding an error or an important exception. Sometimes it is the nature of the relationship that it is valuable and easier to see as a result of graphing.

In those cases where the statistician has little prior knowledge of the response to a survey, preliminary exploration of the early data using graphics can be of significant help in setting up the bounds of the editing process. Also they can be used in imputation and editing itself, monitoring and tuning the editing process and in evaluating imputation procedures by exposing the aggregate properties of imputed data as compared with the overall data set.

The paper by Bienias et al. develops the data exploration theme using graphics and transformations to help expose trends and other properties of the data that may be difficult and more time consuming to find by purely analytical means. Knowledge gained in this way may then be used to define, improve or enhance the editing process.

Graphics can be used in a more specific way as an integral part of the editing process. These processes are essentially identifying exceptional data. They are exceptional either because they have a large effect on the aggregate results of the survey, or because they are erroneous, or both. In either case the graphical process has led the editor to the key data directly and efficiently. The technique of identifying those data that have most impact on the aggregate results is known as 'macro' or 'top down' editing. Such a technique which enables the editors to concentrate their resources on confirming or correcting the most influential data presumes that the remaining large number of small errors are non systematic and have little effect on the result. This has been confirmed by several studies. It does not necessarily mean that editing stops after the largest errors or outliers have been identified. Where the data is to be used for subsequent analyses, especially for purposes other than the initial aggregates it may be essential to complete the process of editing. However, such techniques are an efficient way of reducing the cost and time required to achieve specific results from a survey. The paper by Per Engström of the Swedish Statistical Bureau is a classic example of macro or top down editing.

It is typical of graphical methods that they provide insight, revealing things that were sometimes not dreamt of before the process began. This is a point made by Esposito et al. in describing the ARIES system used in the US Department of Labour. It is thought important that these graphical systems are open ended with flexible facilities to extend and develop new views responding to what the graphical information is revealing. A further example is provided in the paper on Graphical Editing and Analysis Query System by Paula Weir et al. of the US Department of Energy. The PC platform with its Graphical User Interface and an Object Oriented approach would seem to be an ideal environment for these developments.

IMPROVING OUTLIER DETECTION IN TWO ESTABLISHMENT SURVEYS