Macroview

Statistics Netherlands- Wim Hacking

2016 BLAISE 17TH INTERNATIONAL BLAISE USERS CONFERENCE

Abstract:

Why use macro-editing:
Micro-editing is a very time-consuming process. As analysts often want to avoid wrong records to pass the edits, the boundaries for micro-checks are often set conservatively, i.e. only error-less records are accepted. This means that there are often many checks with narrow tolerances resulting in too many mistakes that need to be resolved manually by analysts. An example of a frequently used edit is that the relative change compared to the last period cannot exceed ± 10%. Generally, this approach results in a considerable amount of over-editing.
Another problem with editing at the micro data level is that the analysts cannot assess the relative importance of errors. Each marked item has the same weight and needs the same amount of time for correction. However, many errors have a neglible impact on the final estimates: either they are small or they cancel out.

Why use macro-editing at Statistics Netherlands:
To address the issue of over-editing one can use macro-editing. This method was already in use at a number of places at Statistics Netherlands and other statistical offices, based on applications tailored for specific statistics. At Statistics Netherlands macro-editing is used for error spotting in international trade data. Another example of macro-editing is outlier detection where the (multivariate) distribution determines which records are suspect and possibly need to be corrected. These outliers may be determined automatically and subsequently be displayed, e.g. by coloring these points inside a scatter plot.

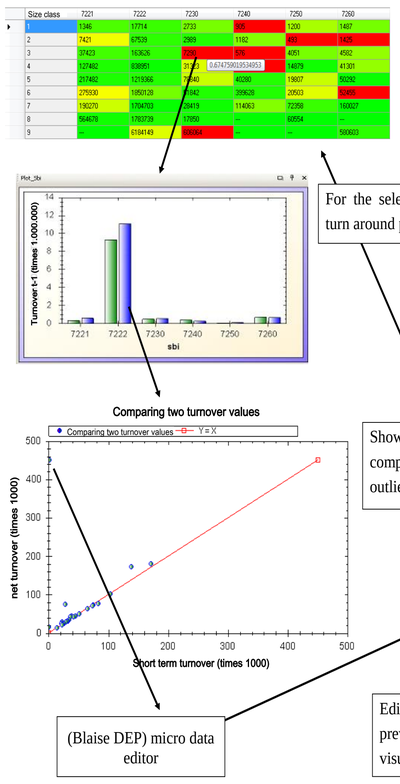
Although applications of the principles above exist at various statistical offices, to the best of our knowledge no general software exists, which can be applied to different statistics. Therefore such a tool called Macroview has been developed at Statistics Netherlands and is currently being applied to data from most economical statistics.

Design considerations:
To assure that Macroview can support all the approaches as discussed above, the software requirements needed for performing all the approaches have been determined.
- The software should be able to compute aggregates from micro data and to compare aggregates using specified rules.
- The software should be able to select those records that contribute the most to errors observed at some aggregate level. In order to asses this error one must have a reference set: this may be the dataset itself (e.g. using outlier detection) or (aggregated) reference data (using t-x data or data strongly correlated to the variables under study). The reference datasets can be used to make a prediction of the current data and the relative differences between the prediction and the actual values can be used to look for possible errors.
- The software should be able to display the data in various ways to enable the analyst to zoom in or analyze the data.
- The analyst should be able to zoom in at a selection of the data.
- The analyst needs to be able to edit a single record.

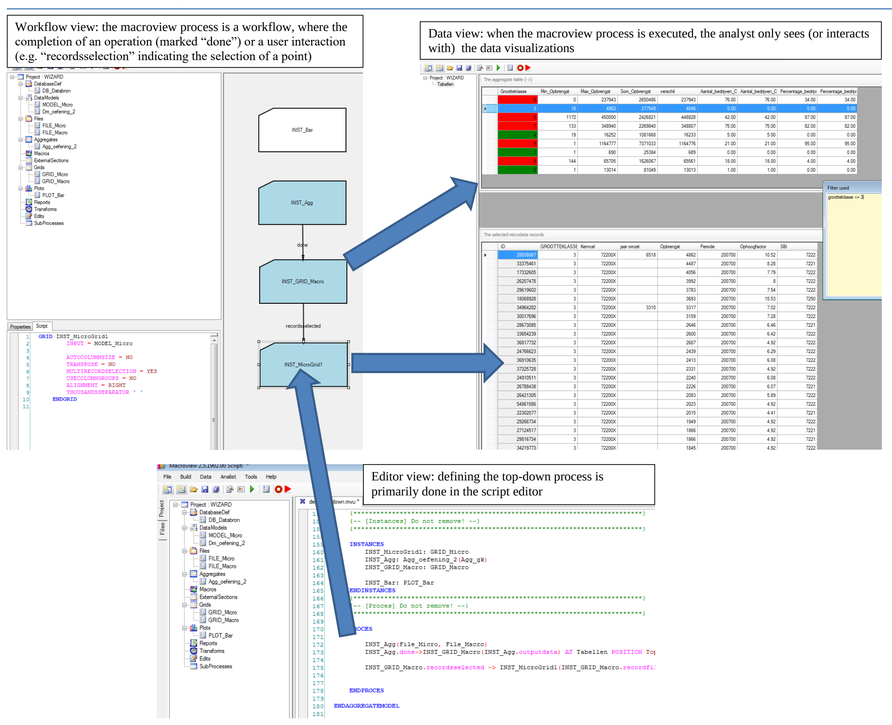
Elements of macroview:
To configure all of these possibilities, the analyst needs to specify the elements of the editing process and, finally, the interactive process itself:
- The input of the editing process: a data model of the micro data input and of the reference (aggregate) data.
- The variables and their derivation in the aggregate(s) (e.g. plausibility functions, distribution moments, …).
- Visualizations: what data need to be displayed, which colors need to be used, etc.
- The interaction of the program when used by the analyst. For example, when the analyst selects a cell from an aggregate: should a sub-aggregate be displayed or a scatter plot?

These possibilities can be specified using a custom scripting language within Macroview; this tool can be used both as an IDE to make the script and as the analysis/editing tool (as defined by the script). An additional and crucial aspect of Macroview is its ability to allow editing of micro data, using an external editor. After the editor has finished, Macroview will update all displayed graphs and plots very efficiently, so that the analyst can judge the impact of the manual edit(s).

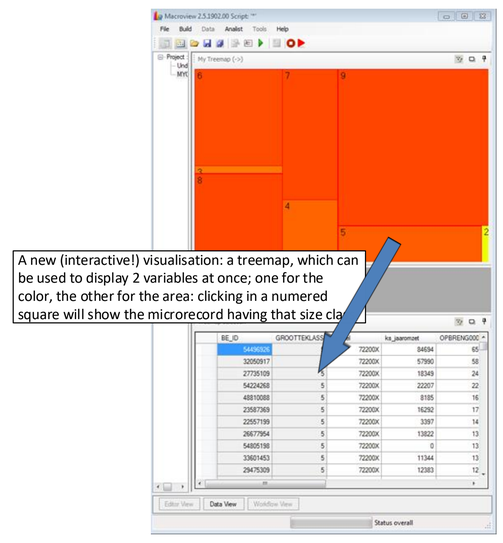
Basic top-down data editing:

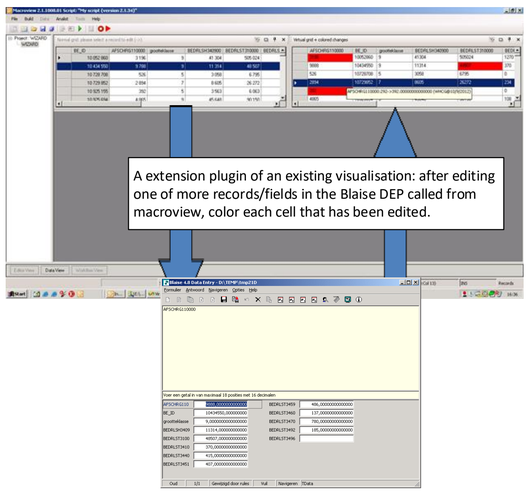


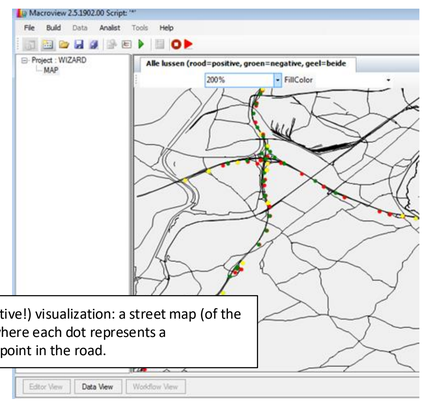
Different views

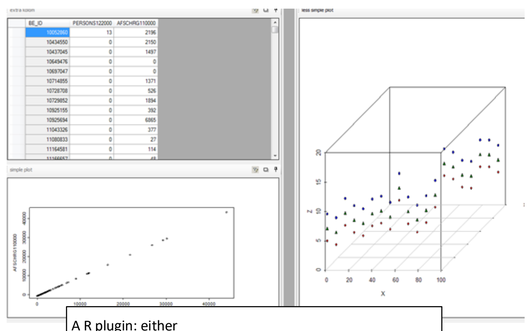


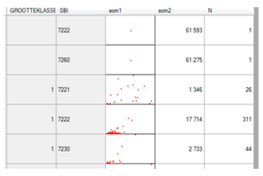
Possibilities using plugins:



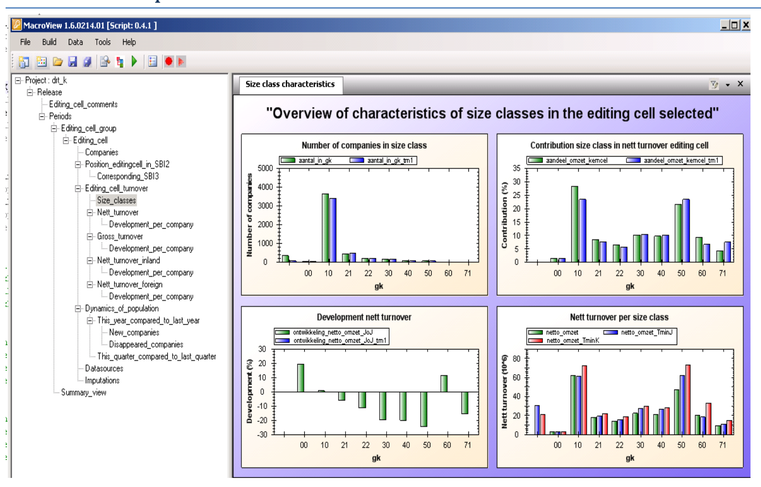




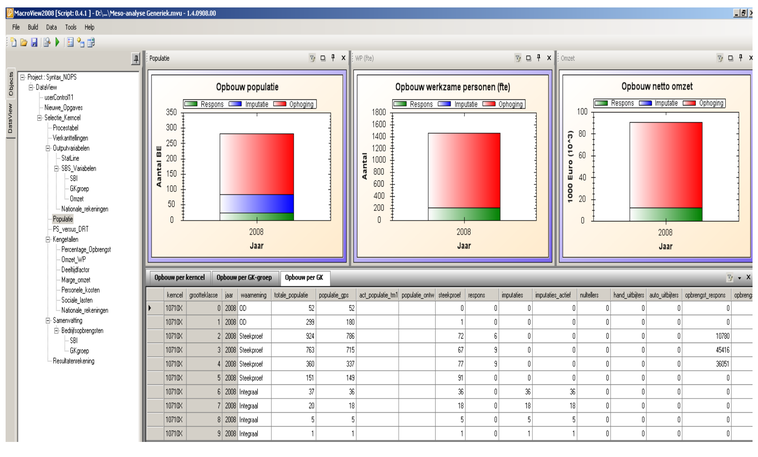




Real life examples:



An example of the analysis screens as used by the DRT project (redevelopment of the short term business statistics).



An example of the analysis screens as used by the NOPS project (redevelopment of the structural business statistics).

Usage and References:

Currently used at SN by

- ABR: Checking/cleaning up Business Registries
- DRT: Short term statistics
- KICR: phase 2 project for DRT
- IHD: International trade in services
- MUST: Environment statistics
- NOPS: A large project to redesign the Structural business statistics

- IHG: International trade in goods, currently being redesigned.
- V&V: Statistics on transport
- Gezo: Statistics on people's health

References:

Work Session on Statistical Data Editing (Ljubljana, Slovenia, 9-11 May 2011) Topic (iii): *Macro editing methods MacroView: a generic software package for developing macro-editing tools*. Saskia Ossen, Wim Hacking, Ralph Meijers, and Peter Kruiskamp, Statistics Netherlands (http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2011/wp.14.e.pdf)

Applying Macro Editing in MacroView. Wim Hacking , Saskia Ossen, Statistics Netherlands, Netherlands (http://ec.europa.eu/eurostat/cros/system/files/S2P2_0.pdf)

Conference of European statisticians Work Session on Statistical Data Editing (Budapest, Hungary, 14-16 September 2015) Topic (i): *Selective and macro editing Changes in macro-editing and score functions for Dutch STS*. Jeffrey Hoogland, Statistics Netherlands, Netherlands (https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2015/mtg1/WP_20_new_Netherlands_Changes_in_macro-editing_and_score_functions_.pdf)

BLAISE by Statistics Netherlands SURVEY SOFTWARE FOR PROFESSIONALS