

Interactive visualisation tools for exploratory data analysis and policy monitoring

Hans-Joachim Mittag
European Commission, Eurostat, Luxembourg
joachim.mittag@cec.eu.int

Growing demand for user-friendly presentation of statistical information

Policy on national and supranational level is nowadays decisively backed by statistical information. Statistical data, in particular data aggregated to indicators, have an importance for evidence-based policy-making that they never had before. They are used to develop strategies and action plans as well as to monitor and to assess the performance of policy in a variety of areas, such as labour market policies, the fight against poverty and social exclusion, the promotion of professional advancement and equal opportunities or the support of sustainable economic growth. The existing huge demand for policy-relevant and reliable statistics is accompanied by a growing demand for easily accessible and user-friendly environments for the presentation and dissemination of data and policy performance measures.

Today's advanced state of information technology has opened up new opportunities to meet these challenges and to communicate statistical information in a way being easily understood not only by experts. User-friendly, interactive visualisation tools are meanwhile available for presenting and communicating official statistics, for example data from statistical surveys or time series on key indicators. The following graphs demonstrate this statement by means of an interactive Java-based visualisation tool developed in connection with a German multimedia project for statistics education (see [1] and [2]). The tool aims at communicating the messages behind statistical information at first glance by means of user-controlled graphs. The user is enabled to immediately explore data without any programming or downloading. The graphs show how the tool can be applied on the presentation of Eurostat's data on structural key indicators "total unemployment rates" and "unemployment rates for women / men".

Figure 1: Use of time series graphs for displaying national unemployment rates

a. Unemployment rates for user-selected countries

b. Unemployment rate with breakdown by sex for one country

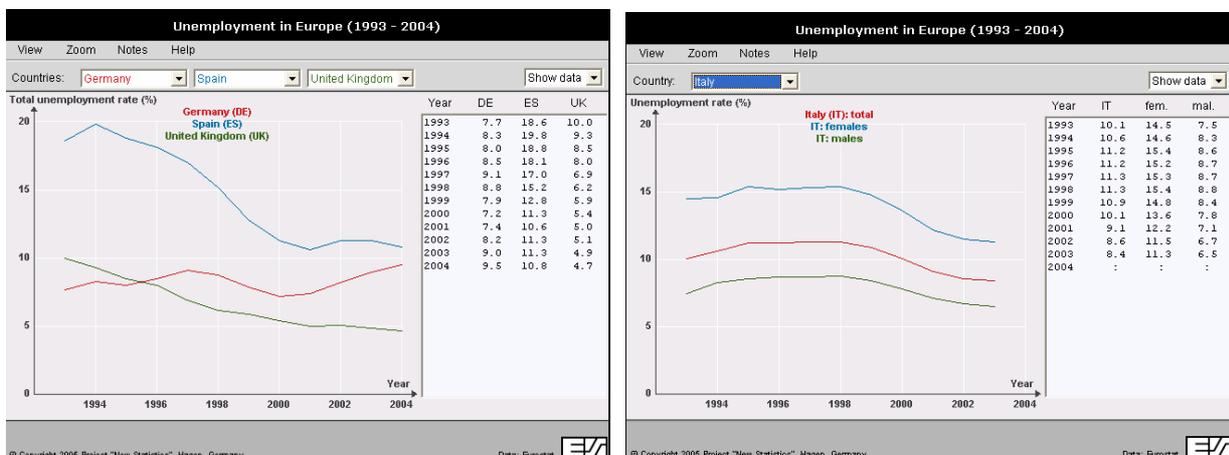


Figure 1a displays time series graphs related to the period 1993 – 2004 for total unemployment in three user-selected countries (European countries and, for benchmarking purposes, the US and Japan) as well as aggregates of European countries. The data itself can be either displayed, as

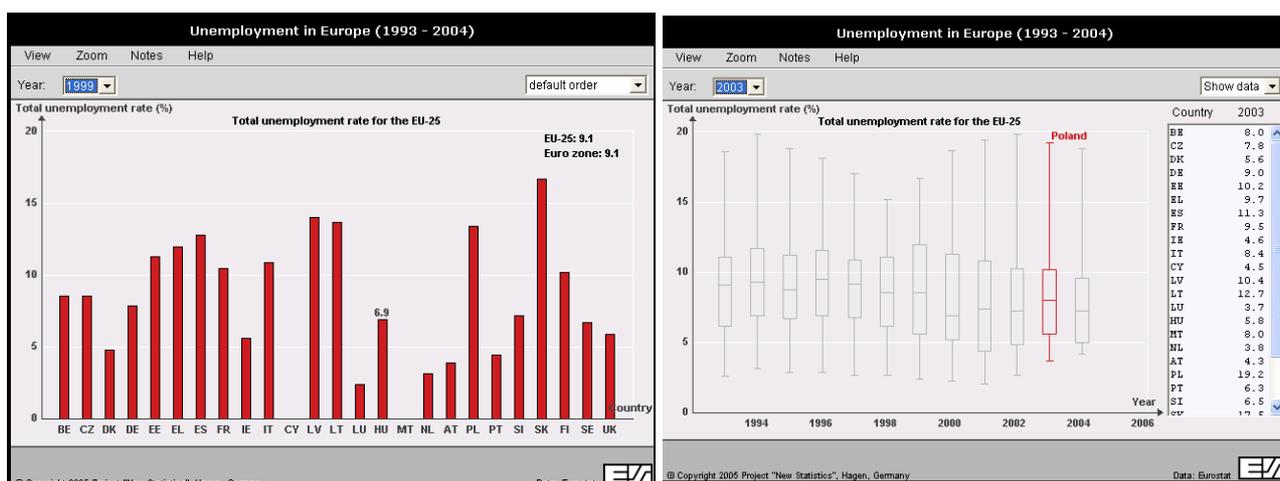
illustrated in Figure 1, or suppressed, depending on the user's choice. The visualisation tool is self-contained with built-in meta data and methodological comments (accessible via "Notes") and a help function with hints for optimal use.

The tool also offers, via a "View" function, the option of an ad-hoc exploration of a given data set aiming at discovering messages behind the data. This exploratory data analysis is performed by viewing the data set from different perspectives and by applying different graphical tools. Figure 2a shows data on total employment in the Member States for a user-selected year, graphically presented in the form of a bar chart. Furthermore, the bars can be presented in ascending or descending order, depending on the user-defined setting in a menu window. The numerical value defining the height of a bar can be displayed by locating the mouse on the top of the corresponding bar. Figure 2b shows boxplots visualising the variation of unemployment rates within the EU-25 at fixed points of time. The user is also enabled to choose one year for which the corresponding unemployment rates defining the boxplot are displayed in tabular form. The countries providing the extreme values of a set of unemployment rates for a specific year are identifiable by putting the mouse on the upper or lower part of the boxplot associated with this year.

Figure 2: Use of bar charts and boxplots for displaying European unemployment rates

a. Total unemployment rates for a user-selected year

b. Variation of unemployment rates between countries over time



Outlook on perspectives for the further development and use of visualisation tools

The visualisation tool presented in Figure 1 – 2 only represents a first prototype. Similar prototypes, most of them with incorporated language change option, have been developed on an experimental basis for presenting data from official statistics. They are meanwhile already publicly accessible (see [3]). The Eurostat Yearbook on Regions 2003 contained one of these interactive environments for exploratory data analysis on a CD-ROM complementing the printed publication (cf. [4]). Eurostat will make further official use of interactive tools in June 2005 for disseminating core results of the latest Structure of Earnings Survey.

In the future, the existing prototypes can be easily modified or complemented by additional functionalities such as interactive maps, depending on the user's needs. They can be applied offline, for example on CD-ROM / DVD or as part of a dynamic PowerPoint presentation. They may be likewise used online on the Web site of statistical offices or national governments or disseminated to press agencies as an e-mail attachment. Furthermore, they can be incorporated into electronic publications in order to take advantage of the opportunities offered by interactivity for online publications, which as yet has scarcely been exploited (see [5]).

Another challenge will be the semi-automatic or fully automatic update of the statistical information displayed via interactive visualisation tools. Semi-automatic update refers to a design where the user only modifies a data set presented, for example, in form of an Excel file. The latter is connected with the visualisation tool which is automatically updated after any manual changes of the Excel file. Full automatic update requires the development of an interface between a large database, for example Eurostat's database for structural indicators. Any update of the database would without any manual intervention automatically update the data to be displayed by the interactive tool.

Acknowledgement and note

The author is indebted to Ulrich Marty, University of Hagen, Germany, for technical support. The views expressed in this paper are those of the author and may not be considered as stating an official position of any institution.

Web links and other references

- [1] <http://www.fernuni-hagen.de/newstatistics/>
- [2] <http://neuestatistik.de/english.php?HP=4>
- [3] <http://forum.europa.eu.int/irc/dsis/wages/info/data/index.htm>
- [4] Regions: Statistical yearbook 2003, Office for Official Publications of the European Communities, Luxembourg, 2003. .
- [5] Mittag, H. -J. (2004). Dynamic data presentation an interactive publication for official statistics. Proceedings of the European Conference on Quality and Methodology; Mainz, Germany, 24 – 26 May 2004.