

THE USE OF ADVANCED VISUALISATION TOOLS FOR COMMUNICATING EUROPEAN DATA ON EARNINGS TO THE CITIZEN

Werner Grünewald and Hans-Joachim Mittag
European Commission, Luxembourg
joachim.mittag@cec.eu.int

This paper illustrates by means of an important European survey on earnings how user-friendly interactive visualisation tools can be applied to communicate results of official statistics and to connect official statistics to the world of statistics education. The visualisation tools presented are self-contained with built-in methodological comments. They can be used offline on a CD-ROM or as part of dynamic PowerPoint presentations. They might be likewise used online on the Web sites of statistical offices, possibly embedded in virtual libraries, or as an integral part of electronic publications.

IMPACT OF TECHNOLOGY ON DATA DISSEMINATION IN OFFICIAL STATISTICS

Eurostat, the Statistical Office of the European Communities, collates and disseminates harmonized European data supporting decision making in policy, economy and other areas directly concerning the daily life of European citizens. An example used here for illustrative purposes is the Structure of Earnings Survey (SES) with its latest results referring to the year 2002. This survey provided detailed information for about 8 million employees on earnings, individual characteristics of the employee and the unit in which they work. The omnipresent news media and the public at large are highly interested in getting the main messages behind the data.

The progress in informatics has opened up new dimensions for communicating official data. The Internet enables the statistical offices to directly and efficiently address the data users. It is not surprising that the recent developments in computer technologies have raised citizens' expectations as regards to the accessibility and comprehensibility of data. The news media tend to accelerate this process by commenting and debating on key statistics and putting the work of statistical offices under a continuous spotlight (Cook and Cowan, 2005). Statistical institutes react on the increasing public interest in their work by adopting a policy of free web-based data dissemination.

The immediacy of access to official data attracts additional users, increases the visibility of the data producers and reinforces the heterogeneity of the data user community. About 25 % of the hits counted for the website of the UK Office for National Statistics are currently linked to school pupils. Eurostat switched in October 2004 to a free dissemination policy and the number of visitors increased almost sixfold during the following first 6 months. These developments require new dissemination strategies which take the needs of new customer groups into account, in particular those of the non-professional users of official data. The latter group covers students and pupils, teachers in any type of educational institution, journalists and the educated citizen in general (Zigure, 2005). In order to successfully communicate with this increasingly important user group, statistical offices need to go far beyond the simple provision of static information and make use of flexible and user-friendly visualisation tools.

Hence, there is still considerable scope for improvements in the accessibility and clarity of statistical information, which are core elements of data quality. A few national statistical offices, such as Statistics New Zealand or Statistics Canada, have opened new channels for the dissemination of statistical information and also involved the educational world in their strategies aiming at an increased use and a better understanding of official data in society. This approach contributes to promoting statistical literacy and to raising the profile of the statistical offices as a key player in national life.

The understanding of statistical information will be fostered by communicating data via well-designed graphs located in the top layer of information, possibly embedded in a story telling environment, and to locate the data itself and the corresponding metadata in a second layer. Well-prepared graphs are much more likely to attract users than mere statistical tables. Interactive

graphs are particularly suitable tools, meeting the needs of non-experts by offering more possibilities for ad-hoc explorations.

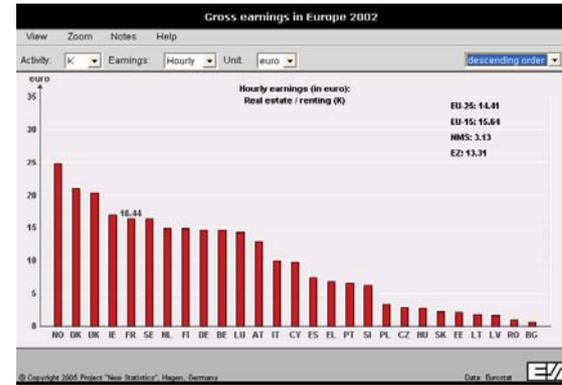
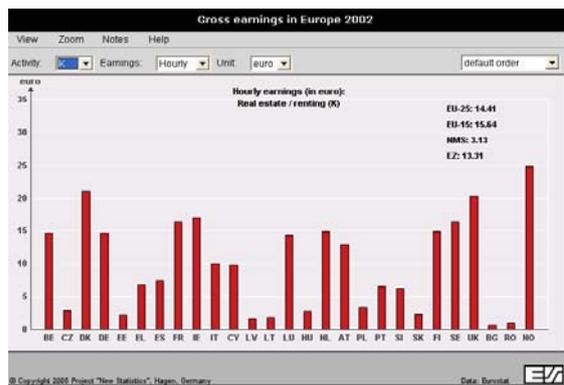
USER-CONTROLLED EXPLORATION OF EUROPEAN DATA ON EARNINGS

Figure 1 shows a Java-based interactive visualisation tool presenting data on gross earnings from the SES 2002. The user can immediately explore the data without any programming or downloading. The data subset to be graphically displayed can be selected via predefined menus. Methodological comments are laid out ready in the background.

Part a of Figure 1 displays a bar chart referring to gross hourly earnings for a user-selected branch. The user may re-order the countries and show the numerical value via the mouse as illustrated in Figure 1b (see the numerical value on top of the bar for France). Furthermore the user is able to graphically display the same data in purchasing power standards (PPS) or to slot in the country code as shown in Figure 1c for Estonia. Finally, as illustrated in Figure 1d, the user has the option to choose another branch or to switch from hourly to monthly or annual earnings. Snapshots of any user-generated graph could be saved, for example by using the “Paint” functionality of Windows or a specialized screenshot capturing program, and exporting to other working environments.

a. Hourly earnings for “Real estate, renting, business activities” (in euro), countries in default order

b. Hourly earnings for “Real estate, renting, business activities” (in euro), countries in descending order



c. Hourly earnings for “Real estate, renting, business activities” (in PPS), countries in descending order

d. Hourly earnings in “Financial intermediation” (in euro), countries in descending order

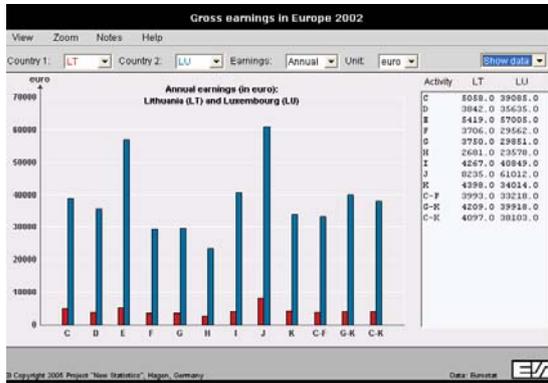


Figure 1: Gross hourly earnings in European countries

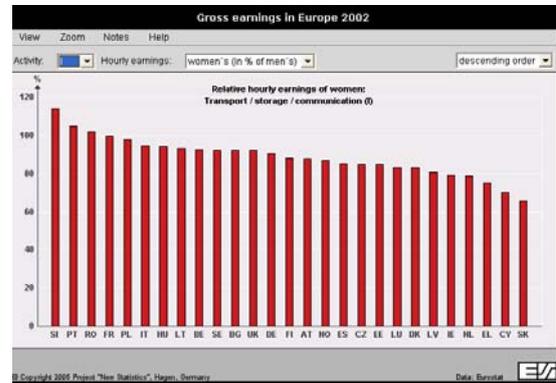
In addition to the options accessible via menu windows, the tool offers a view of the data from different perspectives. The user may, as shown in parts a and b of Figure 2, compare earnings in euro or PPS for self-selected countries with respect to several branches or compare earnings of women and men for a selected industry. Parts c and d illustrate possibilities to also apply other graphical tools such as boxplots. Figure 2c provides a comparison of the variability of

earnings between different branches (within-country dispersion) whereas Figure 2d gives information on the variability between countries for a fixed industry (within-branches dispersion).

a. Annual earnings in Lithuania and Luxembourg for different branches (in euro)



b. Hourly earnings of women in “Electricity, gas and water supply,” in percent of men’s earnings



c. Dispersion of annual earnings (in euro) between branches



d. Dispersion of hourly earnings (in euro) between countries

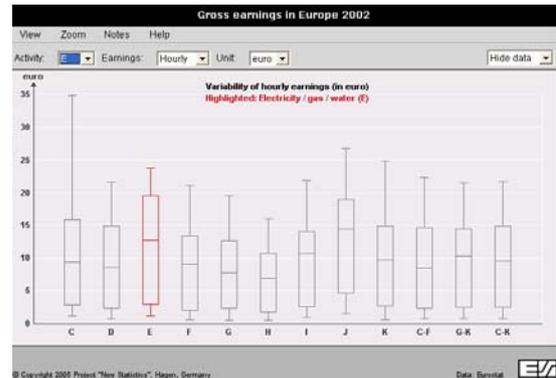
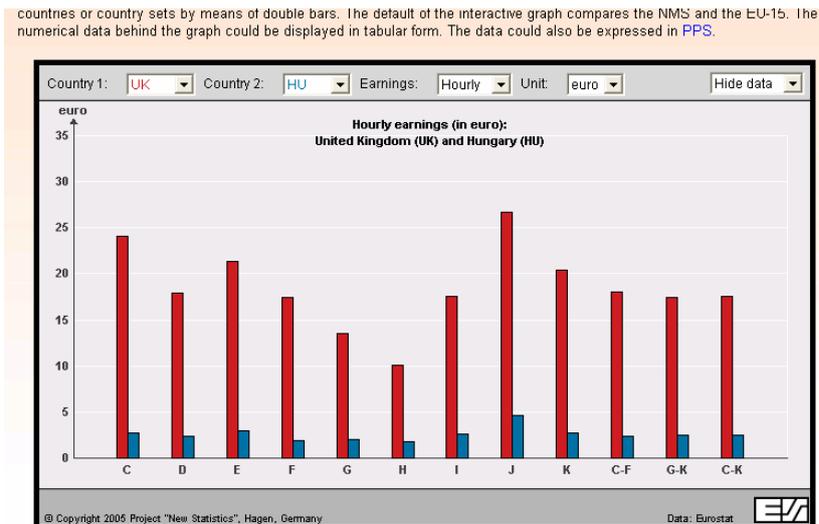


Figure 2: Further user-defined views on European gross earnings data

NEW CHALLENGES FOR ONLINE PUBLICATIONS

User-controlled visualisation tools might be easily incorporated into online publications. Enriching static texts by suitable hyperlinks and dynamic graphs provides sophisticated interactive publications with obvious added value compared to their traditional pdf-type counterparts. Figure 3 shows a screen page presenting an excerpt from an experimental online publication on gross earnings in Europe (Mittag and Marty, 2005, accessible via the Web site <http://forum.europa.eu.int/irc/dsis/wages/info/data/interactive.htm>). The interactive graph gives immediate access to a multitude of different views on earnings data by offering bar charts not only for any combination of countries or country sets but also for hourly, monthly or annual earnings which can be expressed either in euro or in PPS. Furthermore, all graphs could be displayed without or, as carried out in Figure 2a for annual data, with data in tabular form. The interactivity hence opens a completely new dimension for online publishing by drastically increasing the number of graphs which are potentially available for the user.



Suggestions for data exploration:

- Display the numerical values defining the double bars by changing "Hide data" to "Display data".
- Change the defaults for the countries or country sets by means of the menus "Country 1" and "Country 2".
- Change the content of the menu "unit" from euro to PPS.

Figure 3: Interactive graph embedded in a hypertext environment

The text environment is a hypertext which, in this case, provides access to explanatory notes or more extensive statistical tables. The interactive graph could be complemented by a menu containing different suggestions for ad-hoc data exploration. When choosing such a menu option, the user might get a predefined feedback drawing attention to possible observations related to the specific user-selected graphical presentation. The design and implementation of such a tailor-made feedback for a huge number of different combinations of parameters (setting for countries, branches or length of payment periods) defining a specific graph, is not an easy task. This requires an automated or at least semi-automated solution which does not yet exist.

CONNECTING OFFICIAL STATISTICS TO THE EDUCATIONAL WORLD

The ideas presented so far are applicable, offline as well as online, to any area of official statistics. A straightforward application is the deployment of dynamic data visualisation during press conferences, seminars or workshops as integral part of PowerPoint presentations or their use on a CD-ROM. The CD-ROM complementing Eurostat's Yearbook on Regions 2003 contained for the first time an interactive element dealing with data on European capital-city regions (Eurostat, 2003).

A clearly more far-reaching approach is to implement a web-based virtual library containing a collection of independent interactive elements, possibly multilingual and linked to a continuously updated database. Each element visualizes another interesting data set, with meta information on methods and explanatory comments in the background. A nucleus for such a repository is already available at <http://forum.europa.eu.int/irc/dsis/wages/info/data/index.htm>. The self-contained components of the repository can be used by different audiences such as statistical offices, educational institutions or interested citizens. The library can be employed in any educational environment for illustrating the importance of statistics for evidence-based decision making. In addition, the repository can help the non-professional user of official data to understand the world around him or her.

Interactive visualisation has the potential not only to improve the comprehensibility of statistical *data* but also to improve the understanding of statistical *methods* in statistics education. Interactive experiments and user-controlled simulations dealing with basic statistical concepts covering descriptive statistics, probability theory and inferential statistics are valuable as a supplement to traditional lecturing or for self-study purposes. Furthermore, they play a key role in meeting today's growing demand for customized vocational training and life-long learning, in particular for statistical training of employees working in statistical offices. Hence, repositories

containing interactive elements covering the visualization of statistical data and statistical methods help to tie together the still poorly connected worlds of official and educational statistics. A first virtual library containing about 70 interactive elements visualising data from official statistics and methods taught in an introductory statistics course has recently been developed within the framework of a major German multimedia project entitled “New Statistics” for which details are accessible via <http://www.fernuni-hagen.de/newstatistics> (see also Yamaguchi and Mittag, 2004, and Watanabe, 2004).

OUTLOOK

The use of visualisation tools like those presented in this paper is just in its infancy. By far too many institutions dealing with official data do not yet know of, let alone widely use, these or related tools for a better communication to the non-professional user and citizens in general. There is a long way to go and much more effort is needed to allow such tools to become standard in the dissemination process of official statistics.

There are still a few problems to solve. Some have already been raised above, such as the automated production of tailor-made explanatory texts providing guidance for the use of interactive graphs. But there are also further options and opportunities which deserve additional consideration and effort. Classical presentations, for example in printed publications, by definition suffer from print’s static nature. Once the author has decided on the (usually extremely) small data subset to be presented in a (hopefully interesting) graph or map, no further changes or timely updates are possible after starting the production process, at least not at reasonable cost. The situation is different with the online use of flexible visualisation tools. Repetitive data sets such as monthly or quarterly data can be easily updated by just uploading the latest data into the visualisation tools. This flexibility would be particularly welcomed by on-line journals and on-line newspapers. A further application which has been mentioned so far only as a side aspect are interactive maps. Spatial data are often complemented by static maps, i.e., data sets with geographical breakdowns are displayed not only in tabular form or via graphs but also in connection with traditional maps. But the map can also be the starting-point for data exploration. The user may be offered a map and, by clicking on a particular region, the corresponding figures of the statistics under consideration are shown. Furthermore, advanced data visualisation tools for official statistics can incorporate key information on data quality.

All in all, one may conclude that the visualisation tools described in this paper offer a unique way of considerably improving the presentation of statistical information though their use is by far not so widespread as would be desirable. The academic world as well as official statistics are jointly invited to strengthen their efforts to further develop these tools and their use.

ACKNOWLEDGEMENT AND NOTE

The authors are indebted to Ulrich Marty, University of Hagen, Germany, for technical support. The views expressed in this paper are those of the authors and do not necessarily reflect the opinion of the European Commission.

REFERENCES

- Cook, L. and Cowan, C. (2005). The challenge of communicating statistics. Paper presented at the 91st Conference of the Directors-General of the National Statistical Offices (DGINS Conference), Copenhagen, 26-27 May 2005. In *Conference Proceedings*, (pp. 143 – 163). Luxembourg: Eurostat / Office for Official Publications of the European Communities.
- Eurostat. (2003). *Regions: Statistical Yearbook 2003*. Luxembourg: Office for Official Publications of the European Communities.
- Mittag, H.-J. and Marty, U. (2005). Gross earnings in Europe 2002. Interactive version of Eurostat’s *Statistics in Focus* edition 12/2005.
- Watanabe, M. (2004). The “Multimedialization” in statistics education. *Statistics, Journal of the Japan Statistical Society*, 9, 14-22 (in Japanese).
- Yamaguchi, K. and Mittag, H.-J. (2004). The use of multimedia for teaching statistics, *Statistics, Journal of the Japan Statistical Society*, 9, 8-13 (in Japanese).

Zigure, A. (2005). Focusing on non-professional users. Paper presented at the 91st Conference of the Directors-General of the National Statistical Offices (DGINS Conference), Copenhagen, 26-27 May 2005. In *Conference Proceedings*, (pp. 93 – 105). Luxembourg, Eurostat / Office for Official Publications of the European Communities.