HUNGARIAN VISUALIZATION TOOLS ON OFFICIAL SOCIAL-ECONOMIC DATA IN CLASSES

Peter Kovacs
Faculty of Economics and Business Administration, University of Szeged, Hungary
kovacs.peter@eco.u-szeged.hu

Eva Kriszt Sandorne
Budapest Business School

Considering the dependence of Generation Y on IT tools, databases and visualization play a highly important role in tertiary education. In this paper, we will describe and analyze the applicability of new interactive visualization tools (e.g. population pyramid, maps) and datasets of the Hungarian Statistical Office in classes. The applicability of visualization tools will be demonstrated in classes with examples. For instance: 1. Will there still be a state pension for us? We will give a potential answer to this question with the help of an interactive population pyramid. 2. The Hungarian poverty rate is lower than the poverty rate of the EU-28. What does it mean? We will discuss how sensitive data is depending on the definitions of the parameters in question. The conclusions of this examination will constitute parts of lesson plans suitable to promote the importance of visualization and multivariate approaches in Hungary.

INTRODUCTION

The multivariate approach is crucial to understand the phenomena of real life and to interpret the relationships among social and business data correctly. The huge amounts of data, data sources, and visualization tools (for instance Gapminder, OECD, Eurostat, national statistical agencies) on the Internet provide an opportunity to illustrate complex relations with real data relatively easily (Forbes et al. 2014, Hammerman and Rubin 2004). At the same time, the misuse of these tools can lead to misinterpretations. Today's university students are members of the so-called Generation Y, who share special characteristics (Wolberg and Pokrywczynski, 2001) and can be efficiently taught with different methods (Nimon, 2007). IT tools are familiar to this generation, thus the usage of IT tools, the Internet and visualization in class would be effective (Garfield and Ben-Zvi 2007). Visualization tools showing real data could develop the students' complex way of understanding, together with their analytical skills, statistical reasoning and thinking (Chance et al. 2007). Consequently, building these tools into the curricula is essential. Ridgway's (2015) recommendations to revise statistical curricula also include:

- Working with multiple data sources,
- Illustrating and using real problems with real data,
- Using Internet resources,
- Teaching about and with interactive graphics, and familiar technologies,
- Introducing multivariate data at an early stage.

In the case of social and economic issues, it would be practical first to explore our environment, region, and country. As a starting point, national statistical offices provide useful services, data, and graphical tools. The main goal of this paper is to present some visualization tools of the Hungarian Central Statistical Office (HCSO) and demonstrate their applicability for teaching.

In Hungary, business and economic training programs include 2 semester-long introductory statistics courses covering the following main topics: descriptive statistics, theory of indexes, sampling, point and interval estimation, hypothesis testing, and regression models.
At the University of Szeged, Faculty of Economics and Business Administration we use official statistics data sources, Excel functions and Pivot tables, SPSS in the classes both at bachelor and master level. The statistical courses are made up of four hours of teaching a week divided into a 2 hour lecture part for 500 students and a 2 hour computer assisted seminar for 30 students. During seminars we solve both computer and paper-and-pencil based tasks.

In: J. Engel (Ed.), *Promoting understanding of statistics about society.*
Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE),
July 2016, Berlin, Germany. iase-web.org/Conference_Proceedings.php

HUNGARIAN OFFICIAL DATA IN CLASSES TABLES

The broadest collection of the major social and economics official data and indicators is available from the HCSO's site (http://www.ksh.hu). At the beginning of the introductory statistics courses, students meet the HCSO's portal and the use of data sources. At this point the main goal of the in-class exercises is to illustrate and teach the use of data sources and metadata.

STADAT

STADAT is the first data source on the HCSO's website used by the students. It is a free downloadable static table system of Hungarian official statistical data. The data and tables are grouped into main areas:

- Population, vital events
- Society
- General economic indicators
- Sectors of economy
- Environment
- Territorial data
- International comparison

Each area contains annual, monthly/quarterly and long term time series data in separated static tables, which means that the structure of the tables (indicators, grouping variables, rows, columns) are given. Searching for data does not require a high level of statistical knowledge and IT skills, only the searched indicator or the topic should be known. Tables are downloadable in Excel format. In STADAT, it is not possible to generate charts online, but we can access online visualization tools on the HCSO's portal from the tables with links or we can download and visualize the data in our own way.

In order to interpret and display statistical data, it is very important to know the definition of the indicator, and something about the methodological background depending on the target of the application. For a better understanding, the *metainformation system* is available in each table: it contains the definitions of concepts (glossary, changes), the data sources (registers, surveys, sources, data collection, classifications) used in a statistical topic, the applied methods, and the data quality aspects.

In STADAT, both territorial and international data (Hungarian translation of Eurostat data) are available, thus international comparisons can be done with the help of the metadata.

First we ask students about the annual value of the unemployment rate in Hungary and we try to find the exact data together. After that, we discuss the meaning of the data. The introductory statistical courses are scheduled in the third and fourth semesters of the training program, when the students have already accomplished micro- and macroeconomics, thus the students have learnt about the economic definitions of the unemployment rate and unemployment people, so they have enough content knowledge to do this exercise. With this example we can illustrate the different approaches: the ILO's and the national definitions of an unemployment person.

The second exercise is to find the county data of the unemployment rate independently.

The third exercise is to find and download the populations of European countries in Excel format, after that the students should prepare a data table (they have to erase the unnecessary data) and a bar chart in Excel.

Our experience is that using STADAT is not problematic for students. The challenge for them is the preparation of tables and charts in Excel, because several students are afraid of using and applying Excel.

Let us consider another example on demonstrating the importance of discovering metadata in classes. First we ask students about the percent of Hungarians living in poverty, the answers spread between 30% and 70%. Then we check the actual data with STADAT. The Hungarian poverty rate (14.6%) is lower than the previous guesses and also lower than the poverty rate of the

EU-28 (17.2%). These data come as a surprise for students at first time, so the next question is the definition of poverty. This example illustrates the sensitivity of data to definitions and methods. The value of a poverty indicator is based on the following poverty definitions: the metasystem contains the concept of quasi absolute poverty rate, persistent poverty rate, relative poverty rate, subjective poverty, and temporary poverty.

The exact name of the mentioned indicator is relative poverty rate. It is defined by Eurostat as the percentage of people with an equalized disposable income below 60% of the national median equalized disposable income, so it measures only monetary poverty. Furthermore, each country has different threshold values. The value of the indicator is different in the case of different household sizes. Furthermore, if we calculate the relative poverty rate before social transfers, retirement and survivor's pensions, the Hungarian data is 26.3%. The consequence of this task is that poverty is a complex social phenomenon; consequently, one indicator is not enough to describe it.

Population pyramid

In the past few years, more and more interactive charts and maps have been developed on the HCSO's site. These applications use only official statistical data and we do not have the opportunity to upload our data. One of them is the *interactive population pyramid* showing the age structure of the population and the annual changes of the structure between 1870 and 2060 (see Figure 1). The application uses census data between 1870 and 1949, yearbook data between 1950 and 2015, and a forecasting of the Hungarian Central Statistical Office – Demographic Research Institute for the period between 2016 and 2060.

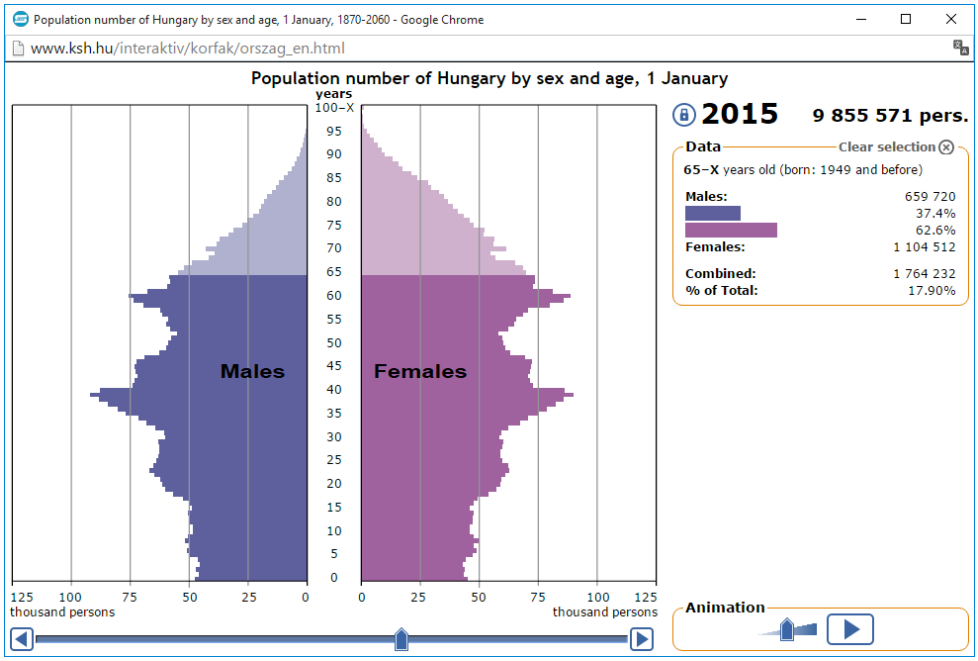

Figure 1 The age structure of the population, Hungary, 2015

The x axis shows the size of the population in each gender group, and the y axis gives the ages. With the mouse an age or an age bracket can be selected, and the proportion and the structure of the selected group are described on the chart. With the animation button we can see the structural changing process of the population. The shape of the pyramid shows the ageing of the population.Consider an application of the pyramid.

Will there still be a state pension for us? The demographic background of this concern is the ageing problem. First we ask students about the meaning and consequences of an ageing society. After that, we discuss how we can measure it. First, students suggest indicators. The typical answer is the ratio of the number of persons aged 65 and over (age when they become generally economically inactive) and the number of persons aged between 15 and 64. We analyze it

and calculate the ratio of the population aged 15-64 (working) and the population aged above 65- (retired) in 2015 and 2060 (see Figure 2 and Table 1).

Table 1 Ageing of the population in Hungary

Year	*Proportion of 15-64 years old people, %*	*Proportion of 65- years old people, %*	*Ratio of 65- /15-64 age brackets*
2015	67.62	17.90	3.8
2060	56.51	30.13	1.9

According to the forecasting, 45 years later one working age person will have to support twice as many retired people compared to now. The usage of the population pyramid does not require a high level of statistical and content knowledge, nor that of IT skills. By locking a particular year, the structure of two different years could be compared.

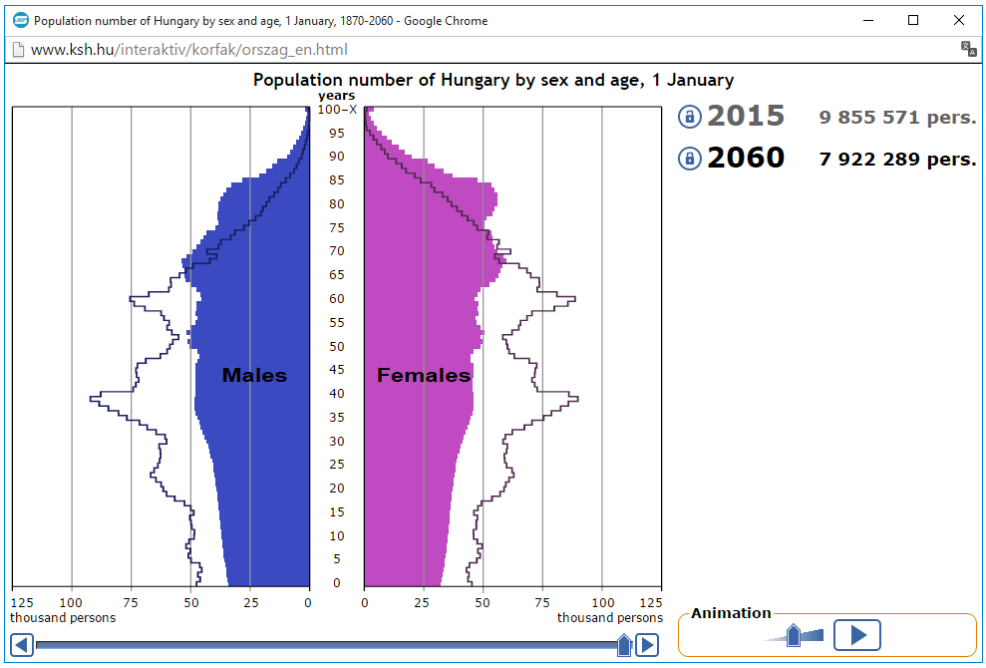

Figure 2 Comparing the age structures of the population, Hungary

Finally we show the official indicators for an ageing society: old age dependency ratio, ageing index, medium age, fertility rate, growth rate of population.

Other sources on HCSO's site

The *dissemination database* is another data handling option on the HCSO's website. This is an interactive tabulation system. Similar to the STADAT, firstly, we need to open a main area of data. In the selected topic we can choose the target indicators, grouping variables, and the structure of the table. The table may be rearranged. Contrary to STADAT, this system can generate online basic downloadable charts (bar, pie, line, etc.) based on our customized table. Using this system requires a higher level of statistical knowledge and IT skills.

The *interactive* Hungarian *maps* visualize 219 Hungarian data on different territorial levels (see Figure 3). The *European maps* application illustrates 273 data of the EU member states. In comparison with the Eurostat maps, the number of the available indicators and the reference time (only 1 year) is much less in the Hungarian system. Furthermore, the Hungarian system displays only concrete data, while Eurostat maps can compare country data to a reference territory or to a reference time and show the comparisons on maps. Set it against the Gapminder or OECD maps, the drawback of this service is that we cannot do real time visual analytics. Furthermore, in Gapminder we have the possibility to display the relations among variables at different times.

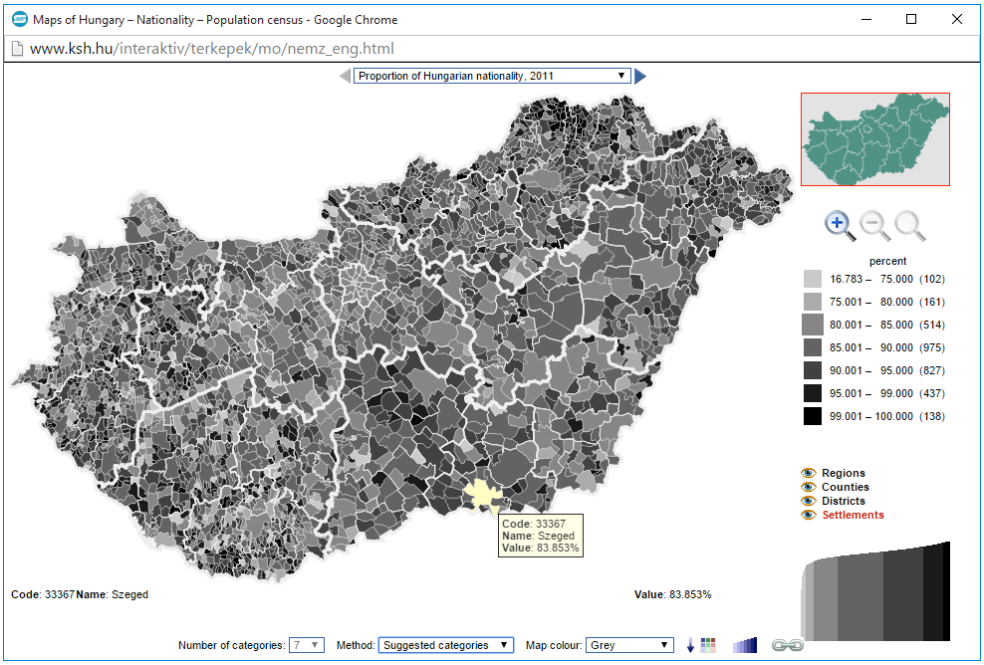

Figure 3: Proportion of Hungarian nationality, 2011 Interactive Hungarian map

The website contains several dynamic *radar charts* (see, for an example, Figure 4), for instance the inflation radar shows consumer yearly price indices by groups of expenditures from January 2000. Each axis measures the relative changes of prices in a group of expenditure. Searching for data does not require a high level of IT skills. On the other hand, understanding the meaning of inflation needs content knowledge, and the analysis of radar charts very often causes a problem for the students. If we combine the inflation radar with the *inflation kaleidoscope* (see Figure 5), which is a tree map, not only the changes of the prices but the structure of this indicator may also be illustrated. The surfaces represent the proportion of the groups within the total expenditures of households.

The radar charts and the population pyramid clearly illustrate multivariate problems.

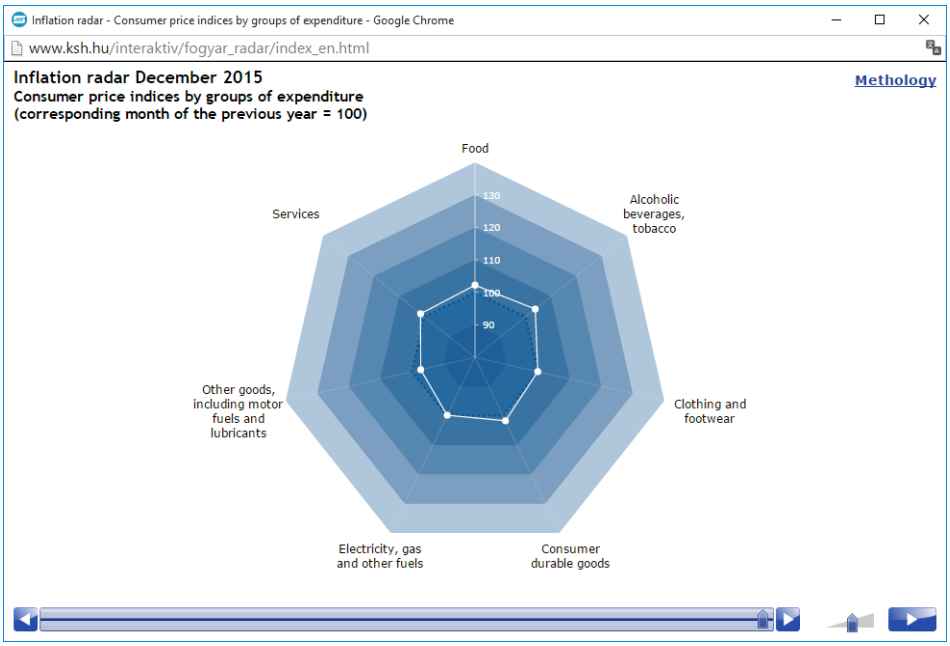

Figure 4: Illustration of Radar chart

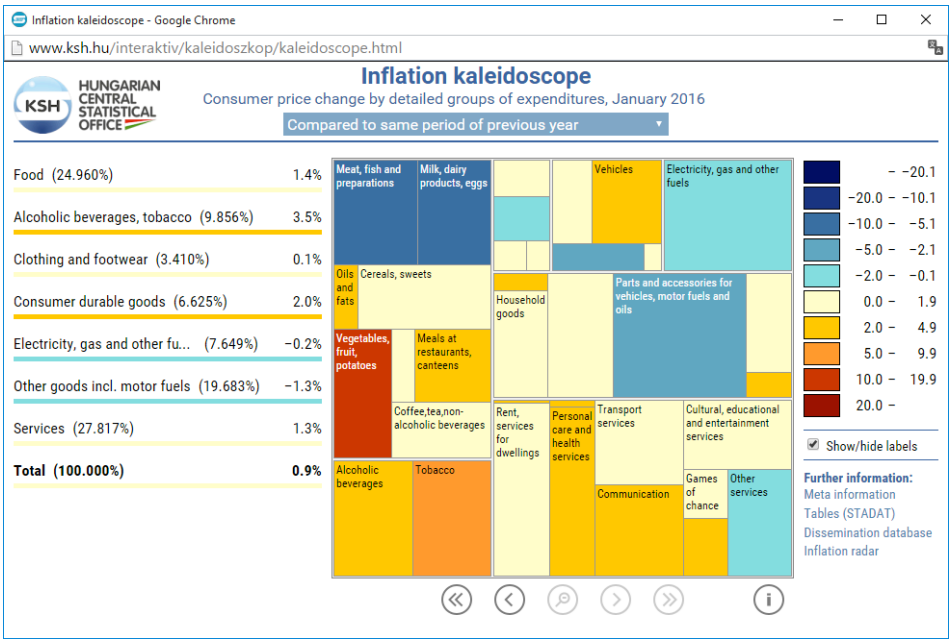

Figure 5: Illustration of inflation

CONCLUSION

The available Hungarian official statistic datasets and interactive visualization tools are suitable to teach them and to use them to illustrate social and economic contents in courses. Detailed metainformation and metadata are available to facilitate the interpretation and the use of the data and charts. The usage of these data sources does not require high IT competencies, but to interpret the result and do international comparison, content knowledge is necessary. Further research could focus on generating more examples applicable in lesson plans and course materials.

ACKNOWLEDGMENT: The work reported in this paper was supported in part by ProCivicStat project, funded by the ERASMUS+ program of the European Commission. However the views and opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007). The Role of Technology in Improving Student Learning of Statistics. *Technology Innovations in Statistics Education*, 1(1) 27.

Forbes, S., Chapman, J., Harraway, J., Stirling D., Wild, C. (2014). Use of Data Visualization in the Teaching of Statistics: A New Zealand Perspective, *Statistics Education Research Journal*, 13(2) 187-201.

Garfield, J. and Ben-Zvi, D. (2007). How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning Statistics, *International Statistical Review*, 75(3), 372–396.

Hammerman, J., and Rubin, A. (2004). Strategies for Managing Statistical Complexity with New Software Tools, *Statistics Education Research Journal*, 3(2) 17-41.

Nimon, S. (2007). Generation Y and Higher Education: The Other Y2K, *Journal of Institutional Research*, 13(1), 24–41.

Ridgway, J (2015). Implications of the Data Revolution for Statistics Education, *International Statistical Review,* doi: 10.1111/insr.12110.

Wolberg, J., Pokrywczynski, J. (2001). A psychographic analysis of generation Y college students, *Journal of Advertising Research*, 41(5), 33–52.