DIDACTICAL ISSUES OF DATA MODELING WITH INTERACTIVE CHARTS

Djordje M. Kadijevich

Institute of Educational Research, Belgrade, Serbia
Email: djkadijevic@ipi.ac.rs

Abstract. This workshop, based on both theoretical considerations and practical work, deals with simple data modeling using pivot charts and dashboards, and the didactical issues of this kind of modeling. The workshop is based upon the author's IASE 2016 Roundtable paper (available at http://iase-web.org/Conference_Proceedings.php), as well as his recent experience in teacher professional development concerning this topic.

Workshop details

Main prerequisites:

- basic knowledge of mathematics and statistics (e.g., absolute and relative frequency, percentage, chart types, the average value);
- basic knowledge related to spreadsheets and relational databases (field i.e. attribute, record–entity, making queries);
- basics of visual programming using the drag & drop approach.

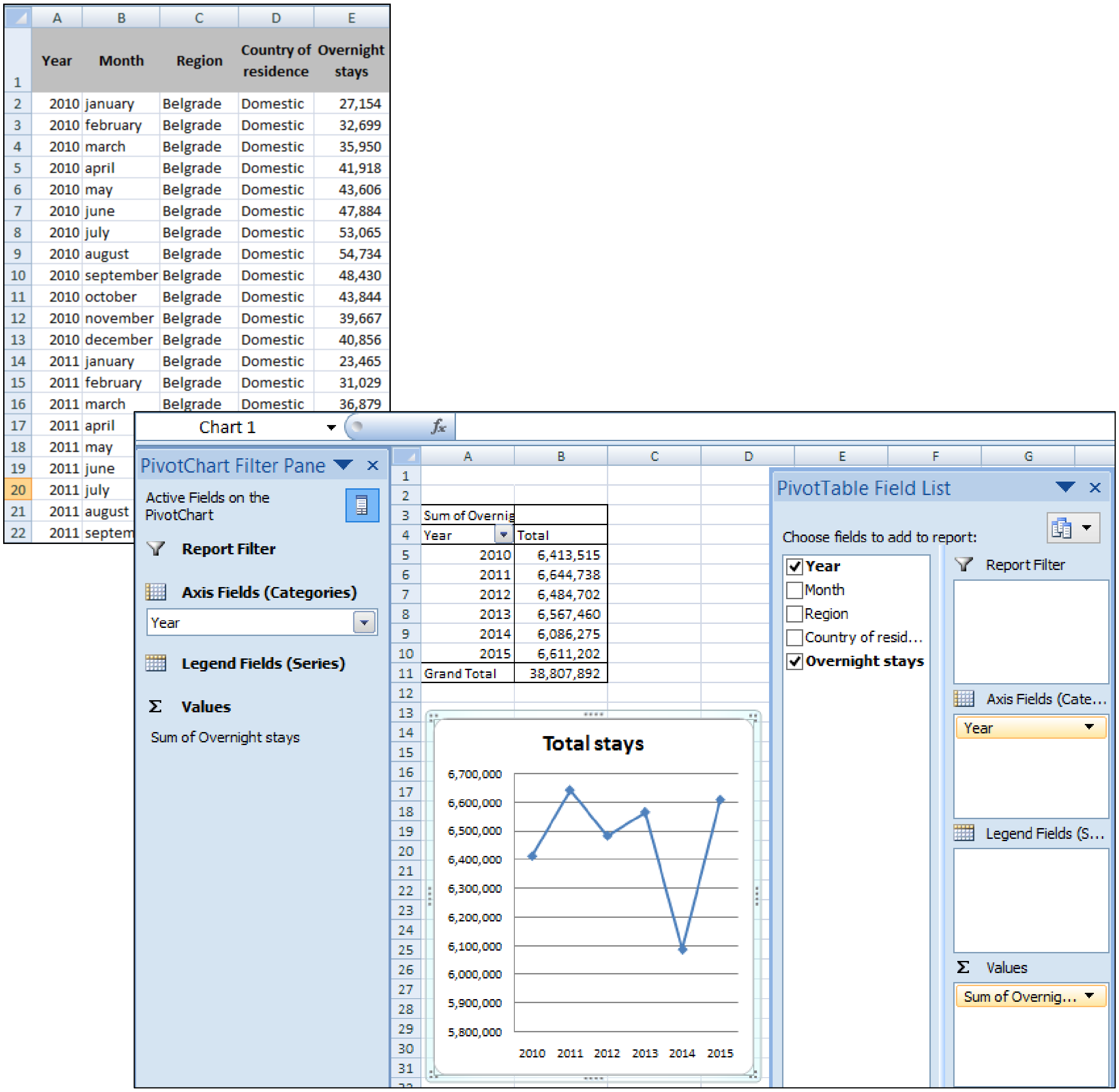
Practice at the workshop: Each workshop participant works on his/her personal computer with a recent version of Microsoft Excel installed. The participant knows how to use this program (at least on an introductory level). In addition (for optimal practice at the workshop), this personal computer has access to the Internet, and the participant has already registered at https://www.zoho.com/reports/dashboard.html to be able to use the ZOHO environment for creating dashboards.

Main outcomes (benefits): The participants improve their knowledge and skills in the following domains:

- the preparation of data for the visualization of relations in those data;
- the visualization of those relations by using one or more charts;
- understanding important statistical concepts, such as size of effect, trend, interaction, and confounding variable;
- applying a methodical approach to the practice of data modeling in mathematics, informatics, and statistics education.

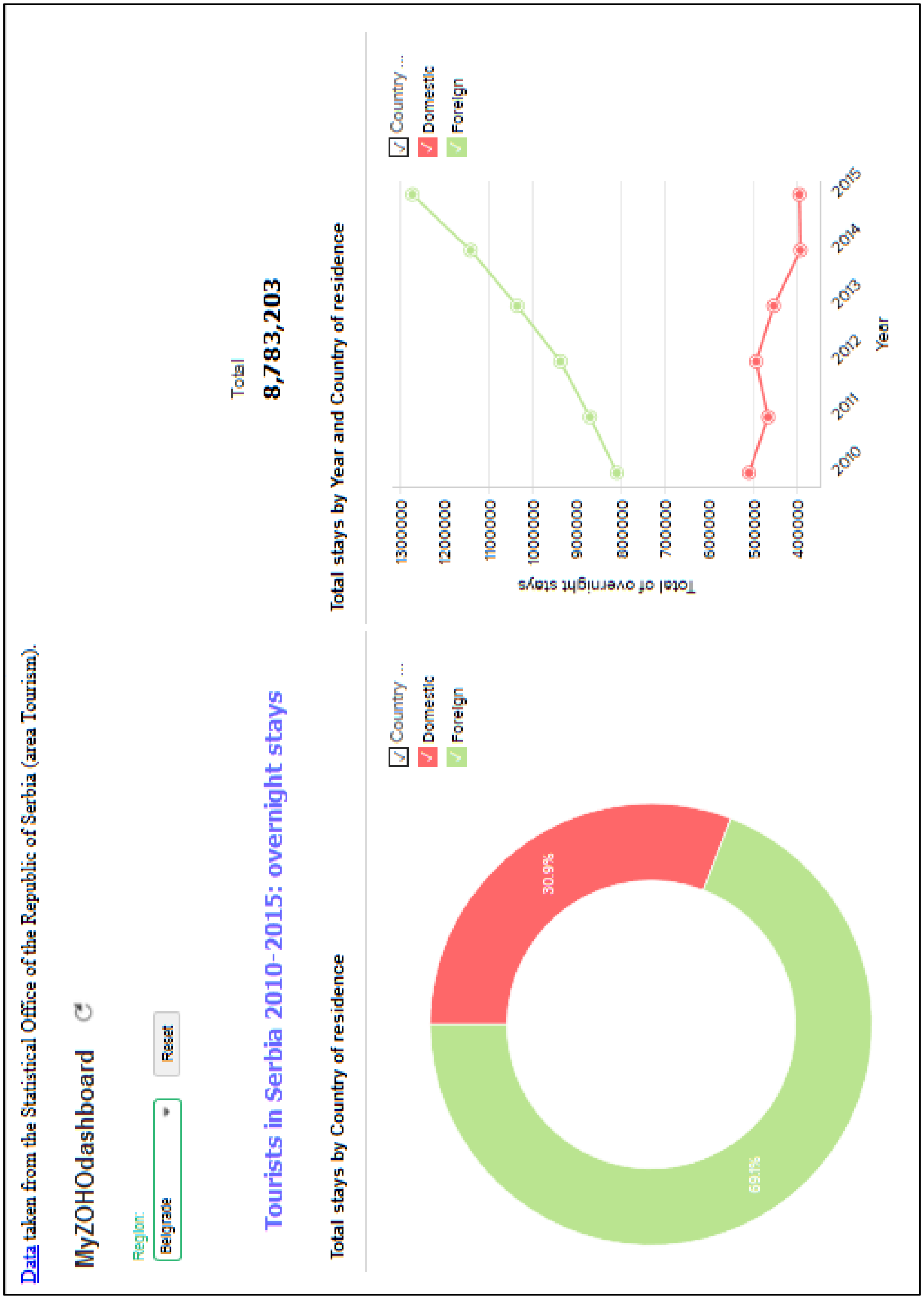
Examples of data modeling

- *Pivot chart*



Internet: www.mi.sanu.ac.rs/~djkadij/Seminar17.xls

- *Dashboard*

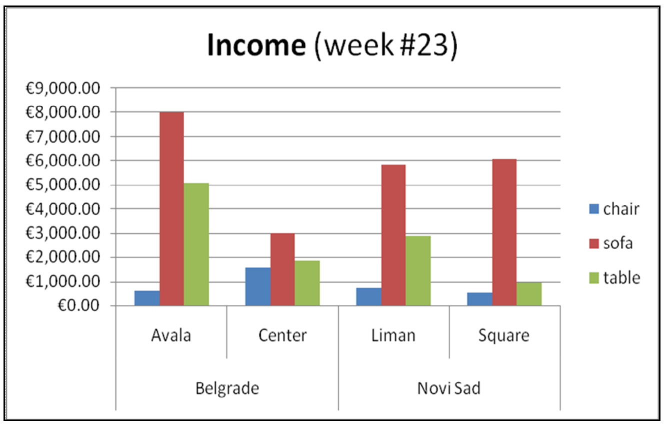


Internet: www.mi.sanu.ac.rs/~djkadij/Dashboard.htm

Pivot chart affordances

Pivot tables and pivot charts are simple tools of business intelligence, which provide summary reports in the form of interactive tables and graphs. These tools allow some summary parameters (e.g., the sum or the average of the values considered) to be shown in relation to different levels of details. The user selects these details with respect to his/her requirements, and, if necessary, further collapses or breaks them down (the so-called roll-up or drill-down views, respectively).

Consider, for example, the following chart regarding the sale of some items.



Is easy to realize that this chart is related to the sales of a number of items, that these items are sold in different stores, and that these stores are located in two cities. Therefore, in addition to the finding of the total income (Level 0), the use of the pivot approach can help its user to analyze the following:

– the total income by city (Level 1 – the first level of details with respect to the categories of one independent variable; also called attribute or indicator), to find out in which city the total income was the highest;

– the total income by city and store (Level 2), to determine in which store and in what city this income was the lowest; and

– the total income by city, store, and item (Level 3), to find out in which store and in what city the total income for a particular item was the lowest.

Participant's self-preparation for the workshop (before it!)
includes the following two tasks:

1. By using a file with some sales data (e.g., www.mi.sanu.ac.rs/~djkadij/LaudonASE1.xls), generate graphs that enable a business analysis like the one given on the previous page. The main question is how to model the realization of this analysis didactically in order to to attain successful students' work on at least two levels of complexity (with simple learning requirements *vs* some more complex ones).
2. By using the visualization titled *Cardiovascular Disease Risk Factors* (available at https://www.dur.ac.uk/smart.centre/freeware/), summarize several findings regarding some, to you important relations in the data analyzed. Through play with this visualization, try to illustrate some of the important statistical notions mentioned above (size of effect, trend, interaction, and confounding variable). To this end, you may use Appendix 2.

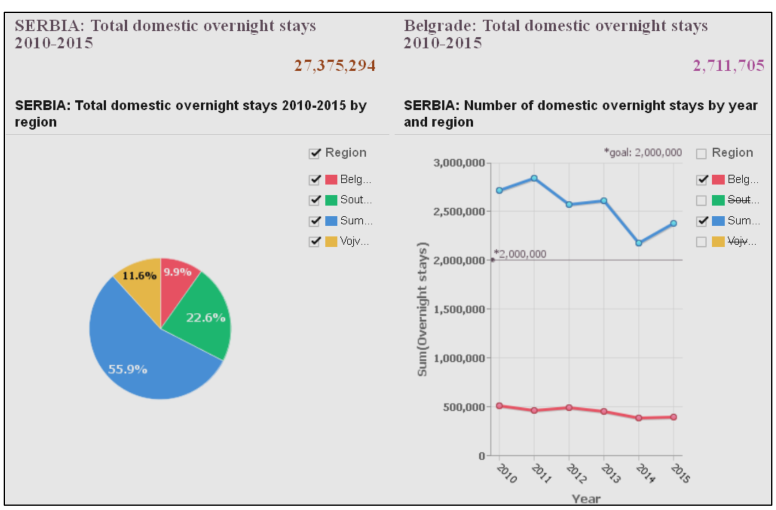
Main task at the workshop

It is planned that each workshop participant, in accordance with a theme chosen by himself/herself, will carry out data modeling by using interactive charts. (If the ZOHO environment is used, helpful instructions for data modeling with dashboards are given in a video available at https://www.youtube.com/watch?feature=player_embedded&v=aPLg4dp-f28). This modeling should be based on a real data set provided by a relevant agency or institution, such as Eurostat, or the World Bank. Particular attention should be paid to the didactical issues of this modeling at a specific educational level (e.g. in a grade 8 mathematics class), to enable successful students' work on at least two levels of complexity (with simple learning requirements *vs* more complex ones).

Appendix 1 – Topic relevance and its educational value

The science of data has in some ways been present in computing for more than a half a century (please recall that the term *datalogy* was introduced by the Danish scientist Peter Naur (1928–2016) almost sixty years ago; see https://en.wikipedia.org/wiki/Data_science), and has been a hot topic in recent years. Having in mind widespread demands for modeling huge amounts of data for business, scientific or other purposes (to obtain from them potentially useful information that may lead to new knowledge), it is not surprising that expertise regarding the study of data (i.e. data scientist post) has been increasingly in demand. Although statisticians usually say that the study of data is in fact the statistical analysis of data (i.e. data science "=" data analysis), this is not strictly true. This is because the science of data requires not only solid knowledge and skills in computing (e.g. programming/ databases) and mathematics (the application of a variety of mathematical and statistical models), but also, among other things, a high degree of creative thinking and communication skills. It is thus not surprising that the job of data scientist is highly paid (frequently over $100,000 annually). As some kind of data modeling is likely to be present in the future professional work of most students, there is a growing call to include basic data modeling at earlier educational levels. Please recall that *Data* has been a content area in a series of international assessments in school mathematics and science known under the acronym TIMSS (http://timssandpirls.bc.edu/). TIMSS tasks in this domain usually deal with tabular and graphical representations of data, drawing conclusions from such representations, which are indeed important aspects of elementary data modeling.

In the last ten years, the educational values of data modeling by using interactive charts (with the application of elementary mathematical or statistical models) have been the focus of study among a number of researchers engaged in improving statistics education, particularly those working at the Smart Centre, Durham University, UK (https://www.dur.ac.uk/smart.centre/). The approach of these British researchers was summarized in a recent paper titled *Visualise then Conceptualise* (available at https://www.dur.ac.uk/resources/smart.centre/Publications/Visualisethenconceptualise_SSTVOL40._3.Summer8-13i.pdf), for example. Their research evidences that this kind of data modeling, even when arranged in an informal way, supports the understanding of important statistical notions, such as size of effect, trend, interaction, and confounding variable. Bearing in mind that interactive objects of the dashboard type consists of two or more pivot charts, these interactive objects enable better data modeling than simply relying on single pivot charts. This is, among other things, because the data modeler can compare the outcomes on several charts, as shown below.



Researchers in statistics education are just beginning to study how to use dashboards successfully in teaching. Bearing in mind the educational value of this kind of data modeling, students should be supported in the creation and use of sets of pivot charts whose structural complexity increases gradually. In doing so, a number of challenges should be addressed by the teacher. These challenges, relating to 1) data to analyze, 2) dashboards to create, and 3) modeling to implement, have not been examined so far (cf. Kadijevich's (2016) contribution to the IASE 2016 Roundtable Conference available at http://iase-web.org/Conference_Proceedings.php).

Appendix 2 – Small statistical reminder

- The *size of effect* describes which variable is more related to the target variable. For example, graphs may show that while smoking increases the risk of an infarct about two times, insufficient physical activity increases that risk even 3–4 times.
- *Trend* tells us in what way the values of one variables change over time. A chart may signal that these values increase or decrease over time, and if a trend is present, it may be linear or not.
- The *interaction* of independent variables exists when the effect of one of them on the dependent variable differs for different values of other variable. Let us suppose that one study found that contrary to Faculty of Mathematics, there were gender differences in the average duration of graduate study at Faculty of Medicine. In that case, there was an interaction between variables gender and faculty. Interaction would be easily recognized in graphs because the distance between the average values of the dependent variable for the values of one independent variable (e.g., gender) changes by changing the value of other independent variable (e.g., faculty).
- *Confounding variable* is a variable that could affect the relationship between variables in question. For example, students in science departments might differ from students in social science departments with respect to their average weight, but this difference might be caused by the fact that these two groups of students differed in their average age. If this applies, variable age is a confounding variable, and its consideration may reduce or annul these weight differences.