DATA MODELING WITH DASHBOARDS: OPPORTUNITIES AND CHALLENGES

Djordje M. Kadijevich
Institute for Educational Research, Belgrade, Serbia
djkadijevic@ipi.ac.rs

Visualizing data about social phenomena has been supported by many important institutions, such as OECD and the World Bank. Although there are several recent papers on data visualization in statistics education, studies on data modeling by interactive displays are rare, especially when dashboards are used. After an examination of data modeling and dashboards, this paper discusses the main opportunities and challenges of data modeling with dashboards, and gives possible reasons for these challenges. This contribution is based on a review of multidisciplinary literature. A summary of findings relevant to statistics education is given in the closing section.

DATA MODELING: RELEVANCE, DASHBOARD-APPROACH

This paper primarily deals with descriptive, exploratory graphical displays produced by the application of simple mathematical models to examined data (these models usually involve absolute and relative frequencies, counts, sums, and means). To denote examining or analyzing data in this way, we use the term *data modeling*, as, for example, done in English (2012).

Viewed from the intersection of mathematics, statistics, and computer (information) science, such delineated data modeling calls for a complex process that involves various activities concerned with data, such as collecting, querying, auditing, analyzing, structuring, representing, transforming, operating, and interpreting data. This kind of modeling has sporadically been examined in research in mathematics and statistics education. A few recent studies are those of English (2012), Mulligan (2015), McNamara and Hansen (2014), and Ridgway, James and McCusker (2013). Despite that, there is no doubt that data modeling is quite relevant to the development of numeracy in general and statistical literacy in particular (Mulligan, 2015). Regarding statistical literacy (SL), data modeling may contribute to the development of various traditional SL components, such as collecting, representing, and interpreting data; finding patterns and trends in data representations; understanding ideas of range, variation, and distribution; and making predictions (English, 2012; Mulligan, 2015). This modeling may also support the development of some (relatively) ‘new’ SL components, such as examining data provenance and quality; understanding ideas of effect size and interaction associated with the analysis of multivariate data (Ridgway, James & McCusker, 2013). Furthermore, data modeling—a key activity in data science that is of increasing importance in today’s business—can contribute to the development of both computational and statistical thinking (McNamara & Hansen, 2014).

Data modeling may follow the PPDAC inquiry cycle (PPDAC is the acronym from Problem–Plan–Data–Analysis–Conclusion; for this cycle see Wild & Pfannkuch, 1999, and Forbes et al., 2011, for example). Main activities in its stages are: Problem–decide on what to study; Plan–decide on what to measure, collect or query; Data–prepare data to analyze; Analysis–create graphical displays, generate and test hypotheses, find patterns and trends; and Conclusion–make interpretations, generate new ideas. Data modeling may also follow key stages that occur in mathematical modeling, such as a) messy real world problem, b) real world problem statement, c) math model, d) math solution, e) real world meaning of solution, f) evaluation, and g) report (Galbraith & Stillman, 2006). This is because many modeling activities are present in both mathematical modeling and data modeling. For example, clarify problem context and questions, identify strategic entities (transition a → b); identify dependent and independent variables, specify calculations needed (b → c); use technology to perform calculations, and produce charts and tables (c → d); give meaning to math results, select arguments to justify interpretations (d → e); and question model adequacy, examine implications of results (e → f).

To enable the viewing (and modeling) of multivariate phenomena in a way that is less complex and more dynamic when compared to the traditional methods used in statistics classes, students should be provided with suitable technology-based tools that are easy to use, such as interactive displays. Two recent studies on computer-based data modeling mentioned above used

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

such displays. Ridgway, James and McCusker (2013) used pivot charts (a type of charts whose contents are defined by selecting fields in a drag-and-drop fashion) with some summary measures. McNamara and Hansen (2014) used dashboards. A dashboard is simply a collection of two or more pivot charts (see Figure 1). As the value of a filter variable (e.g., Region) changes, these charts, and summary measures, if any (e.g., the total), update automatically.

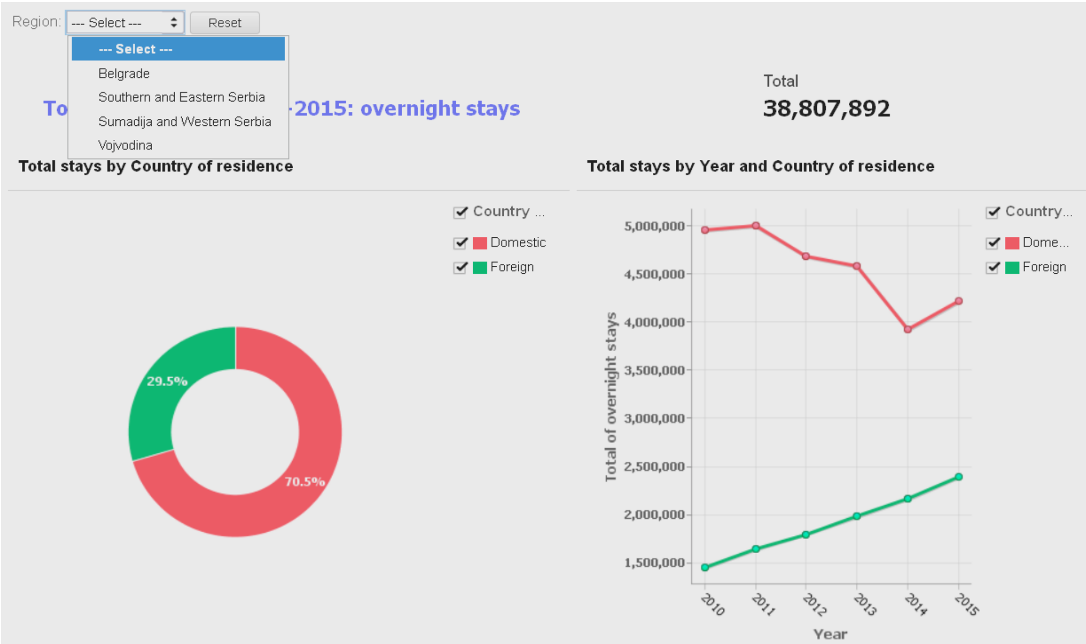

Figure 1. Screenshot of a dashboard available at www.mi.sanu.ac.rs/~djkadij/Dashboard.htm

DASHBOARDS: FEATURES, DIVERSE USE, CREATION, RESEARCH

Dashboard, a relatively simple yet powerful business intelligence tool, is used to display various key performance indicators (KPIs) of business organization to assists managers (usually at operational and tactical levels) to compare actual values of these indicators with past/targeted/forecasted/benchmark values reflecting the objectives and aims of that organization, as well as to signal whether particular numbers or trends are good, acceptable, or bad (Turban et al., 2011). Initially created for the business world, dashboards have entered into various other fields, such as improving patient care, computer-based learning (e.g., learning analytics), and visualizing official statistics (using a dashboard-like tool; Rosling, 2007).

Regarding official statistics, many important institutions, such as the Australian Bureau of Statistics, OECD, and the World Bank, today support the visualization of data concerning social phenomena on their web sites, although most of these visualizations usually make use of just one chart that summarizes values of two or three variables. In other words, the work with business-like dashboards has not been supported, which requires modelers to use flexible dashboards created themselves or provided by other software/web sites in order to do an extensive data modeling.

Dashboards can be created in a number of computer-based environments. One may use a standard software (e.g., Microsoft Excel; Lehman, Lehman & Feazell, 2011), or utilize a Web-based environment (e.g., ZOHO Reports at www.zoho.com, or Statistics eXplorer at ncva.itn.liu.se/explorer/). Some environments (e.g., ZOHO Reports) have versatile drag-and-drop designers that are very easy to use.

As underlined above, data modeling has sporadically been examined in research in mathematics and statistics education, especially computer-supported data modeling. Although there are several recent papers on data visualization in statistics education, studies on data modeling by interactive displays are rare, especially when dashboards are used. To discover whether there are references to teaching and learning about dashboards outside statistics education, we applied an exhaustive search of several bibliographic databases, including EBSCO, ScienceDirect, and SpringerLink. We searched for papers published since 2010 with word ‘dashboard’ in their titles. When such papers were found, we saw whether they examine some aspects of the teaching and learning about dashboards. We found that, in general, research has rarely examined the teaching and learning about dashboards. When it does, it primarily examines opportunities in using these visualization tools. Challenges in using dashboards have sporadically been examined, especially possible reasons of these challenges. Based upon a multidisciplinary literature review (with data modeling in mind), the next section summarizes opportunities in using dashboards, challenges in using them, and reasons for these challenges.

USING DASHBOARDS: OPPORTUNITIES, CHALLENGES AND REASONS FOR THEM

Opportunities

As underlined in the Introduction (see the end of the first paragraph), data modeling is suitable for the development of statistical literacy because it may contribute to the development of various traditional and (relative) 'new' components of that literacy. [For example, the dashboard in Figure 1 revealed the following: (1) for four regions in Serbia, the relative frequency of total overnight stays of foreign tourists was in the range of 14%–69% (range of effect); (2) for three regions, there were many more overnight stays by domestic tourists in each of the six previous years; for the region of Belgrade, the opposite applied (interaction).] Furthermore, data modeling may cultivate not only the use of a data inquiry cycle, but also the application of a mathematical modeling cycle, which would contribute to the development of statistical and mathematical reasoning, respectively. Finally, what has been largely overlooked is that data modeling may also contribute to the development of data retrieving skills (e.g., identification, organization, and retrieval of data; Cox & Nikolopoulou, 1997), because, in some cases, before dashboards can be created and used, data should be prepared by applying database queries, or different organizations of available data (see Figure 2). As many students would probably do some basic data modeling in their future jobs, this modeling seems to be an important goal for higher levels of education, which would, among other things, increase some students' interest in preparing for the (highly paid and valued) career of data scientist.

		REPUBLIC OF SERBIA [RS]			Beogradski region [RS11]			Region Vojvodine	
3		REPUBLIC OF SERBIA [RS]			Beogradski region [RS11]			Region Vojvodine	
4		Total	Domestic	Foreign	Total	Domestic	Foreign	Total	Domestic
5	Tourists overnight stays - monthly data								
6	number								
7	2010/january	339,640	267,449	72,191	63,313	$\boxed{27,154}$	36,159	34,080	24,762
8	2010/february	335,644	257,588	78,056	81,947	$\boxed{32,699}$	49,248	33,562	23,134

	A	B	C	D	E
1	Year	Month	Region	Country of residence	Overnight stays
2	2010	january	Belgrade	Domestic	$\boxed{27,154}$
3	2010	february	Belgrade	Domestic	$\boxed{32,699}$
4	2010	march	Belgrade	Domestic	35,950
5	2010	april	Belgrade	Domestic	41,918

Figure 2. From data available (left) to data to be analyzed (right)

Challenges

Challenges in using dashboards are basically related to (1) data to use, (2) dashboards to use, and (3) modeling process.

Regarding data to use, challenges may be found in querying sources of data (Cox & Nikolopoulou, 1997), organizing available data in different ways (O'Donnell, 2005), and applying various data transformations (e.g., inflation-free normalizations, Tufte, 2001). Even adding new variables may be challenging, especially when complex calculations are required. Concerning data provenance and quality, there are several challenges, including corruptibility of indicators, (Ridgway, James & McCusker, 2013), lack of better indicators (Johnston et al., 2009), and poor data quality (Friedman & Smith, 2011).

Regarding dashboards to use, the requirement to have appropriate presentations of multiple variables (adapted from Tufte, 2001) generates challenges. To find a tool (especially freeware) that can create such presentations may indeed be a struggle (McNamara & Hansen, 2014). Having in mind Figure 1, dashboards should support the work with three or more variables by using various chart types and different summary statistics. Such dashboards can, for example, be created by the ZOHO reports designer, which respects a number of recommendations for dashboard design (e.g., flexible solutions that are easy to upgrade; ability to fit on a single computer screen; simple charts that improve perception; tabular solutions possible; Yigitbasioglu & Velcu, 2013). However, this designer is shareware. In contrast, Statistics eXplorer is freeware for educational and research purposes, but compared to the designer, it is a more powerful tool and thus more complex to use.

Regarding the modeling process, the transition from real world problem statements to math models is one of the most difficult parts of this process (Galbraith & Stillman, 2006). In the case of data modeling, this part requires modelers to identify dependent and independent variables, and specify calculations needed, which may generate challenges (e.g., independent variable used as dependent; selection of variables, especially filters, may not fit specific question to answer; more/less independent variables are needed to answer particular question). Concerning the PPDAC inquiry cycle (Forbes et al., 2011), this requirement is also relevant to its central stage (Analysis) because to find patterns and trends, modelers generate and test hypotheses (e.g., there were more overnight stays of domestic tourists in each region) by using *appropriate* sets of variables. In some cases, asking questions may be a quite challenging part of data analysis (Gould et al., 2016; cf. Schiller & Engel, 2016, for asking questions about data-based statements). Of course, putting the results in context (e.g., telling the story of why certain results are obtained) is, in general, a (very) difficult task (Gal, 2016).

Reasons for challenges

Regarding database querying, students who have not experienced the building of databases may not formulate queries very effectively due to, for example, limited understanding of record and field structures (Cox & Nikolopoulou, 1997). This experience would alleviate challenges regarding organizing data in different ways. Also, when using complex queries, a good understanding of Boolean logic is needed, particularly when negation is used (*Ibid*.)

If modelers create dashboards, then data modeling with these becomes a design task. Its central activity is problem structuring (Jonassen, 2000), which comprises three phases (Dunn cited in Veselý, 2007): problem definition (what to examine, in which terms, and to what extent), problem specification (which input and output variables to use; which measures, charts and tables to use), and problem sensing (whether, and if so, to what extent the formal problem corresponds to the original problem). If we alleviate technology-based challenges (e.g., by using short videos), challenges regarding the modeling process (viewed as a design task) may primarily be caused by lack of knowledge concerning the social phenomena under scrutiny (generalized from Ben-Zvi, 2002), as well as a lack of skills in problem structuring and integrating its parts with respect to questions and hypotheses examined (cf. Kadijevich, 2012). It is context that matters throughout the whole modeling process, and statistical and contextual domains may (and frequently do) interact in a complex way (Makar, Bakker & Ben-Zvi, 2011) that may be characterized as tapestry (Stillman, 1998).

DATA MODELING WITH DASHBOARDS IN STATISTICS EDUCATION

The learning potential of using dashboards (in mathematics, statistics, or computing courses) was stimulated by the author's experience, teaching undergraduate business students the basics of business intelligence (involving pivot charts), at a private university. More than 60 fourth-year students were involved, over two academic years. The experience revealed that most of these students experienced difficulties in answering typical business questions by using pivot charts (even when these charts were given), which is in accord with the findings of O'Donnell (2005). Because dashboards are comprised of several charts, a dashboard clearly offers more learning opportunities, not only in reducing complexity within each chart comprising it, but also in discovering differences and trends in data through examining, in a simultaneous fashion, the outcomes given in two or more charts (see Figure 3). Recall the requirement to use simple charts that improve perception, particularly charts with "a good balance between visual complexity and information utility" (Yigitbasioglu & Velcu, 2013; p. 46).

Most interactive charts used for statistical analyses can deal with just 2–3 variables (Ridgway, James & McCusker, 2013). In line with Tufte (2001), interactive displays provided by the SMART Centre, for example, can deals with four variables by additionally using particular values of one variable as filters (see www.dur.ac.uk/smart.centre/ for such displays), but we can only examine one display at a time. In other words, we cannot examine the outcomes of two displays side by side unless we somehow made copies previously. As a dashboard is basically a set of interactive charts, this problem is removed. Apart from generating various types of interactive charts with different numbers of variables, a dashboard designer (builder) may have additional affordances, such as making queries to produce subsets of the dataset analyzed, which, with some complexity removed, would be easier to model (e.g., the ZOHO reports designer).

Data modeling with dashboards may be practiced when the means of displaying data are examined, which is a traditional topic in introductory or advanced courses on statistics (e.g., Everitt, 1996). Having in mind recent calls to introduce data modeling in secondary education (Davison, 2015; Ridgway, 2015; *The Science Teacher*, *82*(5)), data modeling with dashboards may even be suitable for K-10/12+ statistics education. Some students may build dashboards from scratch or may adapt given dashboards, whereas other students may just examine electronic or paper dashboards created by others. Although we do not have evidence of how this kind of modeling would be taught, the learning potential of using dashboards mentioned above calls for assisting students to build dashboards as sets of charts of increased structural complexity. Of course, possible challenges in using dashboards and the reasons for these challenges, discussed in the previous section in general, should be kept in mind and carefully addressed, if needed. Undoubtedly, some of them would be easier to deal with than others. To support teachers in dealing with them, teacher professional development—possibly arranged in line with McNamara and Hansen (2014) with some Internet-based learning experience—may put into practice a general or specific version of the framework called technological pedagogical content knowledge (Kadijevich & Madden, 2015).

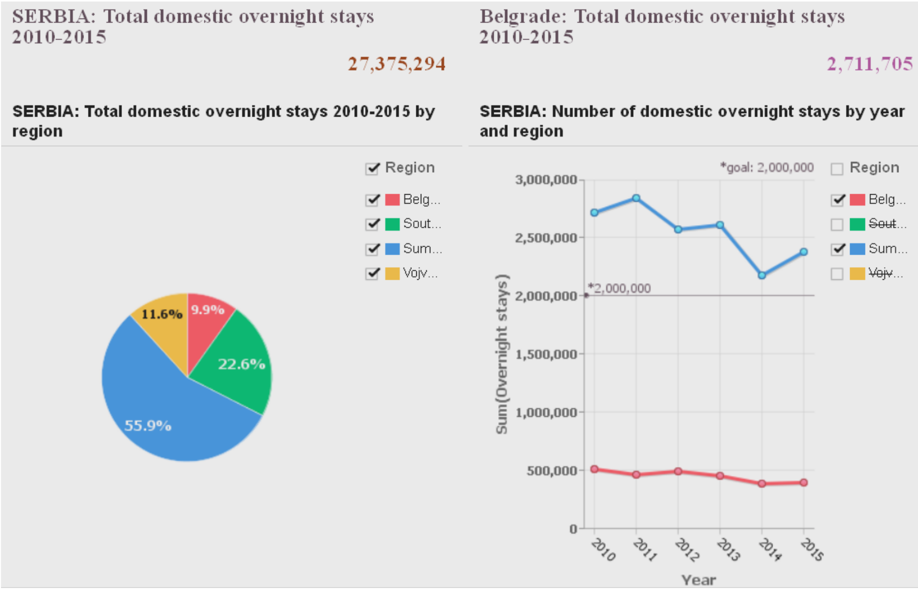

Figure 3. Discovering differences and trends by using simple charts

To make this modeling accessible to students, there are certain computing requirements and prerequisites of statistical nature. Regarding these requirements, simply packing data into proper tables, and using a visual, drag-and-drop reports designer may be needed (if, for example, the ZOHO reports designer is used; a limited edition of this Internet-based designer is available as freeware). Regarding the prerequisites, students should primarily be familiar with basic types of charts, and basic measures of descriptive statistics.

Apart from increasing students' interest in preparing for a career as a data scientist, data modeling with dashboards would particularly cultivate (1) finding patterns and trends in data, and (2) understanding ideas of range, variation, effect size, and interaction. Because observed patterns and trends in social data usually change slightly from year to year, a sudden change may be a sign of some statistical deficiencies (e.g., problematic construct operationalization and measuring; inappropriate data modeling), which would call for improvements, promoting the understanding of statistics about society. (Of course, some statistical deficiencies may be present when wider societal initiatives have resulted in unexpected changes in those patterns and trends or generated no changes in them at all.)

Having regard to the nature of data and statistics concerning society, the teaching of statistics should put less emphasis on inferential statistics, leaving space to deal with issues of data provenance and quality, and promote the understanding of key statistical ideas related to multivariate data modeling, such as effect size, interaction, and confounding variable (Ridgway, James & McCusker, 2013). Regarding future tools for data modeling in this context, apart from providing other measures of descriptive statistics (e.g., median), such tools should be able to (1) assess data quality to some extent; (2) use, if requested, automatic means to combine simple displays into complex ones or decompose complex displays into simple ones; and (3) signal when some common patterns and trends seem to be present on reports generated.

ACKNOWLEDGEMENT

This paper resulted from the author's work on projects 44006 and 47008 supported by the Serbian Ministry of Education, Science, and Technological Development. His participation at the IASE 2016 Roundtable Conference was enabled by a joint grant from the World Bank Trust Fund for Statistical Capacity Building, International Statistical Institute, and International Association for Statistical Education. Any opinions, findings, conclusions or recommendations expressed herein are the author's alone and do not necessarily reflect the views of the funding agencies. The author dedicates the paper to his son, Aleksandar.

REFERENCES

Ben-Zvi, D. (2002). Seventh grade students' sense making of data and data representations. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics* (CD-ROM). Voorburg, The Netherlands: International Statistical Institute.

Cox, M., & Nikolopoulou, K. (1997). What information handling skills are promoted by the use of data analysis software? *Education and Information Technologies*, *2*(2), 105–120.

Davison, C. B. (2015). Addressing the challenges of teaching big data in technical education. *The CTE Journal*, *3*(1), 43–50. Retrieved September 30, 2016, from www.thectejournal.com

English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics*, *81*(1), 15–30.

Everitt, B. S. (1996). *Making sense of statistics in psychology*. Oxford: Oxford University Press.

Forbes, S., Camden, M., Pihama, N., Bucknall, P., & Pfannkuch, M. (2011). Official statistics and statistical literacy: they need each other. *Statistical Journal of the IAOS*, *27*(3/4), 113–128.

Friedman, T., & Smith, M. (2011). *Measuring the business value of data quality*. Stamford, CT: Gartner, Inc.

Gal, I., & Trostianitser, A. (2016). Understanding basic demographic trends: connecting table reading, task design, and context. In Engel, J. (Ed.), *Promoting understanding of statistics about society. Proceedings of the IASE Roundtable Conference*, *July 2016, Berlin, Germany*.

Galbraith P., & Stillman G. (2006). A framework for identifying student blockages during transitions in the modeling process. *Zentralblatt für Didaktik der Mathematik*, *38*(2), 143–162.

Gould, R. et al. (2016). Preparing secondary teachers to teach data science: lessons learned. In Engel, J. (Ed.), *Promoting understanding of statistics about society. Proceedings of the IASE Roundtable Conference*, *July 2016, Berlin, Germany*.

Johnston, R. B. et al. (2009). *Addressing information gaps* (IMF staff position note). Washington, DC: International Monetary Fund.

Jonassen D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63–85.

Kadijevich, Dj. (2012). Examining errors in simple spreadsheet modeling from different research perspectives. *Journal of Educational Computing Research*, *47*(2), 137–153.

Kadijevich, Dj. M., & Madden, S. R. (2015). Comparing approaches for developing TPCK. In Khine, M. S. (Ed.), *New directions in technological pedagogical content knowledge research: multiple perspectives* (pp. 125–146). Charlotte, NC: Information Age Publishing.

Lehman, M. W., Lehman, C. M., & Feazell, J. (2011). Dashboard your scorecard: unleash the power of Excel for visual data analysis. *Journal of Accountancy* (February), 20–27.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, *13*(1), 152–173.

McNamara, A., & Hansen, M. (2014). Teaching data science to teenagers. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics* (CD-ROM). Voorburg, The Netherlands: International Statistical Institute.

Mulligan, J. (2015). Moving beyond basic numeracy: data modeling in the early years of schooling. *Zentralblatt für Didaktik der Mathematik*, *47*(4), 653–663.
O'Donnell, P. (2005). The problem with pivot tables. *Business Intelligence Journal*, *10*(1), 25–31.
Ridgway, J. (2015). Implications of the data revolution for statistics education. *International Statistical Review* (to appear). doi:10.1111/insr.1211
Ridgway, J., James, N., & McCusker, S. (2013). 'Open data' and the semantic web require a rethink on statistics teaching. *Technology Innovations in Statistics Education*, *7*(2).
Rosling, H. (2007). *Gapminder* (computer software). Stockholm, Sweden: Gapminder Foundation.
Schiller, A., & Engel, J. (2016). Civic statistics and the preparation of future secondary school mathematics teachers. In Engel, J. (Ed.), *Promoting understanding of statistics about society. Proceedings of the IASE Roundtable Conference*, *July 2016, Berlin, Germany*.
Stillman, G. (1998). The emperor's new clothes? Teaching and assessment of mathematical applications at the senior level. In Galbraith, P., Blum, W., Booker, G. & Huntley, D. (Eds.), *Mathematical modelling: teaching and assessment in a technology-rich world* (pp. 243−253). Chichester: Horwood.
Tufte, E. R. (2001). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
Turban, E., Sharda, R., Delen, D., & King, D. (2011). *Business intelligence: a managerial approach* (2nd ed). Upper Saddle River, NJ: Pearson.
Veselý, A. (2007). Problem delimitation in public policy analysis. *Central European Journal of Public Policy*, *1*(1), 80−100.
Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265.
Yigitbasioglu, O. M., & Velcu, O. (2013). A review of dashboards in performance management: implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41–59.