THE TEACHING OF MATHEMATICS
2020, Vol. XXIII, 1, pp. 71–80

MODERN STATISTICAL LITERACY, DATA SCIENCE, DASHBOARDS, AND AUTOMATED ANALYTICS AND ITS APPLICATIONS

Djordje M. Kadijevich and Max Stephens

Abstract. With regard to the internationalization of statistics education, this paper considers first a global context concerning modern statistical literacy, data science, and dashboards. Then, it examines data discovery using automated analytics, whereby data insights may be indicated by suitable signals generated by the computer environment used. This theoretical paper, directed towards statistics educators, as well as other educators in relevant high school subjects, should make them (more) aware of this context and such analytics, supporting them to identify issues that need be considered in their teaching (and research) in order to have their students better prepared for the jobs of tomorrow.

MathEduc Subject Classification: K44, R44

MSC Subject Classification: 97K80, 97R40

Key words and phrases: Automated analytics; data science; statistics education; upper secondary.

1. Introduction

In response to internationalization in education, which may basically denote adapting educational products provided by international institutions to particular country needs (e.g., [4]), many countries have applied TIMSS (Trends in International Mathematics and Science Study) and PISA (Programme for International Student Assessment) studies, whose designs and outcomes have reshaped mathematics curricula in many countries around the world (e.g., [19]).

In many countries statistics is seen as a significant component of the school mathematics curriculum; an integrated view is universally accepted for the compulsory years of school, but some differentiation may become more evident in the senior secondary years. Apart from the influence of these international studies on basic data practice during the compulsory years of school (e.g., representing and interpreting data using tables and graphs), there has been additional internationalization in statistics education in senior secondary years, which may lead to a sharper distinction of the place of statistics in these two educational levels. This additional internationalization was recently evidenced by, for example, the International Data Science in Schools Project, whose goal was to develop a pre-calculus data science course and a corresponding teacher professional development program (for more details, see `http://www.idssp.org/`).

To be better prepared for the jobs of tomorrow, students need to become familiar with data science – the science of finding useful information in data by using various methods. Many students may do this at an introductory level, focusing on exploratory data analysis using interactive reports [21]. These reports, which are digital artifacts whose content is updated automatically whenever changes in data or variables occur, are building blocks of dashboards comprising two or more such reports, mostly charts. To cultivate modern statistical literacy, the use of dashboards may be practiced within an appropriate data science cycle [11]. In doing that, dashboards data discovery may be coupled with automated analytics, whereby data insights may be indicated by suitable signals.

The first part of this contribution summarizes the features of modern statistical literacy, underlies an emerging relevance of data science to both education and workplace, and presents how simple data science may be attained through exploratory data analysis using interactive reports. All these issues, some of which readers of this contribution may not be familiar with, may be considered to be an important component of STEM (Science, Technology, Engineering, and Mathematics) education, requiring scientific, technological, mathematical, statistical, and other relevant issues to be examined, utilized, and combined in problem solving, often thereby being a multidisciplinary challenge [24]. The approach to data science presented in this part may be covered within a module of an introductory course on statistics or a related subject in upper-secondary or post-secondary education, and to this end the course instructor may consider and make use of didactic issues examined in Kadijevich [16,17,20,21], for example.

The remaining part of the contribution focuses on automated analytics, whereby data insights may be indicated by the computer environment used. For example, a signal generated by the environment indicates a change in trend direction at some time point (e.g., [31]). Although this promising approach to data practice, introduced at the end of the 2010s, is still being developed, statistics educators, as well as other educators in relevant subjects (e.g., Data mining and Business intelligence), should be aware of this emerging approach, having their students familiar with implemented solutions, some of which they might use in their future jobs. To this end, the course instructor may not only demonstrate, but also arrange and support basic practical work in concrete computer environments, and, in doing so, a critical evaluation of data insights provided by technology should be cultivated (derived from [13]).

In conclusion, this paper links the issues discussed to recent significant rethinking of the kind of education that students will need for their future. To this end, an OECD report [23] is briefly considered.

2. Modern statistical literacy, data science, and dashboards

Modern statistical literacy

In general terms, statistical literacy (SL) denotes a person's ability to understand, interpret, evaluate and communicate statements with statistical elements obtained from various sources [9]. Traditional SL components deal with collecting, representing, and interpreting data; finding patterns and trends in data representations; understanding ideas of range, variation, and distribution; and making predictions [6,22]. Data revolution, based on using authentic (large) and often open data sets [25], whose examination goes (much) beyond the traditional p-value practice [30], has called for some novel SL components. Among them are those related to data provenance and quality, as well as understanding the notions of effect size and interaction associated with the analysis of multivariate data, by using, for example, interactive visualizations [26]. (An interactive visualization may show that males have a 10% higher achievement than females, pointing to an effect size; another may show different achievement patterns for males and females due to e-learning practice, evidencing an interaction between gender and that practice.) Apart from applying exploratory data analysis skillfully (mostly required from an introductory data practice), some familiarity with various machine learning techniques is usually required at an advanced level of data practice to be able to extract more (valuable) information from data (e.g., [30]). Clearly, modern SL includes a number of novel components that may be cultivated through an introductory data science practice, for example.

Data science

Data science is concerned with finding useful information in data, by using exploratory data analysis or other (advanced) mathematical/statistical methods. To support economic growth and better prepare students and workers for (many unknown) jobs of tomorrow, APEC [2] developed a list of Data Science and Analytics (DSA) competencies, including: enhanced skills in data presentation and visualization; versatile applications of data analytics methods; computational thinking and use of algorithms. It is predicted that job candidates with DSA competencies will soon be preferred by most employers in developed economies (for example, see `http://www.bhef.com/sites/default/files/bhef_2017_investing_in_dsa.pdf`). Jobs that embody data science are, according to the Careercast 2019 jobs rates report (`https://www.careercast.com`), very highly paid with a median salary of about $115,000 (USD) annually. Such a high salary level reflects the complexity of knowledge and skills required by data science practitioners: apart from applying a good contextual knowledge of a concrete business, scientific or other area (e.g., insurance), these professionals need solid knowledge and skills in computing (e.g., programming and/or databases) and mathematics (the application of a variety of mathematical and statistical models), as well as, among other things, a high degree of creative thinking and communication skills (e.g., [30]). As workplace contexts change, it is reasonable to expect that students will increasingly be introduced to key elements of data science (at least at an introductory level). Recently, there has been a clear call for the inclusion of data science in secondary education, which according to Gould et al. [11], should be based on the integration of computational thinking and statistical reasoning. A timely example of this integration is the International Data Science in Schools Project (`http://www.idssp.org/`) mentioned above.

Dashboards

Data science at an introductory level may be practiced using dashboards. They are sets of two or more interactive reports, mostly interactive charts, whose content updates automatically whenever changes in data or variables considered occur (e.g., `http://www.mi.sanu.ac.rs/~djkadij/Dashboard.htm`; [17]), making them a powerful tool for visualizing data, i.e. for exploratory data analysis. Dashboards are usually built in a drag-and-drop fashion like the interactive reports of which they comprise. Although initially created for the business world to display values of key performance indicators and see when these values and their trends are good, acceptable or bad, dashboards have entered various industries and areas (e.g., `https://www.idashboards.com/dashboard-examples/`), including education, where they can be used not only to visualize learning data (the so-called learning analytics), but also to promote thinking in a particular discipline [20,21]. For example, apart from traditional SL components (e.g., finding patterns and trends in data), the use of interactive displays (charts and dashboards) can foster the development of some novel SL components, such as the notions of effect size and interaction mentioned above (e.g., [16,25]). Research has evidenced that when using interactive displays (e.g., those available at `https://www.dur.ac.uk/smart.centre/freeware/`), these notions could be grasped by many senior secondary students [25]. To cultivate various SL components, the use of dashboards may be practiced within a data science cycle, whose stages may be *Ask questions*, *Consider data*, *Analyze data*, and *Interpret data* ([11]; for other such cycles, see [30]). In doing that, the use of dashboards can support the understanding of the cycle and the realization of its values in capturing main features of statistical reasoning. Although mostly based on simple mathematical/statistical models (primarily frequencies, sums, and means), the use of interactive displays may generate a number of challenges, regarding, for example, decisions about what variables to use and what charts to apply. Building displays of increased structural complexity may also generate challenges. These challenges can be better understood and pedagogically addressed if examined within a suitable learning cycle with respect to its stages and transitions between them. If dashboards were used to promote disciplinary reasoning (reasoning in a particular discipline), this understanding would contribute to dashboard design issues, which may include some intelligent support, such as signaling possible data regularities in displays produced [21].

3. Automated analytics and its applications

Despite their evident values in summarizing data, dashboards are limited tools for data discovery because the aggregated data they display does not support an in-depth search of the data being aggregated (e.g., [31]). In other words, making dashboards interactive is usually not enough to realize what matters most, especially by novice dashboard users who first need to learn how to read them competently rather than just focusing on single data points displayed. Things get more complicated with real-time dashboards, whose interactive reports including visualizations update automatically on a regular time interval when new values of variables used for these reports become available.

To uncover data insights (patterns and trends in data that matter to the question being considered), a number of business intelligence (BI) vendors have included automated analytics in their BI platforms, such as IBM Cognos Analytics, Tellius, and Yellowfin[1]. This analytical tool is based on machine learning that applies a number of algorithms, such as those related to classification, regression, decision trees, time series, and neural networks [7]. These insights may be personalized if the user can specify what patters and trends are in his/her focus. The BI platform may be instructed to search for only spikes and drops (sudden increases or decreases in the dependent variable) or trend direction changes (changes in trend lines from going down to going up, or vice versa) if these are present [31]. Figure 1 presents a data insight discovered by the Yellowfin platform (`https://www.yellowfinbi.com/`).

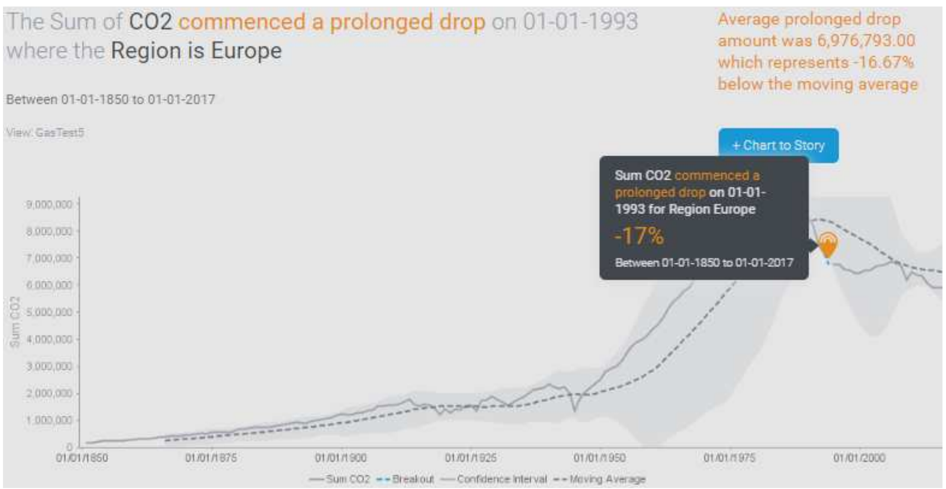

Figure 1. A considerable reduction in annual carbon emission (1850–2017)
Data set: `http://www.mi.sanu.ac.rs/~djkadij/CarbonEmission.csv` (source [12])

Machine learning

As mentioned above, machine learning (ML) is a critical component of automated analytics. Besides post-secondary levels (through courses in data science, data mining, or business intelligence; e.g., [30]), ML can be introduced in the senior secondary years. To this end, a data science module may include basic ML concepts and techniques, but the teaching approach should respect the different levels of mathematical knowledge that students have. For example, for students with a good basic knowledge of probability, association rule learning may be suitable, whereas for students with a solid basic knowledge of functions, iterations, and optimization, simple neural networks (starting with perceptron) may be appropriate (for suitable ML approaches, see [10, 32], for example). Having understood the ML approach (i.e., its key conceptual and procedural issues), students should not only

[1] According to `bi-survey.com`, BI software tools may focus on various data processing issues, such as Dashboarding (DB), Data Discovery (DD), and Embedded Analytics (EA). Examples are Microsoft Power BI, Pyramid A.OS, and Sisense (supporting BD, DD, EA), Tableau (DB, DD), Yellowfin (DD, EA), and Zoho Analytics (DB, EA). To compare BI products with respect to specific focus, the Reader may use a comparison tool available at `https://bi-survey.com/business-intelligence-software-comparison`.

experiment with a software that applies this approach, but may also, according to their computer science knowledge, analyze this software and try to program some of its components.[2] However, for most students (expected only to be future ML users) instruction should primarily focus on basic understanding of ML approaches rather than on mastering particular software solutions (which change as technology advances and data nature evolves). This may help students become more receptive to various ML approaches they might meet in the future.[3]

Despite sporadic initiatives to include ML topics in secondary education (e.g., [8, 32]), research and instructional efforts have been rather meagre, especially regarding a wider context comprising data mining, big data, and relevant ethical and privacy matters (e.g., [5]). A notable exception is the above-mentioned International Data Science in Schools Project, whose framework resulted from wide international cooperation[4] calls for the inclusion of ML in such a context in the senior secondary schools, through, for example, learning about and experimenting with classification and regressions trees (as examples of supervised ML) and K-means clustering (as an example of unsupervised ML). Compared to conventional statistics courses, a greater emphasis is put on data, learning from data, and computing (programming) aspects [15].

Signals approach

As shown above in Figure 1, automated analytics may support the stage of *Analyze data* by generating signals to mark data insights. With the promotion of modern statistical literacy (MSL) in mind, what are the pros and cons of using these signals and desirable improvements to undertake?

Pros and cons. As already mentioned, MSL calls for some novel SL components, including data provenance and quality, as well as understanding the notions of effect size and interaction associated with the analysis of multivariate data. With the signals approach, pointing to various data patterns and trends (e.g., spikes and drops or trend direction changes), the user can be supported to realize useful data insights considering these notions, but also obtain evidence of possible data flaws. For example, when the BI platform generates the signal for volatility changes (changes in the values of interest that rise and fall over a sustained time period) for some data that usually changes slightly from one time period to another, this signal may point out possible data deficiency regarding data provenance/quality.

[2]For example, due to a simple mathematical background (basically Euclidean distance and mean), programming K-means clustering may be an appropriate ML algorithm to begin with [8]. Regarding software applied in general, while freeware may be used for an introductory work, commercial software applied in industry should be used for an advanced work (e.g., [28]).

[3]In doing that, triggered by motivating real-world examples, dealing with ML approaches would become an inspiring computer science topic that does not necessarily require programming [10], whose requirements are usually very demanding (and thus often demotivating) for many students.

[4]The project involves different educational specialists from Australia, Canada, England, Germany, the Netherlands, New Zealand, and the United States. Apart from statisticians and computer scientists, the project team comprises curriculum experts, educators, and school teachers. The project is supported by leading international and national professional societies gathering statisticians and computer scientists.

However, the signals approach, which was introduced about a year ago, has been developed for the business world, relying thus on conceptual and technical solutions that are at present not suitable for statistics education. It may, for example, require large data sets (with 10K+ records) accommodated on professional servers, the use of date fields, and real-time data analyses, being able to display signals for individual interactive charts only. For example, instead of comparing trends in two time series possibly providing evidence for interaction, the trend signal may only compare trends in two periods for the single series.

Desirable improvements. To be more suitable for education applications, a signals approach should highlight interactions when the data for two groups (e.g., European vs Asian countries) are visualized. This means that the signals approach can support the comparisons of trends or effect sizes as well. To ease its comprehension, each signal may be coupled with a suitable graphical cue (for such clues, see [1]). As independent variables are not always time-based, the use of a date field is not always needed. The restriction to apply real-time data analyses should also be removed, because such data sets, reflecting professional needs, are rarely accessible in educational settings. Although uncovering data insights with statistical significance may require voluminous data sets (with 10K+ records, for example), a signals approach should work with data sets of moderate size as well. Vendors of BI platforms may at first sight not be particularly interested in becoming involved in educational matters, but they may indeed see the value of helping students become familiar with the kind of platforms they are likely to meet in the future, and thus offer educational institutions demo versions of these platforms that include some of the desirable improvements listed. As automated analytics works as a black-box, software developers may also consider offering an option that opens this box to some extent in order to encourage and support learning about the structure and key components of the box (derived from [3]), relating relevant procedural and conceptual knowledge (e.g., [18]).

***

Apart from automated analytics, there is an emerging trend in business analytics called augmented analytics, which make use of both machine learning and natural language generation, and are aimed to support all stages of data science cycle, simplifying the workload enterprise for data scientists. Bearing in mind the stages of data science cycle introduced above, the stage of *Ask questions* may be supported by enabling the user to ask questions by voice or by typing them; the stage of *Consider data* by preparing data in the best possible way recommended by the BI platform; the stage of *Analyze data* by applying the automated analytics examined above to uncover data insights; the stage of *Interpret data* by generating stories in natural language about the data insights uncovered (for more details, see [27], for example).

4. Closing remarks

Over forty years ago, Geoffrey Howson remarked wisely that “Mathematics education does not take place in a vacuum. It must reflect or even anticipate changes in the educational and social system" [14, p. 183]. In our rapidly changing world, it is clear is that a static or backwards-looking view of core mathematical knowledge will not equip todays young people for an emerging and difficult-to-predict future. Core disciplinary knowledge will need to expand to include topics that orient students towards their futures: including statistics, data analysis, familiarity with "big data", and techniques like computational thinking that are necessary to deal with an increasingly data-driven world (e.g., [2, 15, 29]).

To have students engage in solving interdisciplinary tasks that require them to apply their mathematical knowledge in authentic settings, certain skills for lifelong and self-regulated learning are needed. These skills have been, in general, carefully elaborated in the OECD Learning Compass 2030 key goals [23], including transformative competencies fostered through the application of Anticipation-Action-Reflection learning cycle, which may be used to support data science practice examined in this paper.

This authentic practice was presented in the following way: after considering a global context concerning modern statistical literacy, data science, and dashboards, data discovery using automated analytics was examined, whereby data insights may be indicated by suitable signals. An implementation of this approach was described, followed by pros and cons of this implementation pointing to desirable improvements. The authors hope that this theoretical account, directed towards statistics educators (as well as other educators in relevant subjects, such as data mining and business intelligence) should make them (more) aware of the context and analytics examined, supporting them to identify issues that need be considered in their teaching (and research) in order to have their students better prepared for the jobs of tomorrow.

Acknowledgements. The authors wish to thank Tatjana Kovačević and Daniel Stanković, from Micro Business Solutions, Belgrade, Serbia, a regional partner of Yellowfin, for their support in the examination of a demo version of a Yellowfin suite containing the Signals product.

The authors also thank the anonymous referee whose suggestions helped to improve the paper.

The research done by the first author was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract No. 451-03-68/2020-14/200018).

REFERENCES

[1] Acartürk, C., *Towards a systematic understanding of graphical cues in communication through statistical graphs*, Journal of Visual Languages and Computing, **25** (2) (2014), 76–88. `https://doi.org/10.1016/j.jvlc.2013.11.006`

[2] Asia Pacific Economic Cooperation (APEC) Agency, *Data science and analytics skills shortage: Equipping the APEC workforce with the competencies demanded by employers*, Singapore, The Author, 2017. `https://www.apec.org/Publications/2017/11/Data-Science-and-Analytics-Skills-Shortage`

[3] Buchberger, B., *Should students learn integration rules?*, SIGSAM Bulletin, **24** (1) (1990), 10–17. `https://doi.org/10.1145/382276.1095228`

[4] Cai, J., & Howson, G., *Toward an international mathematics curriculum*, In: Clements, M., Bishop, A., Keitel, C., Kilpatrick, J., & Leung, F. (Eds.), *Third international handbook of mathematics education*, (pp. 949–974). New York: Springer, 2012. https://doi.org/10.1007/978-1-4614-4684-2_29

[5] Dryer, A., Walia, N., & Chattopadhyay, A., *A Middle-school module for introducing data-mining, big-data, ethics and privacy using RapidMiner and a Hollywood theme*, In: *Proceedings of the SIGCSE '18: 49th ACM Technical Symposium on Computer Science Education*, February 21–24, 2018, Baltimore, MD. New York: ACM, 2018. https://doi.org/10.1145/3159450.3159553

[6] English, L. D., *Data modelling with first-grade students*, Educational Studies in Mathematics, **81** (1) (2012), 15–30. https://doi.org/10.1007/s10649-011-9377-3

[7] Ereth, J., & Eckerson, W., *AI: The new BI. How algorithms are transforming business intelligence and analytics*, Hingham, MA: Eckerson Group, 2018. https://www.eckerson.com

[8] Essinger, S. D., & Rosen, G. L., *An introduction to machine learning for students in secondary education*, In: *Proceedings of the 2011 Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, Sedona, AZ, 2011 (pp. 243–248), IEEE. https://doi.org/10.1109/DSP-SPE.2011.5739219

[9] Gal, I., *Adult statistical literacy: Meanings, components, responsibilities*, International Statistical Review, **70** (1) (2002), 1-25. https://iase-web.org/documents/intstatreview/02.Gal.pdf

[10] Gavaldà, R. (n.d.), *Machine learning in secondary education?*, Paper presented at Teaching Machine Learning Workshop, 6–7 May 2008, Saint-Etienne, France. https://www.cs.upc.edu/~gavalda/docencia/tml08-revised.pdf

[11] Gould, R., Bargagliotti, A., & Johnson, T., *An analysis of secondary teachers' reasoning with participatory sensing data*, Statistics Education Research Journal, **16** (2) (2017), 305334. https://iase-web.org/documents/SERJ/SERJ16(2)_Gould.pdf?1512143292

[12] Gtschow, J.; Jeffery, L.; Gieseke, R., Gnther, A., *The PRIMAP-hist national historical emissions time series (1850–2017)*, v2.1. GFZ Data Services, 2019. https://doi.org/10.5880/pik.2019.018

[13] Hicks, S. C., & Irizarry, R. A., *A guide to teaching data science*, The American Statistician, **72** (4) (2018), 382–391. https://doi.org/10.1080/00031305.2017.1356747

[14] Howson, A. G., *Change in mathematics education since the late 1950's – ideas and realizations in Great Britain*, Educational Studies in Mathematics, **9** (2) (1978), 183-223. https://doi.org/10.1007/BF00221158

[15] International Data Science in Schools Project (IDSSP) Curriculum Team, *Curriculum frameworks for introductory data science*, 2019. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

[16] Kadijevich, D. M., *Data modeling with dashboards: opportunities and challenges*, 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, ISI/IASE, the Haag, the Netherlands, 2016. https://iase-web.org/documents/papers/rt2016/Kadijevich.pdf?1482482774

[17] Kadijevich, D. M., *Data modelling using interactive charts*, The Teaching of Mathematics, **21** (2) (2018), 55-72. http://www.teaching.math.rs/vol/tm2121.pdf

[18] Kadijevich, D. M., *Relating procedural and conceptual knowledge*, The Teaching of Mathematics, **21** (1) (2018), 15-28. http://www.teaching.math.rs/vol/tm2112.pdf

[19] Kadijevich, D. M., *Influence of TIMSS research on the mathematics curriculum in Serbia: Educational standards in primary education*, The Teaching of Mathematics, **22** (1) (2019), 33-41. http://www.teaching.math.rs/vol/tm2213.pdf

[20] Kadijevich, D. M., *Cultivating computational thinking through data practice*, In: Passey, D., Bottino, R., Lewin, C., & Sanchez, E. (Eds.), *Empowering learners for life in the digital age* (pp. 24-33). Cham, Switzerland: Springer, (2019). https://doi.org/10.1007/978-3-030-23513-0_3

[21] Kadijevich, D. M., *Interactive displays: Use of interactive charts and dashboards in education*, In: A. Tatnall (Ed.), *Encyclopedia of education and information technologies*, Cham, Switzerland: Springer, 2019. https://doi.org/10.1007/978-3-319-60013-0_228-1

[22] Mulligan, J., *Moving beyond basic numeracy: Data modeling in the early years of schooling*, Zentralblatt für Didaktik der Mathematik, **47** (4) (2015), 653-663. https://doi.org/10.1007/s11858-015-0687-2

[23] OECD, *OECD future of education and skills 2030: OECD learning compass 2030*, Paris: OECD Publishing, 2019. http://www.oecd.org/education/2030-project/contact/OECD_Learning_Compass_2030_Concept_Note_Series.pdf

[24] Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., Rsken-Winter, B., Tiemann, R., & Upmeier Zu Belzen, A., *A framework to foster problem-solving in STEM and computing education*, Research in Science & Technological Education, **38** (1) (2020), 105-130. https://doi.org/10.1080/02635143.2019.1600490

[25] Ridgway, J., *Implications of the data revolution for statistics education*, International Statistical Review, **84** (3) (2016), 528-549. https://doi.org/10.1111/insr.12110

[26] Ridgway, J., James, N., & McCusker, S, *'Open data' and the semantic web require a rethink on statistics teaching*, Technology Innovations in Statistics Education, **7** (2) (2013). https://escholarship.org/uc/item/6gm8p12m

[27] Sallam, R., Richardson, J., Howson, C., & Kronz, A., *Magic quadrant for analytics and business intelligence platforms*, Stamford, CT: Gartner, 2019. https://www.gartner.com

[28] Satyanarayana, A., *Software tools for teaching undergraduate data mining course*, Paper presented at American Society of Engineering Education Mid-Atlantic Fall Conference, October 2013, Washington, DC. New York: New York City College of Technology, 2013.

[29] Stephens, M., & Kadijevich, D. M., *Computational/algorithmic thinking*, In: Lerman, S. (Ed.), *Encyclopedia of mathematics education*, Dordrecht, the Netherlands: Springer, 2020. https://doi.org/10.1007/978-3-319-77487-9_100044-1

[30] Yan, D.,& Davis, G. E., *A first course in data science*, Journal of Statistics Education, **27** (2) (2019), 99-109. https://doi.org/10.1080/10691898.2019.1623136

[31] Yellowfin, *A guide to automated analytics: How to stop running your business on accidental data discovery*, Melbourne: The Author, 2018. https://www.yellowfinbi.com/

[32] Young, R. & Ringenberg, J., *Machine learning: An introductory unit of study for secondary education*, In: *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*, New York: ACM, 2019. https://doi.org/10.1145/3287324.3293806

D.M.K.: Institute for Educational Research, Belgrade, Serbia
E-mail: djkadijevic@ipi.ac.rs
M.S.: The University of Melbourne, Melbourne, Australia
E-mail: m.stephens@unimelb.edu.au