DISTANCE LEARNING TRENDS: INTRODUCING NEW SOLUTIONS TO DATA ANALYSIS COURSES

Karol FLISIKOWSKI

Politechnika Gdańska, Wydział Zarządzania i Ekonomii, Katedra Statystyki i Ekonometrii
tel.: 58 348 63 12   e-mail: karflisi@pg.edu.pl

Abstract: Nowadays data analysis of any kind becomes a piece of art. The same happens with the teaching processes of statistics, econometrics and other related courses. This is not only because we are facing (and are forced to) teach online or in a hybrid mode. Students expect to see not only the theoretical part of the study and solve some practical examples together with the instructor. They are waiting to see a variety of tools, tutorials, interactive laboratory instructions, books, exams online. In this study I am going to show the latest technical solutions for instructors using R and/or Python in their online data analysis labs.

Keywords: markdown, bookdown, r-exams, tutorials, moodle, e-learning, webexercises, github, data analysis.

1. INTRODUCTION

1.1. E-learning

Currently the field of distance learning, e-learning, is technologically being transformed and becomes a true art and science. Its transformation, because of the pandemic situation, is being noticed not only among professionals but also in higher or adult education and at schools. Suddenly we were moved to the distance learning or hybrid modes of teaching, with or without any trainings. Most of the schools owned at the time learning management system (LMS) like Moodle or OpenOLAT or Canvas, but it was and still is very difficult to fill the course with the proper content (not only with the lecture notes). Very often, it is a piece of art to engage students during online or hybrid classes. In some cases gamification strategies were used successfully, but in courses based on more technically advanced laboratories this kind of learning process started to be especially difficult. Nowadays it is also very difficult to separate technological and methodological site of the teaching and learning processes [1]. Over the last five years we witness an incredibly fast process of distance learning redefinition. New plugins are developed for LMS, new types of content are created by users. All of those encouraged the whole communities of various software users and producers to implement innovative solutions for teaching and learning purposes. This is why the main aim of this article is to present some of the most popular, free of charge, open source based tools invented for courses related to data analysis (statistics, econometrics, machine learning, data mining etc.).

1.2. Data analysis

Data analysis (DA) and all the related courses are now taught with the use of many technological and pedagogical innovative methods. DA learners can access more lessons, more quickly than ever using online streaming services, massive open course services. Teachers on the other hand can have far more detailed insight into how learners work. At the university level DA courses include usually lectures, seminars and laboratory sessions. All three parts should present one topic in three different but at the same time related ways. First, during the lecture, theoretical introduction is presented. There is a huge debate, that this form of DA teaching process should be minimized [2]. Then, students can practice solving challenging, introductory tasks in the seminars' classroom. Finally, in the computer laboratory they may learn all stages of DA using advanced software packages. Now let's switch the mode and try to organize DA online or hybrid, blender learning course. There may exist some scheduling conflicts if we will try to organize synchronous all-class activities like live lectures, so there are many opinions that it should be minimized or shortened. Learners now can organize their own forums on the LMS courses and contribute to collective knowledge, take notes together, serve classroom scripts, problem sets and their solutions. We can conclude, that the way how DA courses are organized move the importance of the learning process to the laboratory sessions. They become more and more important. Here participants can learn together problem solving, share ideas how to process data in various ways, publish scripts or exchange code chunks. The role of the teacher is moved to organize and supervise the course rather than actively participate in the lessons [3]. That is why the new concepts of the DA courses, their tools and technological innovative methods is nowadays more and more significant.

2. NEW TECHNOLOGICAL SOLUTIONS IN DA COURSES

2.1. IDE

The first and the most important step in the DA online course design process is to organize an integrated development environment (IDE) for R, Python or other DA language. In majority of cases it is RStudio, PyCharm, Visual Studio Code etc. We can organize also web-based IDE platforms like RStudio Server, Jupyter Project, Google Colab, Spark Notebook, PyCHarm, Apache Zeppelin etc. It may allow us to use only it remotely, attach to our LMS courses easily, having the same version among learners all the time. IDE includes a console, syntax-highlighting editors, code execution panel, tools for plotting, history,

debugging and workspace management. They are very user-friendly and enable connection with services and repositories like GitHub or RStudio Connect, R Pubs and many others. Thanks to that function students may easily publish and exchange their results – final projects, exams, exercises using just a hyperlink to their account's items. The main window of RStudio used by DA lab's participants is presented on the figure 1.
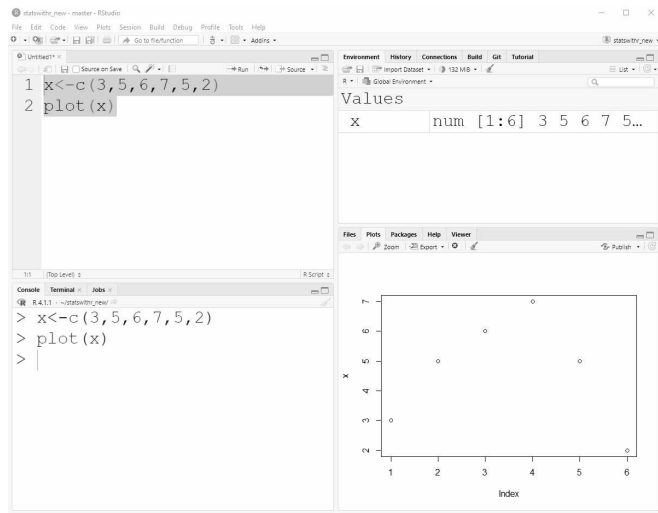

Fig. 1. RStudio main window with 4 sections: R-script (top-left); data environment and preview (top-right); R console (bottom-left); plot's preview (bottom right)

2.2. R-Markdown

The key element in the online DA laboratory is to facilitate peer-to-peer interactions. To aid this, those courses almost always should have some kind of discussion forum [4]. It is observed that learners use these in very different ways, but we should keep them engaged as much as possibilities allow. One solution for that problem is the Markdown format of R-documents. Inside of those reports our students may save and execute code and at the same time generate high quality content that can be shared. Markdown support dozens of static and dynamic output formats like HTML, PDF, Word, PowerPoint, Beamer, R-Shiny apps etc [5].
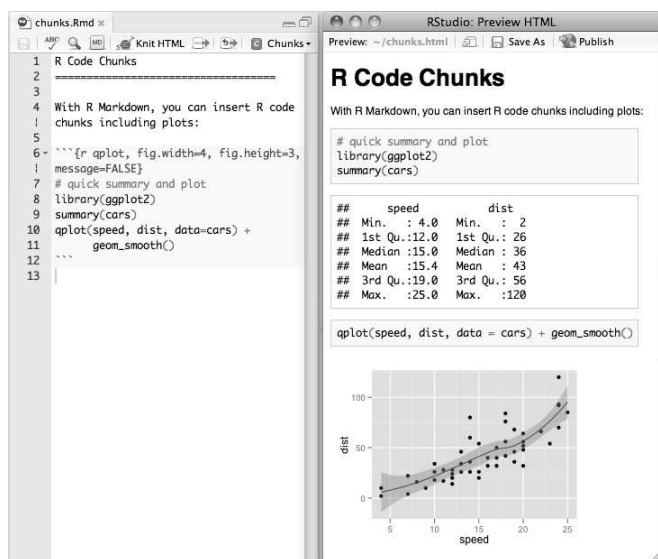

Fig. 2. R Markdown document: Markdown syntax on the left, knitted report as HTML on the right (code chunk computing a data summary and rendering a simple scatterplot)

There are three components of the R Markdown documents: the metadata, text and code. In the online DA lab participant may include their proposed code chunks as solutions for homework assignments, in-class exercises. Other language engines may be used as well (Python, SQL, Rcpp, C, etc.). We can also organize our online class including Markdown instructions with chunks, where some portions of codes are missing, giving hints or sample solutions to students highlighting the most important elements of the lesson (i. e. code chunks, solutions like plots, interpretations). There are many more ways how those reports could be used during our DA online classes. Some, more ambitious projects like HTML widgets or Shiny applications may be prepared with this kind of Markdown syntax as well. What is most important, participant may collaborate publishing their reports to their GitHub, R Pubs accounts. On the figure 2. the basic outline of a very simple Markdown document is presented.

2.3. R-Bookdown

The additional version of Markdown syntax is Bookdown [6]. It facilitates writing books and long-form articles/reports. Instructors can prepare the whole online, asynchronous version of the DA course using e-books compiled directly from Markdown documents. What is more, the HTML version (website) served on some of the free of charge, open servers, enables usage of interactive modules inside it. Students can download multiple version of formats like PDF, LaTeX, HTML, EPUB or Word. Teachers may link the online content of their books, courses, tutorials directly to lecture notes, instructions. A whole collection of R-Bookdown books is available on the Bookdown's archive page: https://bookdown.org/home/archive. My own DA book written in Bookdown is presented on the figure 3.
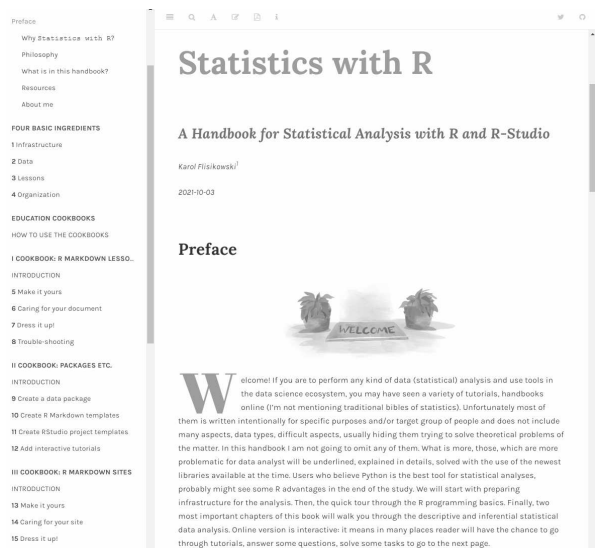

Fig. 3. R-Bookdown based book – "Statistics with R" by Karol Flisikowski published on the R-Connect server and also on Netlify

2.4. Interactive tutorials

In order to minimize cognitive load and provide more opportunities for DA online course participants more teacher-student engaging elements should be included. One of those are interactive tutorials helping students understand problems and learn basic DA programming techniques. We can organize such tutorials with the use of "learnr" or "webexercises" packages built for R [7]. The "learnr"

package is just the next version of the Markdown document that can be easily knitted to an interactive tutorial and published inside a course (nested) or on the website, included in the content of an e-book etc. Tutorials consist of content along with interactive components for checking and reinforcing understanding. The structure of them may be formed freely, from the introductory section, with or without the main menu, through problem solving with hint and solutions (code chunks), ending with the theoretical summaries and interpretations. An additional "gradethis" package may help teachers use a model solution as a templat or write highly customized testing logic to provide specific feedback for common mistakes made by students. Learnr tutorials can include any or all of the following:

- narrative, figures, illustrations, equations;
- code exercises (R code chunks that users can edit and execute on-the-fly inside of our online courses);
- quiz questions;
- videos;
- interactive Shiny components.

On the figure 4. my own DA course (Mathematical Statistics) published as the online course on the moodle platform (eNauczanie.pg.edu.pl) with nested learnr-based tutorial is presented.

Fig. 4. Data basics – learnr interactive tutorial with code chunks' exercises nested inside the moodle course

2.5. R-Exams

Testing knowledge and skills is very challenging when we are only organizing online versions of exams and assignments. It is especially difficult for courses related to DA. Grün and Zeileis created the solution for that problems [8]. They started to work on the "R-Exams" package, which is a smart system for teachers that supports a one-for-all approach to automatic exams generation. Teachers can create a static or dynamic exercise templates to perform large numbers of personalized exams/quizzes/tests. Exams can be created for various systems: PDFs for classical written exams (with automatic evaluation), import formats for learning management systems (like Moodle, Canvas, OpenOLAT, or Blackboard). All of the static or dynamic exercises has its own syntax inside of the separate Markdown file. It is only up to the skills and imagination of a teacher, how the exercise will be organized in the online quizzes (data generation, solutions' calculations, etc.). With R-Exams teachers may create single choice questions, multiple choice questions, true/false questions, open questions, mixed-type (Cloze) questions. The mechanism of the R-Exams questions is always the same and allow us to randomly generate data, then questions in the data bank of the LMS system. Single Markdown file for R-Exams questions include: code for generating data, question text, solution, metadata (settings). One sample question and its R-Exams syntax is presented on the figure 5.

Fig. 5. R Exams syntax – exercise - t-Student test for the single sample: data generation, question, solution and meta-information

After knitting this question, and generating hundreds of its versions using "examstomoodle" function in R, teachers may include in the LMS courses multiple versions of the same tasks without actually preparing them (just once). On the figure 6. we can see the knitted preview of that exercise (students view).

Fig. 6. R Exams based exercise - t-Student test for the single sample: preview.

Of course the solution usually is visible after the exam is closed (if configured).

2.6. Course management

Finally, we can wrap our innovative content into an online course. If our university has its own learning management system like Moodle, Canvas or OpenOLAT you can include and even nest your reports, tutorials, ebooks and exams inside it. Please note, that it can be used in a blended learning or traditional learning form as well. Otherwise, we can freely create and manage our own course and batch of students using i.e. GitHub Classroom systems or other platform like DataCamp (free after registering) or just use GitHub to exchange reports and treat is as repository.

There is one more solution – "ghclass" package. This package is designed to enable instructors to efficiently manage their courses on GitHub (outside of the LMS platform). It has a wide range of functionality for managing organizations, teams, repositories, and users on GitHub and helps automate most of the tedious and repetitive tasks around creating and distributing assignments. A good way to structure our classroom on GitHub may be: one organization (course) per class or one repo per student (or team) per assignment. If we teach our online DA course at the university, this means one semester of a DA course, this would be one workshop. We can also set up teams on GitHub and each team can be given similar repository for team assignments. With ghclass package we can organize authentication, content and manage the students progress, give feedback and grade their projects. This may be the great opportunity for teachers and students very familiar and attached to GitHub repos, without involving LMS courses directly.

3. SUMMARY

Blended-learning and e-learning based data analysis courses are getting more and more popular among university teachers and students around the world. However it is very difficult to design it appropriately, create a content that will satisfy learners and guide them properly throughout the whole curriculum. It is a challenging task to organize online data repositories, lecture notes, instruction manuals, exercises and finally exams or assignments. Thanks to a variety of modern technologies like web-based IDE or reports, applications and exercises based on the Markdown syntax, interactive books (based on Bookdown) and tutorials (like learnr-based ones) the whole process is now much simplified for DA teachers. It is also very simple to create multiple versions of our questions and exercises without actually writing them so many times using R-Exams package. We can finally manage the whole course with the use of all tools of the LMS platforms or without them using packages like ghclass and working with GitHub development platform. The author is aware that only the most popular solutions are presented in this paper, but certainly with all those innovative ways of teaching our students may achieve their learning outcomes much easier if working in the hybrid mode or using fully online courses.

4. BIBLIOGRAPHY

1. Lim, C. P. (2002). Trends in Online Learning and Their Implications for Schools. Educational Technology, 42(6), 43–48. http://www.jstor.org/stable/44428792
2. Anoush, M., Bianco, M., Littlejohn, A., (2015). Instructional Quality of Massive Open Online Courses (MOOCs). Computers & Education, 80, 1. doi:10.1016/j.compedu.2014.08.005.
3. Nilson L. B., Goodson, L. A. (2017). Online Teaching at Its Best: Merging Instructional Design with Teaching and Learning Research. Jossey-Bass, 1119242290.
4. Miller M. D. (2016). Minds Online: Teaching Effectively with Technology. Harvard University Press, 2016, 0674660021.
5. Xie, Y., Allaire, J. J., Grolemund, G. (2019). R Markdown. The Definitive Guide. Boca Raton, Florida: Chapman; Hall/CRC.
6. Xie Y (2021). bookdown: Authoring Books and Technical Documents with R Markdown. R package version 0.23, https://github.com/rstudio/bookdown.
7. Schloerke, Barret, JJ Allaire, and Barbara Borges. 2020. Learnr: Interactive Tutorials for r. https://CRAN.R-project.org/package=learnr.
8. Grün B, Zeileis A (2009). Automatic Generation of Exams in R. Journal of Statistical Software, 29(10), 1–14. doi: 10.18637/jss.v029.i10.

NOWE TRENDY W NAUCZANIU NA ODLEGŁOŚĆ: WPROWADZANIE NOWYCH ROZWIĄZAŃ DO KURSÓW ANALIZY DANYCH

Obecnie analiza danych w różnym środowisku jest zwykle małym dziełem sztuki. To samo dzieje się z procesem dydaktycznym w tej dziedzinie (np. statystyką, ekonometrią itp.). Przyczyną tego stanu rzeczy jest nie tylko fakt, iż w dydaktyce wykorzystujemy e-learning, distance-learning, lub tylko narzędzia online wspomagające co nieco tradycyjne nauczanie. Stopień skomplikowania problemu oraz jego techniczne i programistyczne rozwiązania stanowią spore wyzwanie. Stąd oczekiwania studentów są większe niż dotąd, a samo tradycyjne wprowadzenie teoretyczne do problemów analitycznych (wykład) i ich rozwiązywanie (laboratorium, ćwiczenia) nie jest wystarczające. Instruktor powinien w swojej pracy laboratoryjnej wykorzystywać szereg narzędzi interaktywnych, tj. zadań, tutoriali, instrukcji, egzaminów online. W tej pracy zaprezentowane zostaną najpopularniejsze i najnowsze rozwiązania techniczne z zakresu e-nauczania analizy danych z wykorzystaniem języka R (i/lub Pythona).

Słowa kluczowe: markdown, bookdown, r-exams, tutoriale, analiza danych, moodle, e-learning, github.