

文系学生のためのデータサイエンス教育の事例報告

ゴン ルイ
(教養教育機構)

A Report on a Case Study of Data Science Education for Humanities Students

Rui GONG

(Organization of Liberal Arts Education)

Humanities undergraduates have few mathematics and data-related courses in the classical training program plan. However, with the development of information technology, the moment they graduate and enter the workplace, regardless of whether they have training experience in data science, they have to face processing a vast amount of data in any job. This case was conducted by incorporating automated data processing knowledge into the basic and applied informatics courses to enhance the competitiveness of humanities undergraduates in their future careers. The significance of this case will be that undergraduates without mathematical backgrounds can use data science freely.

Keywords: Data Science, Excel, Humanities, VBA, Undergraduates

1. Introduction

Recently a scientific word has suddenly burst onto our scene and has become popular at a breakneck pace, and this word is data science. Data science is the science of using statistics, algorithms, and other tools to solve real-world issues. Although John Tukey launched the concept in the 1960s, it has become an independent and influential science until recently (Tukey, 1962). Today, data science is an essential component of industry, agriculture, education, medicine, defense, and many other fields (Igual & Seguí, 2017).

Because data science is so strategic, countless universities and colleges have considered it an important part of the composition of modern higher education. However, building data science into the education curriculum is easier said than done because of the complexity and depth of the

background of knowledge required. In addition to the above, computer hardware and a professional instructor suitable for data science are essential.

Japan, as a developed country, started early on in data science education. The government has not only included informatics in university entrance exams in 2025 but has also spared no effort in funding supercomputers and personal computers to be set on campus (Kentaro Komagata, 2021). Mejiro University, which is a major humanities university, is also introducing data science into the new information education system starting in the spring semester of 2022. The following is how Mejiro University offers an entry-level data science course for humanities students who are short on mathematics.

A suitable environment for the program is essential for the smooth operations of data science courses. The Center for Information and Computer

Education at Mejiro University provides the perfect hardware and software setup. Considering that only light data processing is covered in this course, a list of recommended hardware for PC is given below.

Table 1 Recommended hardware of PC

Processor	Core i5 11th Gen
Graphic Card	GeForce RTX 3060
Memory	8GB DDR4-2666MHz
Storage	SSD 1TB
Others	DVD Optical drives
Operating System	Windows 10 Pro 64bit

The computer hardware composition shown in Table 1 is a conservative recipe and does not boldly use the latest versions of CPU or OS. Since this is for undergraduate students majoring in humanities, the software part uses the student version of Office 365. Among the core software are Excel, Access, and the free version of PowerBI (*Office365*). The learning management system (LMS) uses Google Classroom, which is owned by Google. Once the appropriate learning environment is established, a teaching plan will be created and carried out. The core goal of this course is to provide students with an overview of the most basic data science-related knowledge, so the teaching plan is shown below.

2. Teaching Plan

2.1 Teaching Direction

As an optional course, there are certain requirements for students' foundation. It is also stated in the course syllabus that the minimum requirement is to have experience in using Excel because this course is only 15 lessons. In addition, the course is open to all students in grades 2-4, and there are three directions for their further study in this course.

As shown in Figure 1, this course offers both basic and advanced training in the three

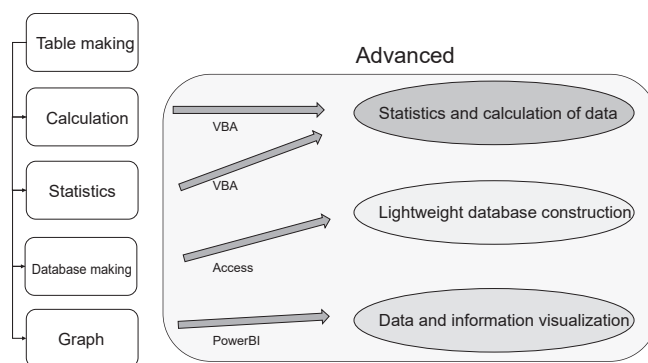


Figure 1 Direction of this course

components of data science, calculations and statistics, databases, and data visualization, as the direction of development. Students are required to complete the basic part of the course, and the advanced part is not forced.

2.2 Textbook and Teaching Methods

A large number of reference books were read during the preparation of this course, but most of them were found to be unsuitable for direct use as textbooks in this course for data science-related books. Three reasons are listed here.

First, even books that claim to be introductory have a mathematical background and programming fundamentals required.

Secondly, the examples and exercises provided in the book are detached from the real life of the students. For example, some books give a lot of examples of business cases, which is difficult for students who have not been in touch with real business.

Thirdly, Office365 will be minorly updated once a month and majorly updated once every half a year, and there is no way for the publication speed of books to catch up with the update speed.

Therefore, this course takes original lecture notes, and books are used as reference books only. Here we especially recommend *The Field Guide to Data Science* written by *Booz Allen Hamilton*, which is open source and can be read online for free.

At the end of this subsection, a brief introduction to Excel and VBA teaching syllabus. As shown in Table 2, the 15 lessons were taught as follows.

Table 2 The teaching syllabus with Excel as the basic part and VBA as the advanced part

Basic part	1 st	Guidance
	2 nd	Quadratic operations by Excel functions
	3 rd	Application of Excel functions with conditional logic theory
	4 th	Edit and organize data based on Excel functions
	5 th	Creating and editing tabulation tables based on Excel data
	6 th	Creating graphs using tabulation tables
	7 th	Basic statistical tools and summary of Excel functions
Advanced part	8 th	How to work with the VBA programming language within Excel
	9 th	Basic understanding of VBA programming language within Excel
	10 th	Manipulation of data using variables by VBA
	11 th	Manipulate data using conditional branching by VBA
	12 th	Manipulation of data using counters by VBA
	13 th	Manipulate data using iterative processing by VBA
	14 th	Drawing graphs using VBA programs
	15 th	High-dimensional array and summary of VBA

2.3 Teaching Methodology in Basic Lessons

There is nothing special about this course in terms of teaching methodology, which still uses the well-known circular learning model. The model unique to this course is shown in the figure below. Each lesson begins with a preliminary study, where students need to pre-read proper nouns and contents of the lecture note.

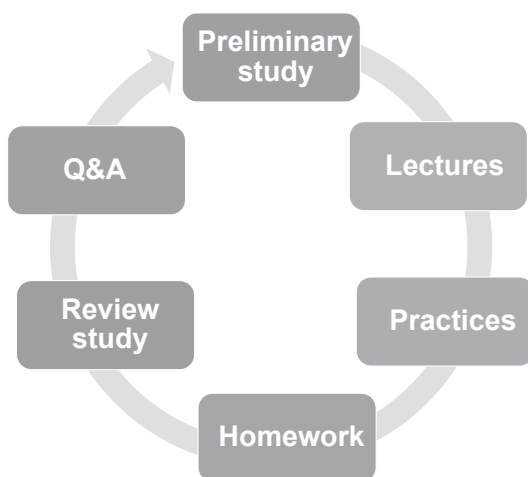


Figure 2 Teaching Methodology

The teaching time is 90 minutes, in which lectures and practices are alternated. The lecturer will start by explaining the main points of a lesson and then give 2-3 example problems as a demonstration. Afterward, there will be similar

practices waiting for students to complete. At this time, the lecturer will give individual instructions according to each student's different pace.

In order to reinforce knowledge, students will be required to complete homework after the lesson. The homework will be slightly more difficult than the previous practice problems, but students are allowed to complete homework assignments by looking up references.

The review study is not enforced, but it is still recommended that students review their past lessons on a regular schedule. If there is any point that students cannot understand, they can always seek answers from the lecturer through the LMS.

2.4 Teaching Methodology in Advanced Lessons

Advanced lessons will be a little more stressful than the basic lessons, as there will be an emphasis on lessons slanted toward practice and skill enhancement. Students must complete assigned tasks by study. The following will explain the teaching process of VBA through an example.

As shown in B of Figure 3, the lecturer gives the lesson goal and starting data, and the students achieve it through VBA code. Students can look up past lecture materials and use the Internet to find other people's code for reference when

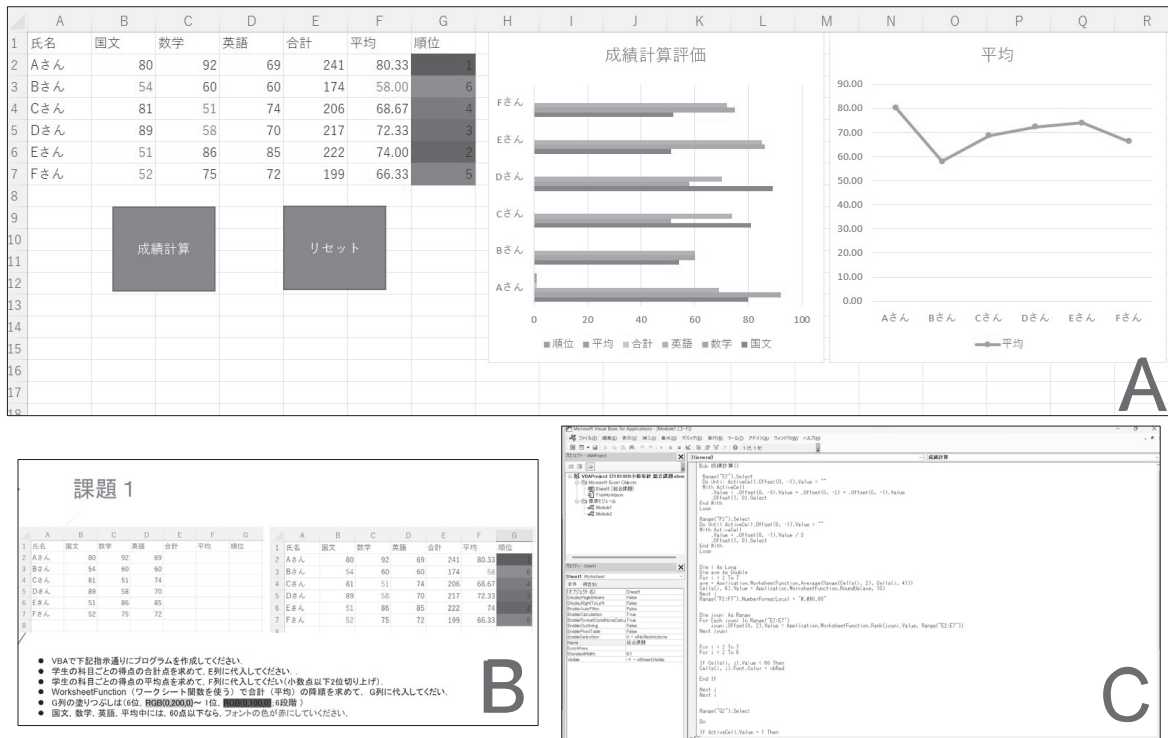


Figure 3 An example of VBA teaching

constructing program structures and writing code. A in Figure 3 is an assignment completed independently by a 2nd-grade student following the requirements of B. The program allows for ranking and charting of individual subject performance via control buttons. C in Figure 3 is the VBA code.

As can be seen in Figure 3 as a whole, the students successfully implemented the code for this function of the grade calculator. Although the construction of the code is still relatively raw and imperfect, the basic functions are complete. These practice questions for advanced lessons certainly lay a foundation for students who are new to coding and boost their confidence to learn data science while developing an interest in coding, which is one of the goals of this course.

2.5 The career advice base on this course

Students are expected to learn various theories and techniques at the university, and it is essential to develop career-related knowledge.

This course is also no exception, even as an entry to data science. However, unlike other fields,

computer science is rapidly evolving, and students will be given career advice that is slightly ahead of the computer skills they must have to graduate. The career advice for this course consists of the following three parts.

The first part is some advice for informatics-related qualifications. Usually, as a college student who is about to graduate and enter the career field, it is possible to be more competitive by having some skills and qualifications. Since the digital office is now a part of all industries, having a qualification in the field of computer science will make for a successful career. Since this course is based on the Office software developed by Microsoft, the Excel and Access expert certificate issued by Microsoft is a good choice.

The second part is for students who do not intend to enter the computer science industry but only use computer skills as a supplement. The lecturer will give advice on software or electronic tools that are relevant to the student's interests and future career visions. For example,

for students who want to be graphic designers in the future, the lecturer teaches students how to download and use some free and open-source design software.

The third part of the advice is aimed at students who are seriously considering entering the IT industry. In addition to developing their logical minds in computer science, these students are assigned a study plan. Although the program of study is a bit too simple for an actual project, if the student can finish the program successfully, it will be very beneficial for their future IT career.

3. Feedback from students

This course was implemented in the spring semester of 2022 a total of 15 times. There are seven basic lessons, seven advanced lessons, and one small exam. Since current 2nd-4th grade students have not taken entry-level data science courses, the advanced lessons are taught in VBA. After 15 lessons were completed, a questionnaire was administered to the students.

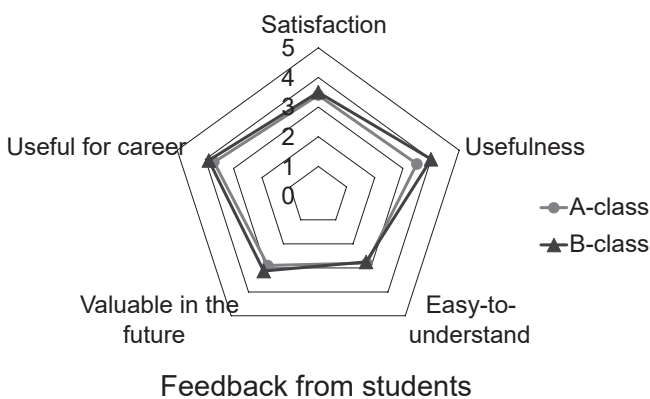


Figure 4 The feedback from students

The questionnaire consisted of 12 questions, and the two classes scored 3.12 and 3.22 out of a possible 5 points. Five of the 12 questions were related to instructional content and were pulled in Figure 4. As seen in Figure 4, the difficulty of the course content that scored the least, students generally found the course a little difficult.

In addition, the instructor also heard from the

students that the pace of teaching was a bit too fast and the amount of content in the advanced courses was too broad.

4. Discussion

This course is set up to teach data science to improve the competitiveness of humanities students. This course is offered as an option for students in grades 2-4 who are currently enrolled in school. After a total of 15 basic and advanced lessons, the results of the questionnaire showed that the students thought the course was too difficult. The authors consider that there are three reasons for the above issues.

First, the transition between the basic and advanced lessons is not very well done. The previous basic lessons were too easy, and the later advanced lessons became more difficult. Therefore, the lecturer is considering adding VBA programming-related knowledge to the next year in basic lessons.

Second, there is a large gap in the knowledge base of mathematics and logic among the students. For the same conditional statements, some students can understand them without any problem, and some students cannot understand them no matter what. Therefore, the lecturers decided to introduce a new textbook in the next year to assist the students in understanding logic in programming.

Finally, students generally felt that the amount of practice was a bit too much. The next thing to think about is how to ensure that students consolidate what they have learned while reducing the amount of practice and homework.

5. Conclusion

This report describes a course that teaches humanities students data science-relevant knowledge and skills. Students generally agreed that the course is useful and helpful for their future careers, but still felt that the course content was

a bit difficult. In the future, teaching data science content more rationally and simply to humanities students will continue to be challenging.

《References》

- Igual, L., & Seguí, S. (2017). *Introduction to Data Science*. Springer International Publishing.
<https://doi.org/10.1007/978-3-319-50017-1>
- Kentaro Komagata. (2021). *Explanation of the "Information" entrance examination for the common test starting in 2025 (in Japanese)*. <https://project.nikkeibp.co.jp/pc/atcl/19/06/21/00003/100400271/office365>. Retrieved September 24, 2022, from <https://www.office.com>
- Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, 33 (1), 1-67.
<https://doi.org/10.1214/aoms/1177704711>
- (受付日：2022年10月10日、受理日：2022年11月25日)