

Brief Paper:

An Analysis of Curricula for Data Science Undergraduate Programs

Soosun Cho^{1*}

Abstract: Today, it is imperative to educate students on how to best prepare themselves for the new data driven era of the future. Undergraduate education plays an important role in providing students with more Data Science opportunities and expanding the supply of Data Science talent. This paper surveys and analyzes the curricula of Data Science-related bachelor's degree programs in the United States. The 'required' and 'elective' courses in a curriculum for obtaining a B.S. degree were evaluated by course weight to indicate its necessity. As a result, it was possible to find out which courses were important in Data Science programs and which areas were emphasized for B.S. degrees in Data Science. We found that courses belong to the Data Science area, such as data management, data visualization, and data modeling, were more required for Data Science B.S. degrees in the United States.

Key Words: Data Science, Education, Curricula Analysis, Undergraduate Programs, Required Courses.

I. INTRODUCTION

Now a days, Data Science is affecting almost every sector of the industry. Work across nearly all domains is becoming more data driven, affecting both available jobs and required skills. Accordingly, the demand for Data Science talent is growing rapidly, and academic institutions are actively offering relevant degree programs.

In particular, Data Science is an emerging discipline that was born in the United States, so the effort of universities in the U.S. to provide educational services are leading the way. 'Discover Data Science [1]' reported that five years ago, a bachelor's degree in Data Science was nearly nonexistent but now, each academic year, colleges and universities are adding new Data Science programs (DS programs). In fact, there are over 50 schools across the U.S. offering a Data Science major. Commonly, the four-year undergraduate DS programs are offered by either the Computer Science, the Mathematics, the Statistics, or newly established

Data Science department. It is also provided as an interdisciplinary program in collaboration with multiple departments [2].

According to the Korea Council of University Educations [3], as of 2022, there are 18 Data Science-related recruitment units at four-year universities in Korea, including AI·Data Department at Korea National University of Transportation. Unlike the U.S. universities, Data Science-related undergraduate programs are still in their infancy in Korea. The newly established Data Science department is facing difficulties in creating an essential curriculum and course objectives.

This paper introduces the results of survey and analysis of the curricula of Data Science undergraduate programs in the United States. In particular, we investigated the characteristics of the curriculum according to the hosting department of DS program. Also, we analyzed how much emphasis is placed on courses belong to the Data Science area, such as data management, data visualization, and data modeling. It is intended to be helpful to understand Data Science undergraduate programs and to develop them for better education in this attractive new discipline.

II. RELATED WORKS

The Data Science programs are multidisciplinary and a single subject domain is not enough to cover the magnitude of content and skills needed for DS programs [2]. In the United States, most of today's DS program courses cover three foundation areas; Mathematics, Statistics, and Computer Science [4]. In addition to the foundation courses, several Data Science curriculum guidelines emphasize Data Science specific courses such as data curation, data visualization, data modeling, and data mining [4-6]. In addition, it is strongly recommended to include courses outside of Mathematics, Statistics, and Computer Science to accommodate a wide range of Data Science applications [4-5]. Also, the 'National Academies of Sciences, Engineering, and Medicine [4]' focuses on developing data insights

Manuscript received March 10, 2022; Revised April 20, 2022; Accepted April 27, 2022. (ID No. JMIS-22M-03-008)

Corresponding Author (*): Soosun Cho, +82-31-460-0584, sscho@ut.ac.kr

¹Major of Data Science, Korea National University of Transportation, Uiwang, Korea, sscho@ut.ac.kr

through a curriculum including important Data Science concepts, applications to real-world problems with an understanding of limitations, and ethical issues related to Data Science. It is widely adopted as a guideline for curriculum in Data Science undergraduate program [7].

According to the ACM Taskforce on Data Science Education [8], since industry employers require more computing skills than statistical or mathematical skills, undergraduate programs should strengthen students' computing skills.

The authors in [2] reported 107 Data Science academic programs in the United States. Although DS programs can be introduced as graduate, undergraduate, certificate and minor, they only considered the regular programs at B.S., M.S., and Ph.D. levels with 49, 45 and 13, respectively. The 49 B.S. programs consists of 25 programs from Doctoral Universities, 16 programs from Master's Colleges, and 9 programs from Baccalaureate Colleges. They also found the distribution of the hosting departments of 49 B.S. programs. The standout departments were Computer Science and Data Science with 20 and 8, or 40% and 16%, respectively.

In the research [6], the 101 B.S. and B.A. degrees in Data Science were reviewed. They showed that almost all of the DS programs included introductory statistics, and more than 90% of the programs included introductory computer programming.

In addition to reviews of the distribution of the hosting departments or subject domains on DS programs, research [7] quantified the relative amount of coursework in three categories: computer science, statistics/mathematics, and domain knowledge. They investigated 16 Data Science undergraduate programs from Doctoral Universities in the U.S. For the quantitative evaluation of DS programs, they adopted an explicit framework describing the components of Data Science education; the framework presented in the National Academies of Sciences, Engineering & Medicine (NASEM) report [4]. For the framework, they developed a rubric and coded the undergraduate Data Science curricula on a four-point scale using direct survey methodology indicating the familiarity with a topic that could be expected from a student graduating from the program. As the results, training in Computational Foundations was one of the highest scoring areas. Most programs also scored well in Statistics and Mathematics training.

Because the DS programs are multidisciplinary many curricula have balanced composition in the disciplines of Computer Science, Mathematics and Statistics. In recent years, however, independent Data Science-related departments have rapidly emerged to offer Data Science degree programs. This paper analyzes how the curriculum differs for independent Data Science departments.

III. ANALYSIS OF CURRICULA IN DATA SCIENCE PROGRAMS

3.1. Target Programs for Analysis

For the curricula analysis, we focused on the 40 campus-based B.S. programs in 'Data Science Colleges and Universities [9]'. Like the research [2] we chose the regular B.S. programs in Data Science. And unlike the research [7], we investigated all Data Science undergraduate programs from Doctoral Universities, Master's Colleges and Baccalaureate Colleges in the U.S. See Appendix for details.

The distribution of the hosting departments is shown in Fig. 1. The 40 B.S. programs were offered from 11 Computer Science, 8 Data Science, 6 Mathematics, and 5 Statistics departments. Similar to research [2], the prominent departments were Computer Science and Data Science, accounting for 27.5% and 20%, respectively.

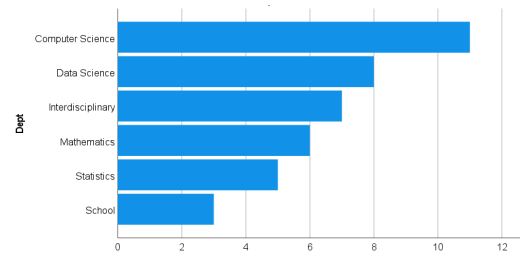


Fig. 1. Departments offering 40 B.S. programs in Data Science.

3.2. Curricula Analysis

For the analysis of courses in DS program, 870 course titles were collected from 40 programs' course credits of B.S. degrees. In many cases, course titles vary from program to program. For example, courses related to 'Programming' have different titles such as 'Introduction to Programing', 'Applied Programming', 'Programming Methods', etc. These courses renamed to 'Programming'. In order to preserve the original meaning of the courses, minimal changes were made.

Table 1 shows 17 popular courses which appear more

Table 1. Popular courses in Data Science B.S. programs.

Courses	Freq.	%	Cumul. %
Calculus	48	5.8	5.8
Linear algebra	36	4.3	10.1
Statistics	34	4.1	14.2
Data science	32	3.8	18.0
Programming	23	2.8	20.8
Computer science	22	2.6	23.4
Database systems	21	2.5	25.9
Machine learning	21	2.5	28.5
Probability	20	2.4	30.9
Discrete mathematics	17	2.0	32.9

than 10 times in 870 cases.

In order to define subject area for analysis of DS curricula, we used the guidelines presented in the NASEM report [4] as research [7]. The NASEM guidelines lists 10 key concept areas for describing the components of Data Science education. As shown in Table 2, we made up the 5 subject areas (or disciplines) from the 10 concept areas.

Table 2. 10 NASEM concept areas and 5 subject areas for analysis.

10 concept areas	5 subject areas
Mathematical foundations	Mathematics (MATH)
Computational foundations	Computer science (CS)
Statistical foundations	Statistics (STAT)
Data management and curation	Data science (DS)
Data description and visualization	
Data modeling and assessment	
Workflow and reproducibility	Others
Communication and teamwork	
Domain-specific considerations	
Ethical problem solving	

To find the distribution, all courses were grouped in five subject areas in Table 2. Then, a Cross-Tab Analysis was performed using two variables, 'Department' and 'Area'. Fig. 2 represents that all departments have the highest number of courses in CS area and the second is STAT except for 'Mathematics' department.

The ACM Taskforce on Data Science Education [8] states that undergraduate programs should strengthen students' computing skills because industry employers require more computing skills than statistical or mathematical skills. The result of the analysis show that the surveyed programs are meeting this need.

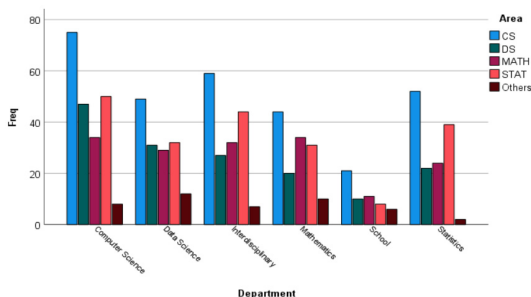


Fig. 2. Distribution of course areas along to departments.

3.3. Analysis of Required Courses

All courses in DS programs are of varying importance depending on 'required' or 'elective'. Moreover, among 'elective' courses, some of them can be chosen from two

courses, others from three or more courses. For example, in the required courses of one DS program, if three or more courses appear after a sentence such as "Take one of the following courses:", these courses can be chosen from three or more courses. The necessity of a course can indicate its importance in DS programs. So, we give a 'weight' to each course according to its importance. In the weight range, '3' means 'required', '2' means 'elective' from two, and '1' means 'elective' from three or more.

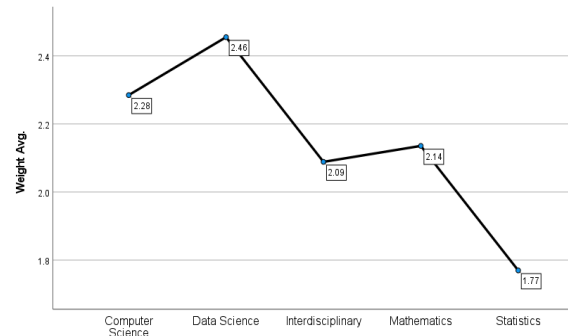


Fig. 3. Graph of course weight averages in 5 different departments.

After assigning weights to each course, we compared how the weight average varies by hosting department. As shown in Fig. 3, courses in the department of Data Science have the highest weight average. This shows that Data Science departments have more compulsory courses than other departments. It can be thought that more effort is needed in the new Data Science department than in the traditional department to get a B.S. degree in Data Science.

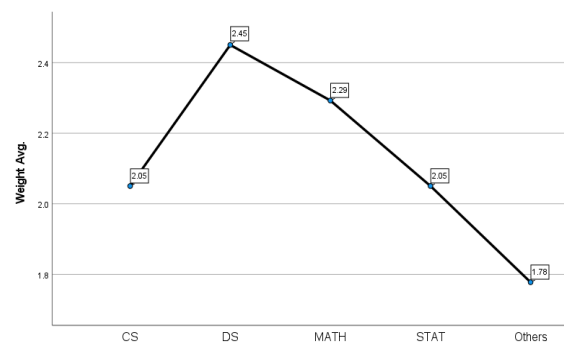


Fig. 4. Graph of course weight averages in five subject areas.

To determine the importance of subject areas, i.e. disciplines in Table 2, we compared how weight averages varied across five disciplines. As shown in Fig. 4, courses in DS area have the highest weight average. This means that across all departments, there are many required courses in DS area rather than in the traditional area such as CS, MATH, or STAT. Contrary to research [7] showing that Computational Foundations was one of the highest scoring

areas, we found that Data Science was the highest scoring area with course weight indicating its necessity.

IV. CONCLUSION

Due to the expansion of the demand, growing number of U.S. universities are offering undergraduate programs in Data Science. In this paper we analyzed the characteristics of the curricula for Data Science B.S. programs in the U.S. We investigated the characteristics according to the hosting department of DS program. Also, we analyzed how much emphasis is placed on courses belong to Data Science area. For quantitative evaluations, we used course weight to indicate its necessity.

First, as shown in Fig. 2, all departments have the highest percentage of CS area which means that the largest number of CS courses are being offered in 40 programs. This result shows that the surveyed programs are meeting the need from the ACM Taskforce on Data Science Education [8]. The need was that undergraduate programs should strengthen students' computing skills.

Second, when weight were given according to the necessity of each course and then the averages were compared by department, the weight average of the Data Science department was found to be the highest as shown in Figure 3. Although the number is smaller than that of Computer Science as shown in Fig. 1, the Data Science department is a newly established department that fits the Data Science discipline and growing fast. Compared to the existing departments offering DS programs, these new departments are operating a larger scale of required courses to nurture talents suitable for their purpose.

Third, when course weights were compared by disciplines, the weight average of DS area was found to be the highest as shown in Fig. 4. This shows that although many MATH-related courses are placed in foundation courses, the required courses are more in DS area. And unlike to research [7], Data Science was the highest scoring area with course weight indicating course necessity. From the result it can be suggested that courses belong to Data Science area, such as data management, data visualization, and data modeling should be treated with great importance in the Data Science B.S. curriculum.

The limitation of this curriculum analysis is that it did not include all courses of Data Science programs. Capstone courses required in most programs were excluded, and courses related to Data Science application fields recommended as electives were also excluded from the analysis. Case analysis studies that include a wider range of courses are needed in the future.

ACKNOWLEDGEMENT

The research was supported by a grant from the 2021 program for visiting professors overseas in Korea National University of Transportation.

REFERENCES

- [1] Discover Data Science, Bachelor Degree in Data Science - Guide to Choosing a Great Program, 2022. <https://www.discoverdatascience.org/programs/bachelors-in-data-science/>
- [2] I. B. Hassan and J. Liu, "Data science academic programs in the U.S.," *The Journal of Computing Sciences in Colleges*, vol. 34, no. 7, pp. 56-63, 2019.
- [3] KCUE, University Information Portal, 'adiga', 2022. <https://adiga.kr/EgovPageLink.do?link=EipMain>
- [4] National Academies of Sciences, *Engineering and Medicine, Data Science for Undergraduates: Opportunities and Options*. Washington, DC: The National Academies Press, 2018.
- [5] R. De Veaux, M. Agarwal, M. Averett, B. S. Baumer, A. Bray, and T. C. Bressoud, et al, "Curriculum guidelines for undergraduate programs in data science," *Annual Review of Statistics and Its Application*, vol. 4, no. 1, pp.15-30, 2017.
- [6] I. B. Hassan, T. Ghanem, D. Jacobson, S. Jin, K. Johnson, and D. Sulieman, et al, "Data science curriculum design: A case study," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (SIGCSE '21)*. Association for Computing Machinery, New York, NY, pp. 529-534, Mar. 2021.
- [7] J. C. Oliver and T. McNeil, "Undergraduate data science degrees emphasize computer science and statistics but fall short in ethics training and domain-specific context," *PeerJ Computer Science*, vol. 7, 2021.
- [8] A. Danyluk, P. Leidig, S. Buck, L. Cassel, A. McGettrick, and W. Qian, et al, ACM Data Science Task Force: Computing Competencies for Undergraduate Data Science Curricula, 2019, <https://dstf.acm.org/>
- [9] RyanSwanstrom, 2022. Data Science Colleges and Universities, <https://ryanswanstrom.com/colleges/>

APPENDIX

Target program list from ‘Data Science Colleges and Universities [9]’.

School	Program	Department
1 California Polytechnic State University	Data Science Minor	College of Science and Mathematics
2 University of California-Irvine	Data Science	Statistics
3 University Of San Francisco	Data Science	Department of Mathematics and Statistics
4 Yale University	Statistics and Data Science	Department of Statistics and Data Science
5 Florida Polytechnic University	Data Science	Data Science and Business Analysis
6 Chaminade University	Data Science	School of Natural Sciences and Mathematics
7 Luther College	Data Science	Computer Science
8 Brigham Young University - Idaho	Data Science	Department of Mathematics
9 University of Evansville	Statistics and Data Science	Mathematics
10 Indiana University-Purdue University Indianapolis	BS in Applied Data Science	School of Informatics and Computing
11 Valparaiso University	Bachelors of Science in Data Science	Interdisciplinary
12 Northern Kentucky University	Data Science	Computer Science
13 Worcester Polytechnic Institute	Data Science	Data Science
14 Smith College	Statistical & Data Sciences	Statistical & Data Sciences
15 University of Michigan-Ann Arbor	Data Science	Computer Science & Engineering
16 Winona State University	Data Science	Department of Mathematics and Statistics
17 Elon University	Data Science Minor	Computer Science
18 University of Nebraska at Omaha	Data Science Concentration	Mathematics
19 University of New Hampshire	Analytics & Data Science	Applied Engineering & Sciences
20 Thomas Edison State University	Data Science and Analytics	Heavin School of Arts & Sciences
21 New York University	Applied Data Analytics and Visualization	School of Professional Studies
22 Fei Tian College	Data Science	Data Science
23 University of Rochester	Data Science	Interdisciplinary
24 Case Western Reserve University	Data Science	Data Science
25 Denison University	Major in Data Analytics	Data Analytics
26 Miami University of Ohio	Data Science and Statistics	Statistics
27 The Ohio State University	Data Analytics	Interdisciplinary
28 Pacific University	Data Science	Mathematics and Computer Science
29 Juniata College	Data Science	Data Science
30 University of Texas at Dallas	Bachelors of Science in Data Science	Interdisciplinary: Mathematical Science/Computer Sciences
31 Pennsylvania State University (Computational Data Sciences)	Data Sciences	Interdisciplinary: Statistics/Computer Sciences/Information Science & Tech
32 Pennsylvania State University (Applied Data Sciences)	Data Sciences	Interdisciplinary: Statistics/Computer Sciences/Information Science & Tech
33 Pennsylvania State University (Statistical Modeling Data Sciences)	Data Sciences	Interdisciplinary: Statistics/Computer Sciences/Information Science & Tech
34 Elizabethtown College	Data Science	Computer Science
35 College of Charleston	Data Science	Computer Science
36 Augustana University	Data Science	Computer Science
37 Westminster College	Minor in Data Science	School of Arts and Sciences

38 George Mason University	Computational Data Sciences Minor	School of Physics, Astronomy, and Computational Sciences
39 University of Mary Washington	Data Science Minor	Department of Computer Science
40 University of Wisconsin-River Falls	Data Science and Predictive Analytics	Department of Computer, Information, and Data Sciences
