Graduate Certificates in Data Science: Miss or Match for Data Scientists' Jobs in Logistics and Operations?

Catherine Chen*
Ball State University, Muncie, Indiana, USA

Haoqiang Jiang
Northern Kentucky University, Highland Heights, Kentucky, USA

With the explosion of the Internet of Things (IoT), big data and data science have created unprecedented career opportunities in all fields. To help professionals in logistics, supply chain, and operations management to determine if graduate data science certificate was a valid option to gain needed data-related skills for the fields, this study analyzed the contents of graduate certificates and the skills needed for data scientists in the fields to reveal if the certificates covered the skills needed. The study found that a high percentage of data-related skills in logistics, supply chain, and operations management were covered in graduate certificates and that certificates in the Business area had the highest percentage of 'match' between skills needed and course topics.

* Corresponding Author. E-mail address: cchen@bsu.edu

I. INTRODUCTION

With the IoT, the number of devices connected to the internet had reached 22 billion worldwide at the end of 2018 and was estimated to reach 38.6 billion by 2025 (Mercer, 2019), which was in line with IHS Markit's projection of 30.7 billion by 2020 (Nordrum, 2016); as a result, the amount of data collected has been exploding. The phenomenon was coined as 'Big Data' with high volume, high velocity in which the data is generated, and a greater variety of data formats (De Manro, Greco and Grimaldi, 2016; McAfee, Brynjolfsson and Davenport, 2012; Watson, 2014). This phenomenon prompted new job demands and a new job title: data scientists (Davenport and Patil, 2012; Pranavathiyani, 2017). However, Donoho (2017) contended that data science was not a new phenomenon with big data; instead, it had roots in statistics. Others asserted that data science was a "concept to unify statistics, data analysis, machine learning and their related methods in order to understand and analyze actual phenomena with data" (Hayashi, 1998). Regardless if data science was a new field, it received a boost in the early 2000s from the aforementioned 'big data' phenomenon.

With the growth of big data, the demand for data scientists is expected to grow. In 2016, McKinsey estimated the need for 1.5 million managers and analysts to analyze big data for decision making by 2018 (Manyika et al., 2011). Because the environment is dynamic and competition is high, organizations that want to tap into the potential of big data will need professionals who can manage the four V's issues of big data to support decision making. Due to the challenges and complexity of the data, a successful data scientist usually requires training from multiple disciplines (Chen and Jiang, 2018). The fundamental responsibility of a data scientist is to find meaningful results while "swimming in data." Such work requires people to identify different data resources, to collect and consolidate these raw data, to clean the data and manage potential error and missing data, and to explore a different type of analysis for decision-maker; some of these tasks might require coding ability (Davenport and Patil, 2012).

To respond to the high job market demand, many universities have created new data science programs or infused big-data related curricula into existing ones (Horton, Baumer and Wickham, 2015, Watson, 2014). A recent large investment in data science in academia was a $100 million 'Data Science Initiative' made at the University of Michigan with a plan to hire 35 new faculty (Donoho, 2017). Data science programs are offered in both graduate and undergraduate levels.

In the initial stages of data science program development, the focus was placed at degree programs, not certificates. Although graduate certificate as a data science option was mentioned in articles (Cao, 2017; Song and Zhu, 2017; Chen, Chiang and Storey, 2012), few studies have investigated the current status of graduate certificates in data science. With the projected need of 1.5 million managers who would need data analytics skills, graduate certificates seemed to be a valid alternative for working professionals who may not want to commit to a full-time, two-year graduate degree program to update their skills. Graduate-level certificate programs can play a critical role in helping professionals with business knowledge and working experience to become data scientists; yet, the certificate would need to provide data analytics skills that complement the students' professions (Chen, Chiang and Storey, 2012).

The field of logistics, supply chain, and operations management have recognized the benefits of using data science (Govindan, Cheng, Mishra and Shukla, 2018; Schoenherr and Speier-Pero 2015). It would be beneficial for professionals in these fields to know if existing graduate certificate programs help them gain the analytic skills that they need. This study aims to provide that information.

II. LITERATURE REVIEW

This section provides an overview of the previous studies about data science; required job skills; existing education programs; data scientist job requirements for logistics, supply chain, and operations; and the research gap.

2.1. The Data Science Field

The voluminous data collected in high velocity and a variety of formats distinguished itself from the traditional data collection and data structure; therefore, the data cannot be processed and analyzed using only traditional tools (Manyika et al., 2011; Zikopoulos, Eaton, Deroos, Deutsch and Lapis, 2012). From activities point of view, data science could be categorized into six different divisions: 1) data exploration, including gathering data, dealing with raw data issue, and exploring potential analysis; 2) transformation, including consolidating raw data from different resources and translating them into structured data; 3) computing, such as using different computer programming languages such as R or Python to do the analysis and process; 4) modeling, such as using statistics knowledge to provide model for descriptive or predictive analyses; 5) visualization, including presenting results into a meaningful, easy-to-understand way by using different graphic representations; and 6) "science about data science" (p.756), which occurs when data scientists find recurring patterns and are able to generalize them for future applications (Donoho, 2017).

Due to the great potential of gaining information and knowledge from big data, data science could be applied to many different fields, such as education, public health, medicine, e-commerce, e-government, security, etc. (Chen, Chiang and Storey, 2012; Donoho, 2017). In these fields, data scientists could use descriptive analysis, predictive analysis, or prescriptive analytics (Dhar, 2012) to support important business decision making (Cao, 2017) and as a consequence, to provide competitive advantages to an organization (Chen, Chiang and Storey, 2012).

2.2. Required Job Skills

Skills in many domains are needed to become a successful data scientist. Gardiner, Aasheim, Rutner, and Williams (2018) conducted a content analysis on 1,216 job advertisements containing 'big data' in the job title and found several skill categories, including analytics, database/data warehouse, large data, emerging technologies, statistical and mathematical, and technologies. On the other hand, Radovilsky, Hegde, Acharya, and Uma (2018) used the keywords 'business data analytics' and 'data science' to collect 1,050 job requirements and qualifications. They identified four skill domains: technical, analytical, business, and communication.

A review of literature identified six categories of skills needed by data scientists.

- ❖ Business: Many articles stressed the importance of data scientists' ability to speak the language of business. For data scientists to advise executives and product managers on the implications of the data for products, process, and decisions, they must understand business operations and have specific business knowledge (Davenport and Patil, 2012; Harris, Murphy and Vaisman, 2013; Harris and Mehrotra, 2014; McAfee, 2012; Phillips-Wren, Iyer, Kulkarni and Ariyachandra, 2015).
- ❖ Data-related skills: These skills include data storage, data access, and data manipulation. Data scientists should be able to handle large amounts of data (Harris, Murphy and Vaisman, 2013; McAfee, Brynjolfsson and Davenport, 2012; Phillips-Wren, Iyer, Kulkarni and Ariyachandra, 2015; Watson, 2014).
- ❖ Statistics, machine learning, and model building: Instead of the traditional hypothesis-based approach to statistical analysis, big data analysis is more likely to involve machine learning (Gardiner, Aasheim, Rutner and Williams, 2018). Many agreed that analytical abilities are a key element of a data scientist's job and that they need a firm grounding in statistics and should be fluent with a wide range of statistical techniques (Harris and Mehrotra, 2014; McAfee, Brynjolfsson and Davenport, 2012). Others suggested that knowledge in machine learning and modeling complex data is important (Harris, Murphy and Vaisman, 2013; Phillips-Wren, Iyer, Kulkarni and Ariyachandra, 2015). Although statistics are important, key big-data techniques are rarely taught in traditional statistics courses (McAfee, Brynjolfsson and Davenport, 2012).
- ❖ Coding: Some suggested that data scientists should have strong programming skills (Davenport and Patil, 2012; Harris, Murphy and

Vaisman, 2013; Radovilsky, Hegde, Acharya and Uma, 2018). The suggestions include Java, Python, and R (Watson, 2014; Radovilsky, Hegde, Acharya and Uma, 2018).

- ❖ Problem-solving: Data scientists agreed that they often address new and complex problems (Harris and Mehrotra, 2014; Watson, Wixon and Ariyachandra, 2013). A deep understanding of problem formulation is needed to help leaders reformulate their challenges in ways that big data can tackle (McAfee, Brynjolfsson and Davenport, 2012; Pattin et al., 2014; Radovilsky, Hegde, Acharya and Uma, 2018).
- ❖ Communication: Presenting insights to upper management requires good communication skills. Data scientists must be able to communicate in language that all their stakeholders understand (Davenport and Patil, 2012; Phillips-Wren, Iyer, Kulkarni and Ariyachandra, 2015; Watson, 2014). They need to be able to craft narratives, to use visualization tools (McAfee, Brynjolfsson and Davenport, 2012), and to creatively display information (Davenport and Patil, 2012).

Beyond the aforementioned basic foundations and domain skills mentioned, Provost and Fawcett (2013) stressed the importance of following the fundamental principles of data science: 1) well-defined stages are available to solve business problem systematically; 2) results must be evaluated in appropriate business context; 3) problems and solutions can be decomposed into subproblems, 4) technology can be used to analyze large volume of data; 5) similarity can be identified between known and unknown; 6) the data analysis might not be generalizable beyond the current observation; and 7) underlining assumptions are critical in drawing causal conclusions.

2.3. Existing Education Programs

Due to the complexity and generalizability of the data science technology (Davenport and Patil, 2012), several guidelines have been developed to explore teaching strategies and curricula issues (Anderson, Bowring, McCauley, Pothering and Starr, 2014; De Veaux et al., 2017; Horton and Hardin, 2015).

The majority core courses for a data science major in computer science department are divided into three major domains: data science domain, computer science domain, and mathematics domain. The data science domain usually was covered by data organization and data management courses; the computer science domain was covered by programming, data mining, algorithms, and artificial intelligence courses; and the mathematics domain was covered by calculus, linear algebra, statistical method, and statistical learning (Anderson et al., 2014).

Horton and Hardin (2015) recommended that statistics in undergraduate curriculum teach students how to think with data, including statistic foundation, algorithmic thinking, and multivariate thinking. On the other hand, De Veaux et al., (2017) suggested six main subjects for a data science major: data description, math foundation, computational thinking, statistical thinking, data modeling, and communication. Compare to computer science and statistic majors, the data science program required comprehensive skills including math, computer, and communication.

To meet the market's demand, many universities in the U.S. provide doctoral, master, bachelor, and certificate programs in data science majors with data analysis, data mining, or big data concentrations and in both on-campus and online formats (Chen and Jiang, 2018; De Veaux et al., 2017; Donoho, 2017; Hassan and Liu, 2019). Many studies investigated the course contents and requirements of a data science program (Baumer, 2015; Song and Zhu, 2017), while others attempted to identify the sequencing of courses in undergraduate programs (De Veaux et al., 2017). In addition, some studies focused on the process in which data analytics was incorporated into curricula (Wymbs, 2016). Differences exist in undergraduate and graduate programs. Undergraduate programs tended to focus on understanding of tools (Mitri and Palocsay, 2015), while graduate data-science programs focused more on advanced and applied concepts, such as advanced statistics, big data analysis, and machine learning to help students understand techniques and to develop application to solve real business problems (Gupta, Goul and Dinter, 2015; Ortiz-Repiso, Greenberg and Calzada-Prado, 2018; Schaus, Van Hentenryck and Régin, 2009; Song and Zhu, 2015; Wixom et al., 2014). In summary, most studies in data science education focused on undergraduate and graduate program development (Cao, 2017; De Veaux et al., 2017; I. Y. Song and Zhu, 2015), and few studies had been devoted to certificate programs of data science (Gupta, Goul and Dinter, 2015; Hassan and Liu, 2019).

2.4. Requirements for Data Scientist Jobs in Logistics, Supply Chain, and Operations

Big data and data science have been increasingly used in logistics, supply chain, and operations management. According to the research conducted by Schoenherr and Speier-Pero (2015), there were many benefits of using data science in supply chain management (SCM), with informed decision making, increased visibility, better supply chain management as the top three. Big data was reportedly used in manufacturing to generate new product ideas (Mishra, Gunasekaran, Papadopoulos and Childe, 2018), and data analytics was critical in modern manufacturing (Bi and Cochran, 2016).

Data science in general focuses on broad areas with different types of analysis, such as descriptive analysis, predictive analysis, and prescriptive analytics (Akter and Wamba, 2016; Bertot and Choi, 2013; Dhar, 2012; Kundu and Garg, 2015). However, data science applied in logistic, supply chain, and operations has its unique focus (Chae, Olson and Sheu, 2013; Mishra, Gunasekaran, Papadopoulos and Childe, 2018).

The major focus of the data science in logistic include data collection, data quality, and the use of predictive model. Data in this area usually were collected from multiple sources, such as POS (point of sales), RFID (radio frequency identifications), and GPS (global positioning system) (Mishra, Gunasekaran, Papadopoulos and Childe, 2018; Queiroz and Telles, 2018); therefore, the ability to consolidate data from multiple sources in different format is required. Hazen, Boone, Ezell, and Jones-Farmer (2014) advocated the importance of addressing data quality issues in supply chain research because the quality of a business decision was based on the quality of the information generated by data analytics methods. The quality issues that they recommended to monitor and to control included accuracy, timeliness, consistency, and completeness related to inventory, customer address, and delivery dates. In addition, data scientists in logistic, supply chain, and operations needed to know how to use predictive model to plan future development or to optimize the progress (Govindan, Cheng, Mishra and Shukla, 2018; Waller and Fawcett, 2013).

2.5. Research Gap

Mishra, Gunasekaran, Papadopoulos, and Childe (2018) conducted research to evaluate the most influential articles on big data and supply chain published from 2006 to 2016. They used 'Big Data' and "Big Data and SCM' to search for scientific publications in peer-reviewed journals and identified 286 articles that contained the most influential works and researchers. Of the six clusters that they identified from these articles, the only one cluster concerning big data in SCM was on problems associated with data quality and its relevance in resolving problems of the supply chain. Therefore, research on data science in the field of logistics, supply chain, and operations management is needed.

Data science, by its definition, is a comprehensive subject that overlaps computer science, mathematical, specific business domain knowledge, and communication; and data science education from universities should provide learning opportunities to students with different backgrounds (Chen and Jiang, 2018). Certificate programs offer hands-on data analysis projects that provide students real practice experience to strengthen their knowledge and build strong relationships with different industries. They are usually short programs with five or six courses delivered online or on-campus (Chiang, Goes and Stohr, 2012). Thus, certificate programs, especially graduate-level certificates play an important role to train students with different backgrounds and with rich working experience to become data scientists (Chen, Chiang and Storey, 2012). However, few studies have been devoted to graduate certificate programs, as aforementioned.

To fill the gap of whether graduate data science certificates meet the needs of logistics, supply chain, and operations management professionals, this research investigated the current status and contents of graduate data science certificates; data science skills needed in logistics, supply chain, and operations management; and if the certificate contents meet the needs of the skill requirements.

III. METHOD

This section describes the process used for data collection and the data-process methods, which consisted of three main steps: data pre-processing, keywords extraction, and analyses; all were conducted using Python in Jupyter Notebook, an Integrated Development Environment.

3.1. Data Collection

To investigate graduate data science certificates that meet the needs for logistics, supply chain, and operations jobs, two sources of data were used: graduate certificates in data science and data scientist jobs in logistics, supply chain, and operations.

Google searches with keywords 'Graduate Certificates' combined with various data science keywords, such as 'Data Science,' 'Data Scientists,' and 'Big Data,' were conducted in late August and early September of 2019 to gather graduate certificates in the data science area. To ensure that the lists are current, "2019" was used as part of the keyword searches. The keyword 'Business' was not used as part of the searches to make certain that the searches would generate search results in all discipline areas. Several lists were obtained with a combined 249 unique graduate certificates and their URs, which formed the final certificate list. Data were then gathered manually from these URLs, including the universities' names, titles of the certificates, description of the certificates if available, number of required credits to complete the certificates, online or on-campus format, the school and department in which the certificates were offered if available, and required and elective course titles. Several certificates in this list did not have specific required courses; instead, students would select a fixed number of courses from a long list of courses from multiple discipline areas. For example, one certificate required students to select four courses from a list of 47 courses; another required students to select 12 credit hours from any graduate-level statistics courses. Providing a wide range of course selection did not appear to have the objective of enforcing the 'interdisciplinary' nature of the certificates because the selection of courses was not necessarily from different disciplines. Since one of the objectives of this study was to investigate the certificates that met the needs of specific jobs, these certificates were dropped from data collection due to the lack of specificity in terms of courses required.

If a URL led to an unfunctional web page, best efforts were given to find the certificate information from the university's home page and to collect data from the university's other web pages. However, this effort was occasionally fruitless. After the best attempts, data from 179 graduate certificates in data science from 132 universities in the U.S. were collected. Then, the certificates were categorized into 11 discipline areas, based on the titles of the certificates. If the discipline area is evident in the title of a certificate, the certificate is categorized into that discipline area. For example, if "Business" is included in the title of a certificate, the certificate is categorized as 'Business." If a certificate title is generic, the certificate was categorized using the name of the Department or School in which the certificate was offered. For example, a "Graduate Certificate in Data Analytics" certificate offered in a School of Engineering was categorized as 'Engineering.' Certificates that could not be categorized using the aforementioned criteria and with very few instances were categorized into one category: 'Other.' This included titles with words such as "Earth Data," "Game," "Geographic," "Government," "Higher Education," "Spatial," "Urban," and "Urban Science". In addition, two certificates with generic titles offered in iSchools were also included in the 'Other' category because of the very low instances. The data gathered from the 179 graduate certificates yielded 899 courses; 588 of which were required courses, and 311 were elective courses. Despite the interdisciplinary nature of big data, a great majority of graduate certificates were housed in a single department. Although a few certificates required students to take some courses from different departments, the majority of courses were from the same department.

To find appropriate job titles in logistics, supply chain, and operations areas that required data analytics skills, searches were conducted on the ONetOnline site, the Occupational Information Network developed under the sponsorship of the U.S. Department of Labor/Employment and Training. Keywords used to search for relevant job titles included 'operations,' 'logistics,' 'supply chain,' 'operation data scientists,' 'logistics data scientists,' 'supply chain data scientists,' 'operation analysts,' 'logistics analysts,' 'supply chain analysts,' 'operation specialists,' 'supply chain specialists,' 'logistic specialists,' 'data analysts,' 'business data analysts,' 'data scientists,' 'data engineers,' and 'research analysts.' These searches yielded three job titles in the logistics, supply chain, and operations management areas that required data analytics skills: Logistics Analysts, Operations Research Analysts, and Quality Control Analysts. These three job titles were, then, used as keywords to gather job links from Indeed.com job site. Table 1 shows the number of job links found and the unique URL count for each job category. Some job links in one job title were listed in other job titles; therefore, there were duplicate URL job links in the combined pool of links. After duplications were eliminated, there were a total of 2,845 unique job links and 2,318 unique job titles for the three job categories combined.

TABLE 1. LOGISTIC AND OPERATION JOB COUNTS IN LINKEDIN

Job Title	*URL Count*	*Unique URL Count*
Logistics analysts	7,940	1,076
Operations Research Analysts	17,490	925
Quality Control Analysts	13,750	910

3.2. Keywords Extraction

From the graduate certificate data source, keywords were extracted from the 899 course titles gathered from the 179 graduate certificates. For example, the course title "Data Mining for Business Intelligence" was extracted into two keywords: "Data Mining" and "Business Intelligence." The course title "Introduction to Applied Statistics" was extracted to "Introduction" and "Applied Statistics." This step extracted 1,477 keywords from the course titles, with many duplicate keywords since a keyword might exist in multiple course titles. Duplications were then eliminated to form 337 unique keywords.

3.3. Creation of a Comprehensive Job Skills for Course Title Keyword and Job Skill Validation

A comprehensive list of data scientist skills in the job markets was needed to validate the topics extracted from course titles and skills extracted from Indeed.com job sites. Since no such list was available, several steps were taken to construct such a list. First, a complete LinkedIn skills list available on a public domain was downloaded from http://tech.bragboy.com/2016/11/crawl-all-linkedin-skills.html, which contained 51,515 of all skills, not only data-related skills, found in LinkedIn in 2016. To ensure that current skills were included in the list, an additional 177 data science concepts from Data Science Ontology at datascienceontology.org were obtained and presented to a data scientist with more than 5 years of experience in data science area and more than 31 years of experience in information systems in general for review. The data scientist added 21 skills to the list. Finally, this list was combined with the LinkedIn list of 51,515 skills. After duplicates were removed, the final list contained a total of 51,704 skills in all areas, not limited to data-related jobs. This list was then used in keyword extraction; keyword not found in this list were not considered a skill or data-related topics, including articles, prepositions, pronouns, and irrelevant nouns and verbs.

3.4. Certificate Keyword Validation

To validate the 337 unique keywords extracted from the 899 course titles were job-skill related topics, an algorithm was used to find if any of the 337 topics existed in the list of 51,704 comprehensive skills. All 337 data science topics were validated.

3.5. Job Skills Validation

From the logistics, supply chain, and operations job postings data source, skills were obtained by scraping the 2,845 job links. This extraction yielded a large number of skills for each of the job search keywords, with many duplications within each job category. Each of these skills was checked against the 51,704 skills in the complete list of all skills. After the elimination of duplicates, each category had more than 8,000 unique skills within the category, as shown in Table 2. After combining all unique skills in these three categories, duplication again was found. After eliminating the duplications, the total number of unique skills for these three job categories combined was 13,202. It's important to reiterate that these were all skills, not only data-related skills, in data scientist jobs in logistics, supply chain, and operations.

TABLE 2. ALL SKILL COUNTS OF DATA SCIENTIST JOBS IN LOGISTICS/OPERATIONS.

Job Search Keyword	*Skill Count*	*Unique Skill Count*
Logistics analysts	769,398	8,075
Operations Research Analysts	1,924,924	8,540
Quality Control Analysts	1,475,465	8,405
Total	4,169,787	13,202 unique skills

The total skill count of 4,169,7878 non-unique skills obtain from Indeed.com was then processed against the validated 337 topics obtained from graduate data science certificates, resulted in 52,579 non-unique skills. After removing duplications, 290 unique skills were found. Since the initial Indeed.com skills were checked against the validated 337 certificate topics before eliminating duplication, this meant that 290 of the skills found in Indeed.com were also found in the unique keywords extracted from the graduate certificates.

3.6. Merging of Logistic/Operations Data Scientists Skills with Discipline Areas and Topics

To analyze which discipline area offered the data science skills needed in logistics and operations jobs, a list with all the relationships between each job skill and the discipline area/course in which the same skill was taught was needed. This merging process matched each data-related skill in each logistics/operations job to every course in every discipline area that taught the same skill, as illustrated in Figure 1.

This merging process yielded 427,721 items. Table 3 shows these counts by job category and discipline area.

IV. FINDINGS AND DISCUSSION

This section reports the findings and discussion on graduate certificates in data science, data scientist skills in logistics and operations, and the discipline areas in which these skills were taught.

4.1. Graduate Certificates in Data Science

Of the 179 graduate certificates gathered in data science, the average number of credits required to complete the certificates was 14 credit hours, which was approximately equivalent to five courses. In terms of discipline areas in which the certificates were offered, the Business area had the highest number of graduate certificates in data science, despite the fact that the word 'Business' was not included in the Google searches for graduate certificates in data science. Health and Math/Statistics were the areas that offered the second and third highest numbers of graduate certificates, respectively. Together, these three discipline areas offered 68% of graduate certificates in data science. Table 4 shows the discipline areas and the certificate count in each area. The finding of 39% of graduate certificates were in the Business area concurred the findings in a 2016 study that found approximately 40% of data science courses focused on business and social science (Cao, 2017).

Table 5 shows data-science keyword counts and the percentage of keywords found in each discipline area. Since 'Business,' 'Health,' and 'Math/Statistics' offered the highest numbers of certificates, they also have the highest numbers of data-science keyword counts.

In terms of keywords counts in all discipline areas, the top five highest numbers of keywords found in the 899 course titles were 'Introduction,' 'Data Mining," and 'Analytics.' This implied that many courses were introductory courses. Future research is needed to gather certificate admission requirements to investigate if 'Introductory' courses were offered to help students with a non-data-science background to build the foundation in a new field.

The keywords "Data Mining," however, had the second-highest counts in all courses, which indicated that "Data Mining" could be the most commonly offered topic in graduate data science certificates. Some defined 'Data Mining' as "the science of extracting useful knowledge from huge data repositories" (Chakrabarti, et al., 2016, p. 1). This definition seemed to meet the need for data analytics in the Big Data era, so it was not surprising to find that the keyword 'Data Mining' had such a high count.

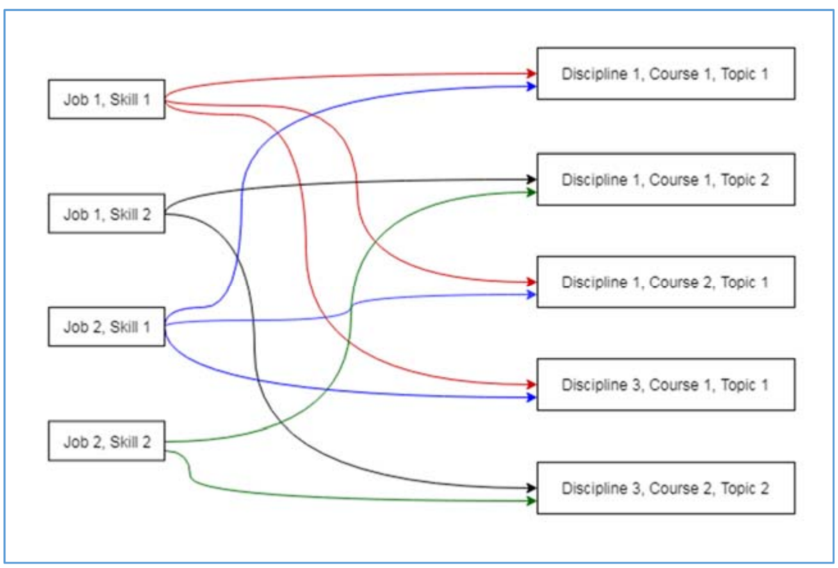

FIGURE 1. PROCESS OF MERGING JOB SKILLS WITH COURSES/DISCIPLINES/TOPICS.

TABLE 3. COUNTS OF LOGISTICS/OPERATIONS SKILLS TAUGHT IN VARIOUS DISCIPLINE AREAS

Job Search Keyword	*Discipline Area*	*Non-Unique Skill Count*	*Unique Skill Count*
Logistics Analysts	Business	67,346	141
	Computer Science	11,504	64
	Data Science	3,626	20
	Engineering	4,232	33
	Extended Education	2,785	23
	Health	27,405	87
	Information Sciences	3,118	16
	Math/Statistics	14,145	65
	Others	13,366	55
	Professional Development	1,464	19
Operations Research Analysts	Business	65,191	146
	Computer Science	12,101	67
	Data Science	3,945	21
	Engineering	4,703	35
	Extended Education	3,530	23
	Health	25,885	101
	Information Sciences	3,259	19
	Math/Statistics	17,960	69
	Others	16,827	57
	Professional Development	2,054	19
Quality Control Analysts	Business	50,191	133
	Computer Science	9,627	61
	Data Science	3,046	19
	Engineering	3,543	31
	Extended Education	2,412	23
	Health	25,457	100
	Information Sciences	2,493	16
	Math/Statistics	13,828	63
	Others	11,539	54
	Professional Development	1,139	19
Total		427,721	1,599

TABLE 4. COUNTS OF CERTIFICATE IN VARIOUS DISCIPLINE AREAS.

Discipline Area	*Certificate Count*	*Percent*
Business	69	39%
Health	29	16%
Math/Statistics	24	13%
Computer Science	17	9%
Engineering	8	4%
Information Sciences	7	4%
Data Science	6	3%
Professional Development	5	3%
Extended Education	4	2%
Other	10	6%
Total	**179**	**100%**

TABLE 5. COUNTS OF KEYWORDS EXTRACTED FROM VARIOUS DISCIPLINE AREAS.

Discipline Area	*Keyword Count*	*Percent*
Business	537	36%
Health	257	17%
Math/Statistics	186	13%
Computer Science	145	10%
Engineering	58	4%
Information Sciences	41	3%
Data Science	35	2%
Professional Development	39	3%
Extended Education	51	3%
Other	128	9%
Total	**1,477**	**100%**

Taking the design of the data mining curriculum as an example, the ACM SIGKDD Curriculum Committee intentionally divide the curriculum into the 'Foundations' part that contained basic data mining materials that should be taught in an introductory course. The second part of the curriculum contained 'Advanced Topics' that can be used in an advanced course (Chakrabarti, et al., 2016). To investigate if differences exist between 'required' versus 'elective' courses, keyword counts were analyzed accordingly. Figure 2 illustrates the keywords that had 10 or more counts in the 588 required courses, and Figure 3 shows the keywords that had five or more counts in the 311 elective courses. The keyword counts in the required courses were in line with the aforementioned top keyword counts: including 'Introduction,' 'Data Mining," and

'Analytics.' However, the keyword counts for elective courses deviated from this.

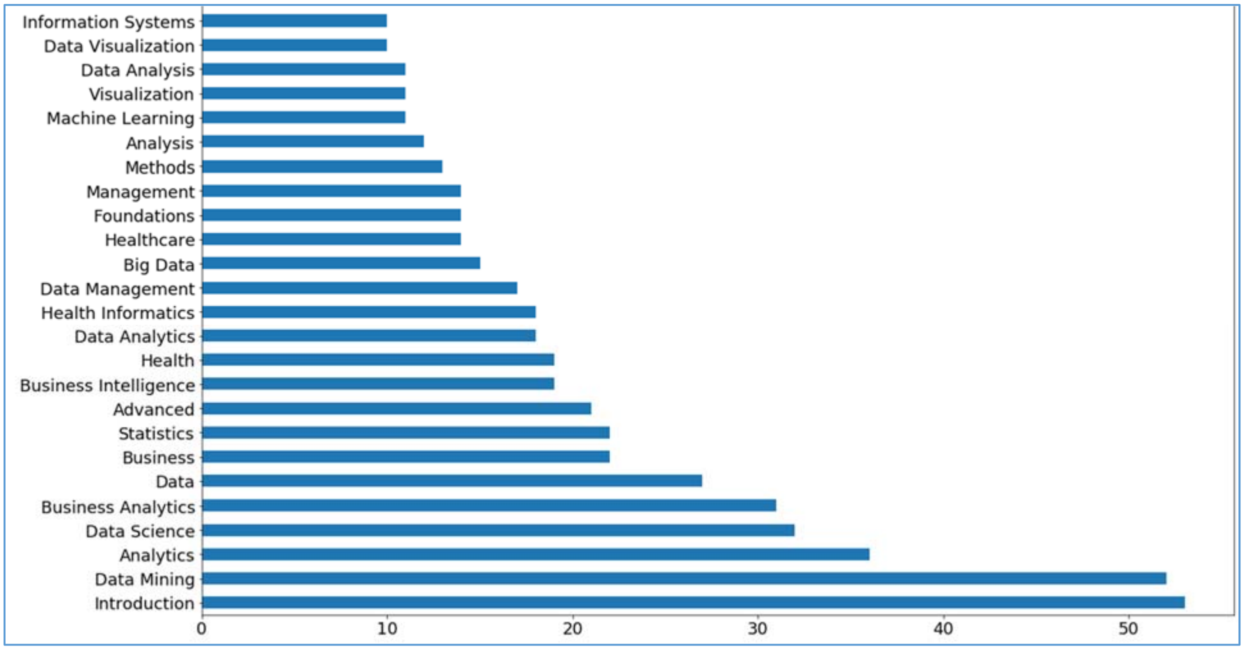

FIGURE 2. TOP REQUIRED COURSE KEYWORDS—WITH TEN COUNTS OR MORE.

As shown in Figure 3, the keyword 'Advanced' became more prominent in elective courses, which seemed logical. After students had taken required, 'Introductory' courses, they had the liberty to take more advanced courses to further their skills. Although 'introduction' remained as with a high count, this might be an introductory course of more advanced topics. In addition, more technical keywords, such as programming, Python, and SAS, appeared in elective courses, but not in required courses.

When considering the top 20 keywords in required courses and the frequency of the same keywords in elective courses, percentage, not counts, were used. Due to the large difference between the number of the required courses (588) and the number of elective courses (311), the percentage would depict a more acute picture. As shown in Figure 4, health-related keywords had a much lower percentage of keyword counts in the elective courses than in the required courses. This indicated that certificates with a health focus might have the same common required course as those in other discipline areas and that the focus of health-related topics was taught in elective courses. 'Data mining' had relatively high percentages of keywords in both required and elective courses. As aforementioned, keywords such as 'introduction' and 'foundation' had higher percentages in required than in elective courses, and 'advanced' had a higher percentage of keyword counts in elective courses.

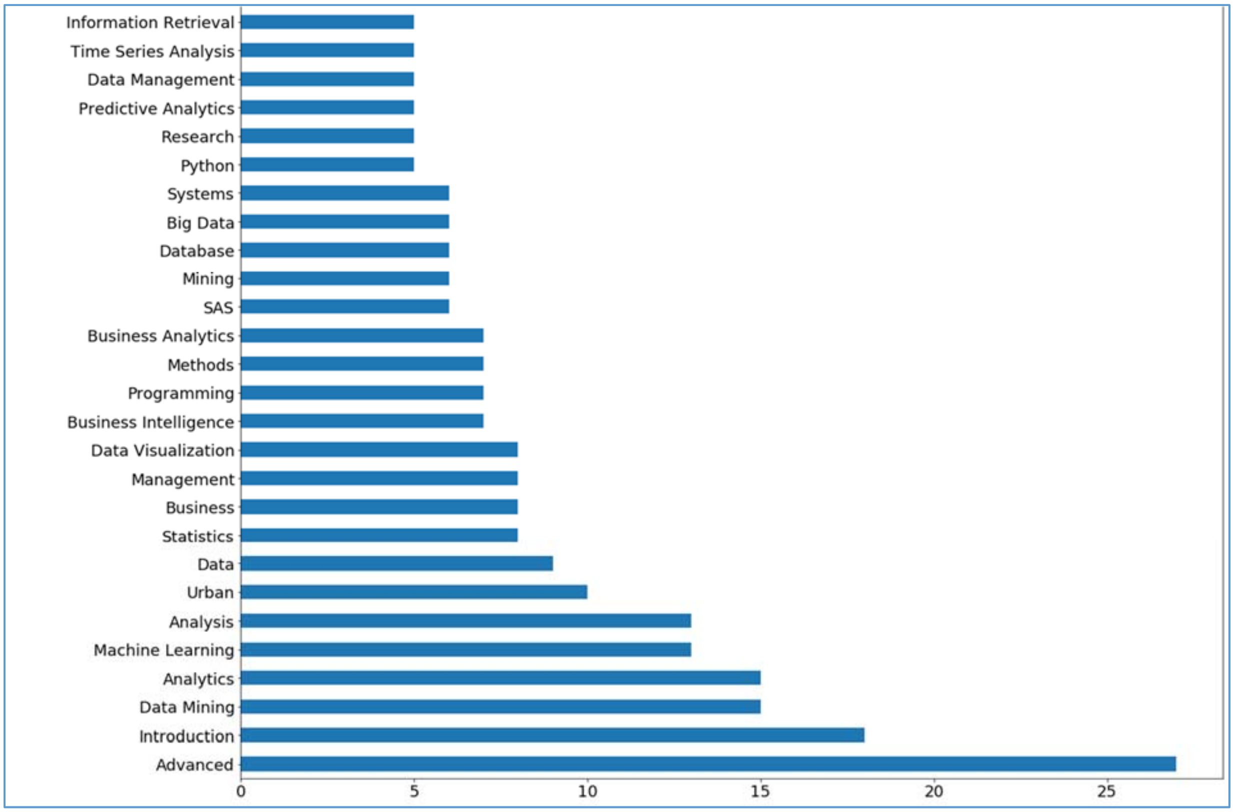

FIGURE 3. TOP ELECTIVE COURSE KEYWORDS—WITH FIVE COUNTS OR MORE.

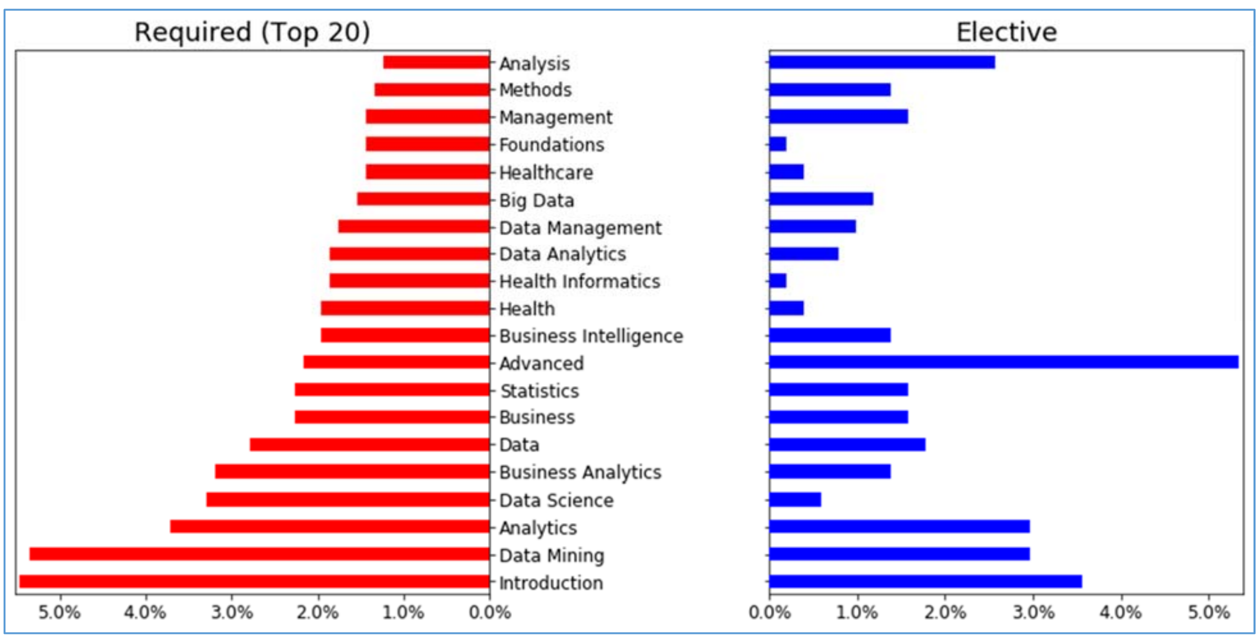

FIGURE 4. TOP 20 KEYWORDS IN REQUIRED COURSES IN COMPARISON WITH KEYWORDS IN ELECTIVE COURSES.

On the other hand, when considering the top 20 keywords in elective courses and the frequency of the same keywords in required courses, more technical keyword 'SAS,' 'programming,' and 'machine learning' had a much higher percentage of keyword counts in elective courses than in required courses, as shown in Figure 5. The keyword 'urban,' surprisingly, had a relatively high percentage, the 7th highest, of keyword counts in elective courses. Considering there were only two graduate certificates that contained the word 'urban' in their titles, further investigation would be needed for a better understanding of this finding.

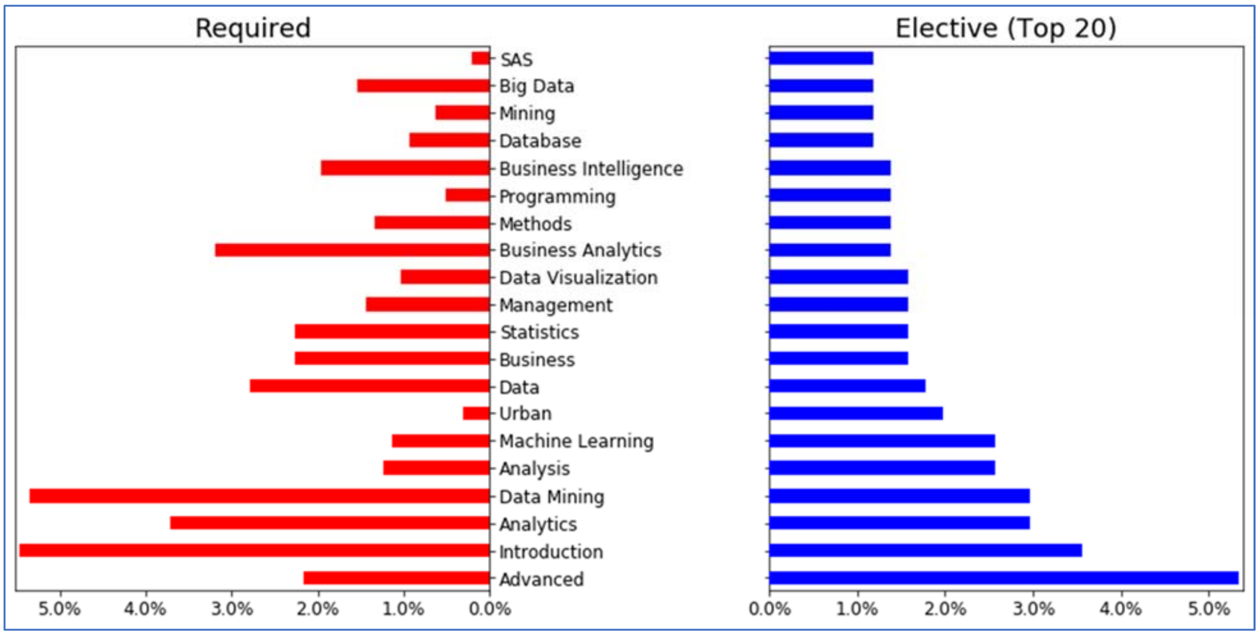

FIGURE 5. TOP 20 KEYWORDS ELECTIVE COURSES IN COMPARISON WITH KEYWORDS IN REQUIRED COURSES.

As aforementioned, the total unique keywords for all discipline areas was 337. Since many of these keywords appeared in multiple discipline areas, the total unique keywords across all 11 discipline areas were much higher: 1,477. To illustrate the keyword overlapped in multiple discipline areas, Figure 6 shows the number of same keywords found in different discipline areas. For example, the first column in the figure shows that there were 47 common keywords between 'Business' and 'Computer Science,' and there were 44 common keywords between 'Business' and 'Health.' Since the top four discipline areas that offered the highest numbers of certificates were Business, Health, Math/Statistics, and Computer Science, it was not surprising to find that 'Business' had the most common keywords with these three discipline areas. However, this finding also indicated that there was a fair amount of overlap in terms of keywords in various discipline areas. Of the 1,477 unique keywords found in the 899 graduate data science certificates, 54.84% of the keywords were found in more than one discipline area. While in the process of assigning certificates to different discipline areas, there was little indication that the courses were offered from different departments to form the 'interdisciplinary' nature of the programs. This finding, however, might provide the reason for the lack of involvement from multiple disciplines in one certificate: the 'interdisciplinary' nature of the certificates was implemented by covering topics from multiple areas in the courses offered in one department, instead of requiring students to take courses from multiple departments.

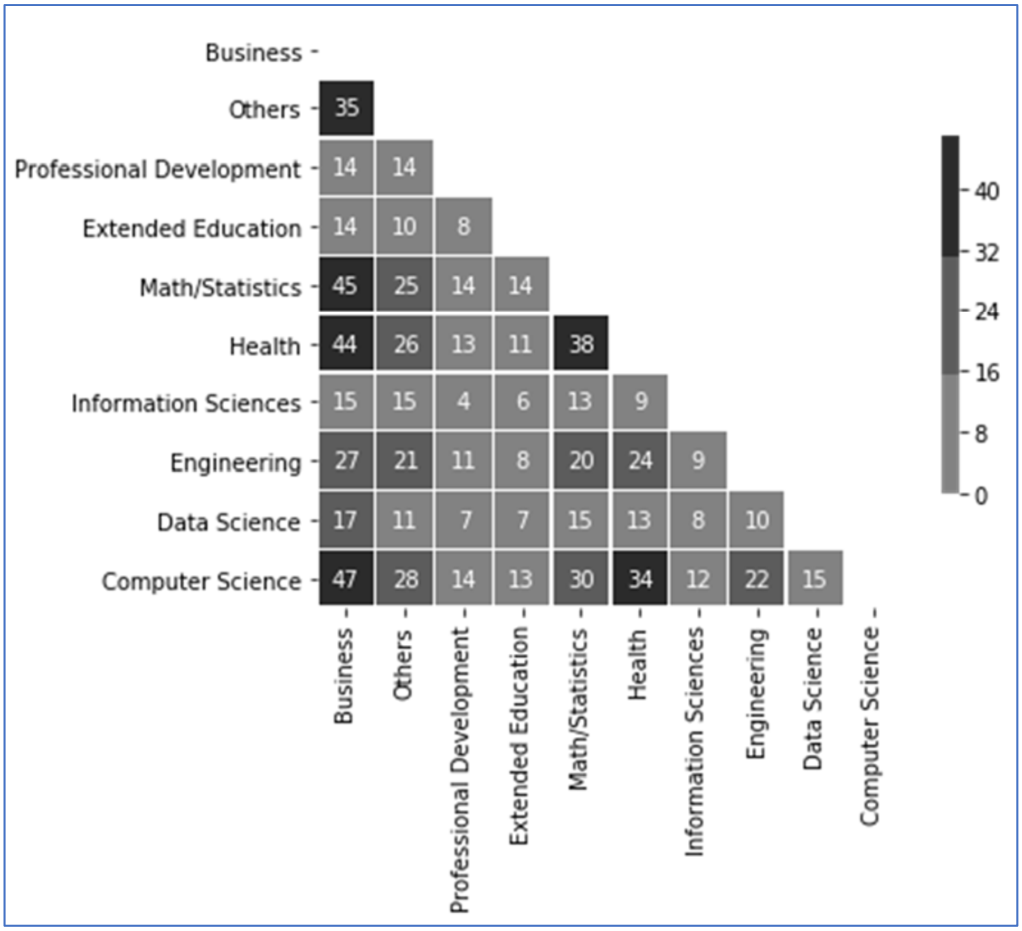

FIGURE 6. THE NUMBER OF SAME TOPIC KEYWORDS APPEARED IN MULTIPLE DISCIPLINE AREAS.

4.2. Logistics/Operations Data Scientists Skills

As mentioned in the Method section, 13,202 unique skills were obtained from 2,845 job links. A Venn diagram was created to show how these skills overlap among Logistics Analysts, Operations Research Analysts, and Quality Control Analysts jobs. As shown in Figure 7, 4,383 of the 13,202 unique skills were common for jobs in all three categories. This finding indicated that 33% of skills were common in all three job categories. The Venn diagram also indicated that the percentages of skills that were unique in each job categories were similar, with 1996 (15.12%), 1979 (14.99%), and 1792 (13.57%) and that skills that overlapped between two categories were also similar: 1152 (8.73%), 1009 (7.64%), and 891 (6.75%).

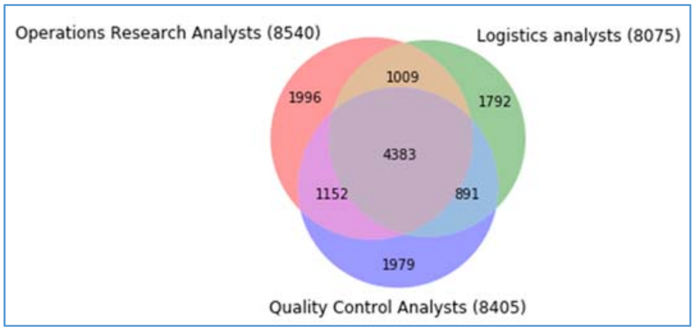

FIGURE 7. VENN DIAGRAM OF SKILLS IN LOGISTICS ANALYSTS, QUALITY CONTROL ANALYSTS, AND OPERATIONS RESEARCH ANALYSTS.

Skill keywords extracted from job links on Indeed.com were matched to a list of topics extracted from 899 courses, which resulted in 52,579 non-unique skills. Using this list, skills were grouped by job category. Figure 8 shows the counts of top 20 skills for Logistics Analysts in comparison with the counts of the same skill for Operations Research Analysts and Quality Control Analysts. 'Management' skill was most important for Logistics Analysts with 803 counts, in comparison with 585 for Operations Research Analysts and 619 for Quality Control Analysts. 'Business,' 'Planning,' and 'Communication' were more important for Logistics Analysts than for Operations Research Analysts and Quality Control Analysts jobs, as well. 'IT' skills was one in the top 20 most important skills for Logistics Analysts that had relatively equal importance in Operations Research Analysts and Quality Control Analysts jobs, with 388, 358, and 371, respectively. 'Technology' skill reinforced this assertion with relatively close counts in all three areas: 254 in Logistics Analysts, 275 in Operations Research Analysts, and 227 in Quality Control Analysts. Using 'Tools' was another skill that had relatively equal importance, with counts of 252, 272, and 265, respectively. These findings indicated that technology skills and the ability to use tools, presumably technology tools, were equally important for data-related jobs in all three categories.

Figure 9 shows the counts of top 20 skill keywords for Operations Research Analysts in comparison with the counts of the same skill keywords for Logistics Analysts and Quality Control Analysts. Obviously, 'Research' was the top skill needed for Operations Research Analysts with 769 counts. The counts for this skill were drastically lower for Logistics Analysts and Quality Control Analysts, 194 and 180 respectively. 'Strategic' was one of the top 20 skills for Operations Research Analysts but did not appear as one of the top 20 skills for Logistic Analysts, and it has a fairly low count for Quality Control Analysts, as well.

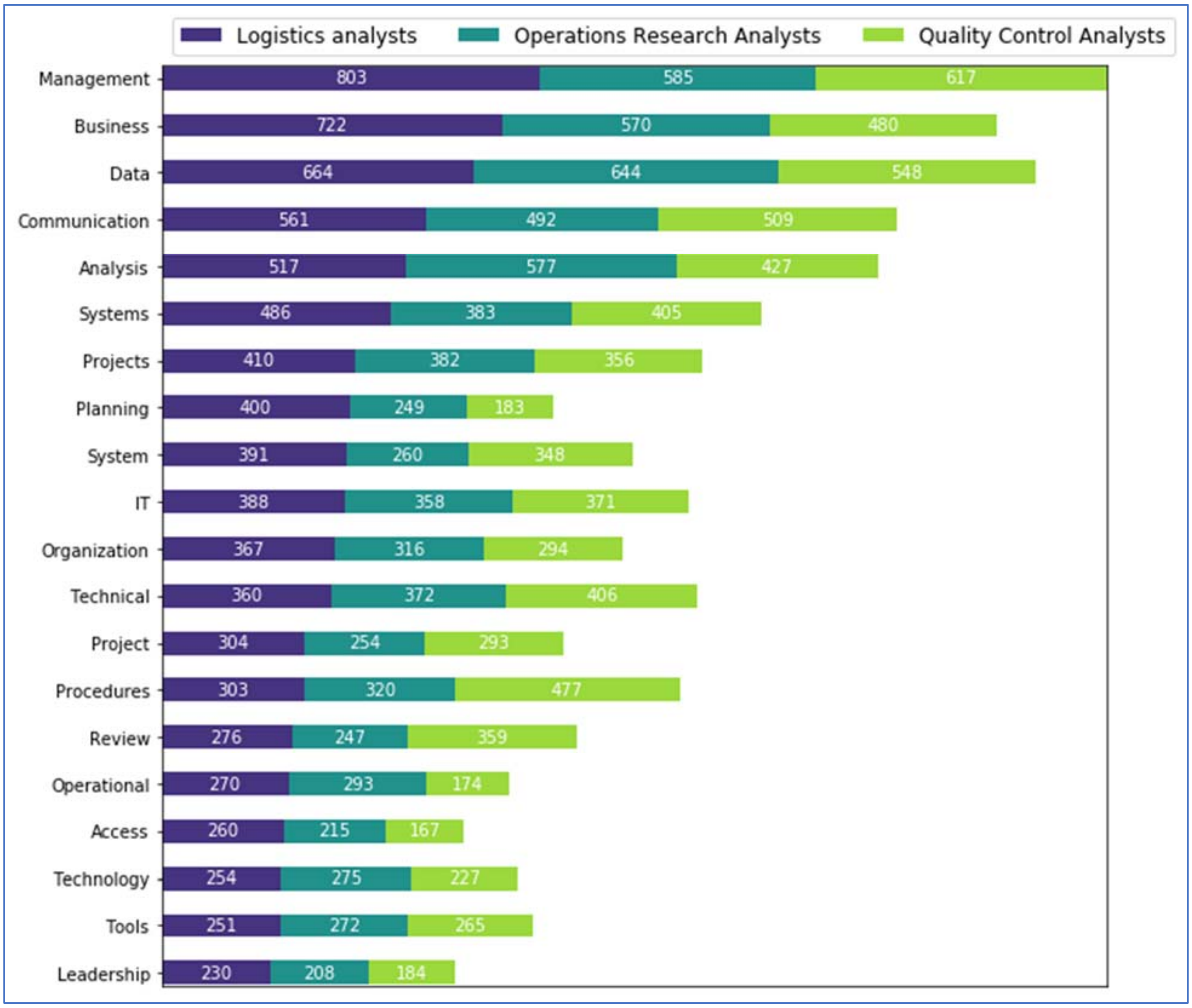

FIGURE 8. TOP 20 SKILL KEYWORDS FOR LOGISTICS ANALYSTS IN COMPARISON WITH THE COUNTS FOR OTHER JOBS.

FIGURE 9. TOP 20 SKILL KEYWORDS FOR OPERATIONS RESEARCH ANALYSTS IN COMPARISON WITH THE COUNTS FOR OTHER JOBS.

Figure 10 shows the counts of top 20 skill keywords for Quality Control Analysts in comparison with the counts of the same skill keywords for Logistics Analysts and Operations Research Analysts. Although the count of 'Management' was not as high as that of Logistics Analysts jobs, it had the highest count for Quality Control Analysts jobs. 'Testing' had a fairly high count of 403 for Quality Control Analysts but were drastically lower for Logistics Analysts and Operations Research Analysts. 'Quality Control' had the same phenomena. 'Health' and 'Application' appeared in the top 20 list for Quality Control Analysts but not in the top 20 for both Logistics Analysts and Operations Research Analysts jobs.

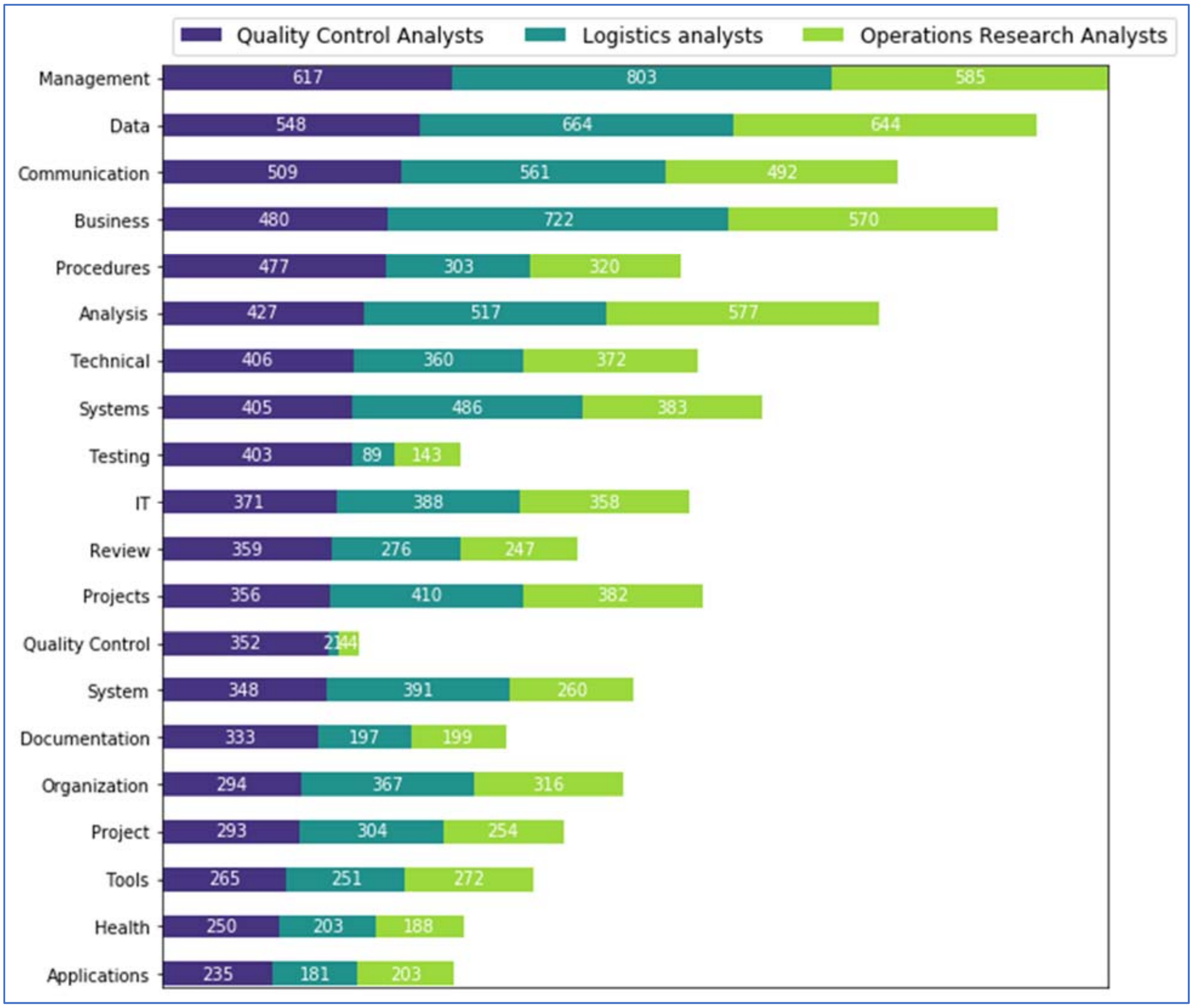

FIGURE 10. TOP 20 SKILL KEYWORDS FOR QUALITY CONTROL ANALYSTS IN COMPARISON WITH THE COUNTS FOR OTHER JOBS.

4.3. Logistics/Operations Data Scientists Skills Taught in Various Discipline Areas

Of the 290 unique Logistics/Operations Data Scientists skills that matched the topics taught in graduate data scientist certificates, 215 (74.13%) were common in all three job categories, as shown in Figure 11. Operation Research Analysts job had the highest number of skills (91.38%) covered in graduate certificates, with Quality Control Analysts being the second with 86.21% and Logistics Analysts the third with 85.17% of data-related skills covered in graduate data science certificates.

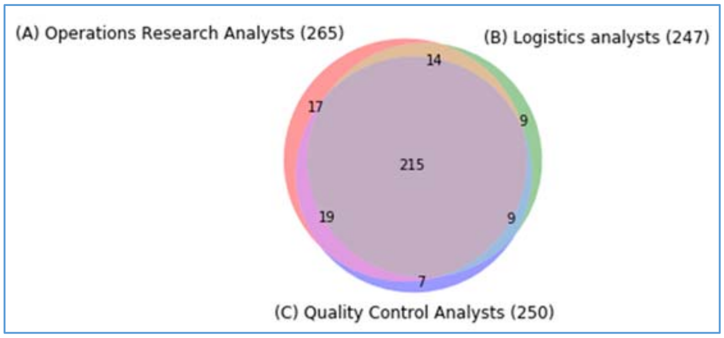

FIGURE 11. NUMBER OF SKILLS IN DIFFERENT JOB CATEGORIES COVERED IN CERTIFICATES.

Many of the 290 skills were covered in multiple discipline areas, Figure 12 shows the counts of unique skills taught in each discipline area for each of the job categories. Since the top three discipline areas that offered certificates were Business, Health, and Math/Statistics, these three areas had the highest counts.

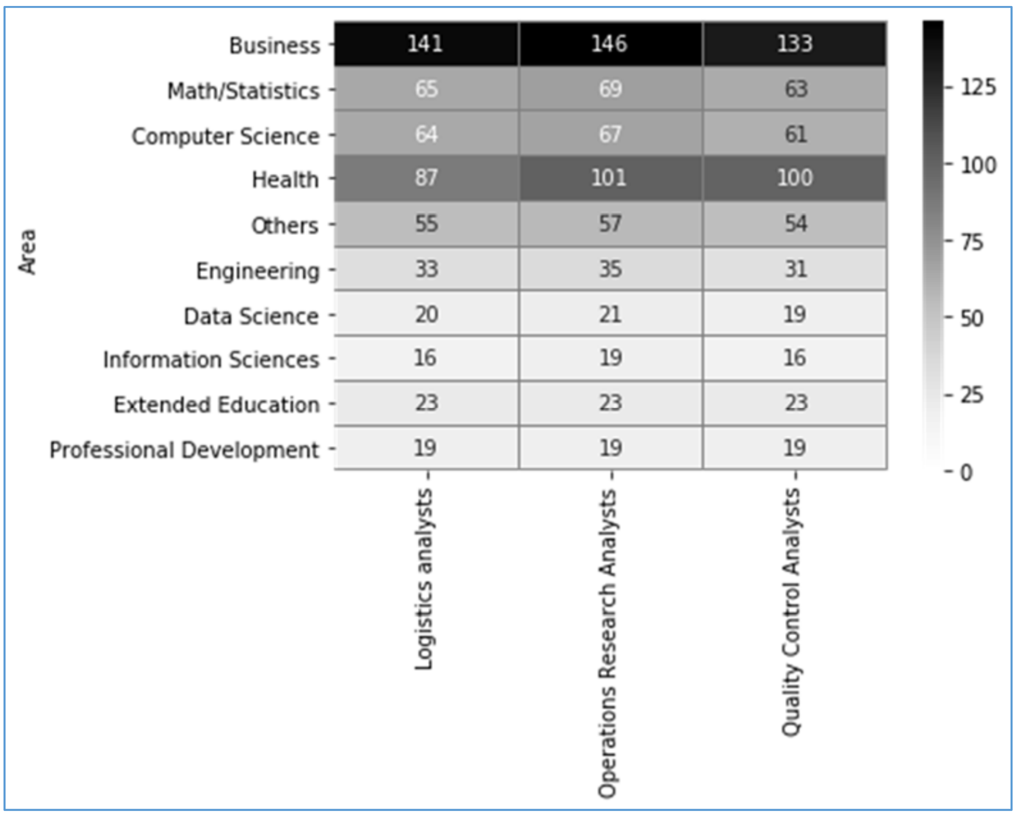

FIGURE 12. COUNT OF UNIQUE SKILLS IN EACH JOB CATEGORY TAUGHT IN VARIOUS DISCIPLINE AREA.

In terms of the percentage of data science skills for logistics/operation jobs that are taught in different discipline areas, 15.7% of Logistics Analysts data skills were taught in Business graduate certificates, and 6.4% can be found in Health graduate certificates, as shown in Figure 13. Business remained to be the top provider for Operations Research Analysts and Quality Control Analysts as well; one main reason for this phenomenon was the simple fact that Business offered the highest number of graduate certificates in data science. Although Health has the second-highest number of certificate offerings, Business probably was more related to logistics, supply chain, and operations management than Health.

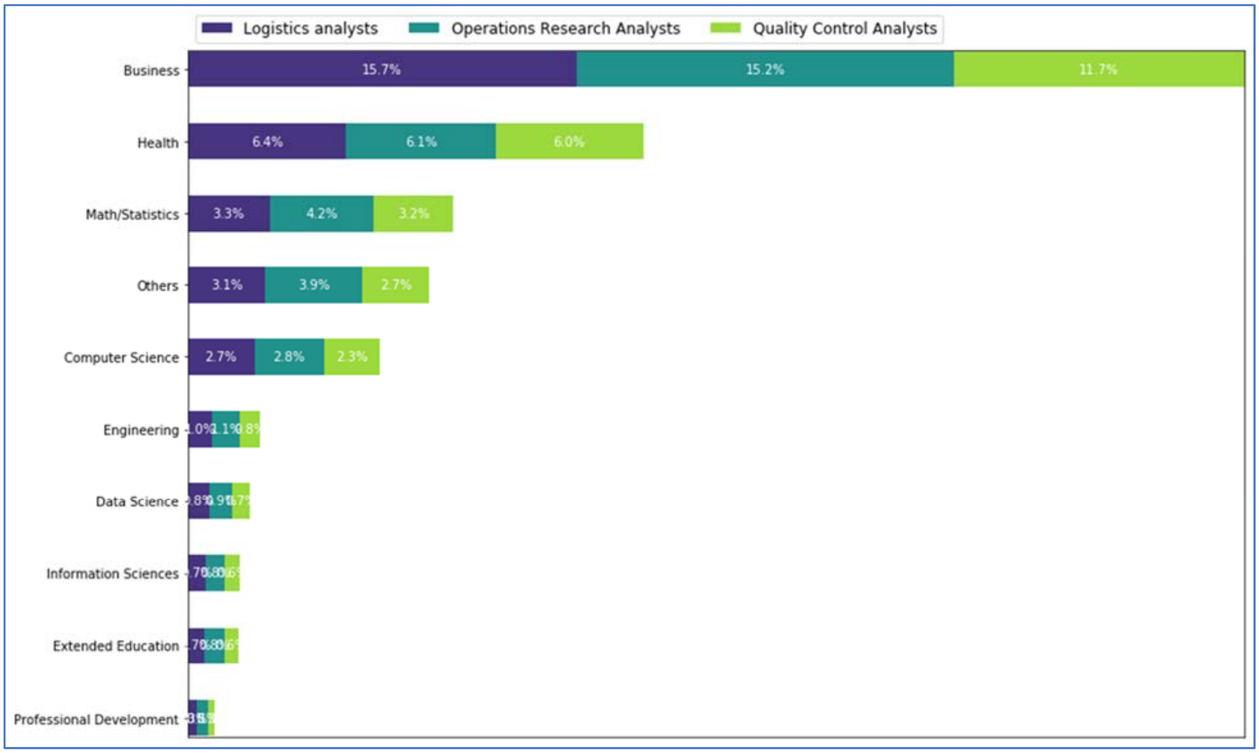

FIGURE 13. PERCENTAGE OF DATA SKILLS TAUGHT IN DISCIPLINE AREAS.

VI. CONCLUSION

Amplified by the big data phenomenon, jobs in the data-science area are expected to grow rapidly. Data-science degree programs are growing in tandem at both graduate and undergraduate levels to prepare the future workforce to meet the needs. However, a graduate certificate in data science provides an excellent alternative for working professionals to gain data-science skills without the commitment to a longer, full-time degree program. This study investigated 'miss' or 'match' between the data science skills needed in logistics, supply chain, and operations management and the courses offered in graduate data science certificates in the U.S.

The findings show that the majority of graduate data science certificates were offered in Business,' 'Health,' and 'Math/Statistics.' Many required courses were introductory or foundation courses, and many elective courses were advanced courses. Technical courses, such as programming, Python, and SAS, were more likely to be elective courses, not required courses. An interesting finding was that more than one-half of topics were covered in multiple discipline areas, but yet courses needed to complete a certificate tended to be offered in the same department. This might suggest that the 'interdisciplinary' nature of the certificate content was implemented not by requiring students to take courses from multiple departments but by infusing topics from multiple discipline areas into the courses in the same department.

Of the data-related skills in Logistics Analysts, Quality Control Analysts, and Operations Research Analysts, the graduate certificates as a whole covered 93.38% of the data-related skills needed for data scientists in Log Logistics Analysts, covered 86.21% of the skills needed for data scientists in Quality Control Analysts, and covered 85.17% of the skills needed for data scientists in Operations Research Analysts. These high percentages of coverage suggested that graduate certificate was a valid, desirable option for professionals in these fields to gain data-related skills. When professionals in these areas are looking for a graduate certificate, they would probably find it in Business programs because 39% of all graduate data science certificates were offered in this discipline area.

Certain data-related skills in Logistics Analysts, Quality Control Analysts, and Operations Research Analysts were specific to its job category and rarely covered in data science graduate certificates. For example, 'natural language processing,' 'survival analysis,' 'health informatics,' and 'computer security' were skills required only in Logistics analysts and occurred only once in the 899 certificate course titles. 'Categorical data analysis,' 'social network analysis,' and 'statistical computing' were skills required only in Operation Research jobs and occurred only once in the certificate course titles. For Quality Control Analysts jobs, 'scientific programming,' 'probability theory', 'public health informatics,' and 'statistical learning' were the specific skills required only in this job category and occurred only once in the certificate course titles.

REFERENCES

Akter, S. and Wamba, S. F., "Big Data Analytics in E-Commerce: A Systematic Review and Agenda for Future Research", *Electronic Markets*, 26(2), 2016, 173–194. https://doi.org/10.1007/s12525-016-0219-0

Anderson, P., Bowring, J., McCauley, R., Pothering, G. and Starr, C., "An undergraduate degree in data science: Curriculum and a decade of implementation experience", *Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, SIGCSE, Atlanta, Georgia, 2014, 145–150. https://doi.org/10.1145/2538862.2538936

Baumer, B., "A Data Science Course for Undergraduates: Thinking with Data", *American Statistician*, 69(4), 2015, 334–342. https://doi.org/10.1080/00031305.2015.1081105

Bertot, J. C. and Choi, H., "Big data and e-government: Issues, policies, and recommendations", *The Proceedings of the 14th Annual International Conference on Digital Government Research*, ACM, Quebec, Canada, 2013, 1-10. https://doi.org/10.1145/2479724.2479730

Bi, Z. and Cochran, D., "Big Data Analytics with Applications", *Journal of Management Analytics*, *1*(4), 2014, 249-265.

Cao, L., "Data Science: A Comprehensive Overview", *ACM Computing Surveys*, 50(3), 2017, 43:1–43:42, https://doi.org/10.1145/3076253.

Chae, B. K., Olson, D. and Sheu, C., "Understanding Big Data Analytics Capabilities in Supply Chain Management: Unravelling the Issues, Challenges and Implications", *International Journal of Production Research*, 52(16), 2013, 4695–4710, https://doi.org/10.1080/00207543.2013.861616.

Chakrabarti, S., et al., "Data mining curriculum: A proposal (Version 1.0)", *Intensive Working Group of ACM SIGKDD Curriculum Committee*, 140, 2006, 1-10.

Chen, C. and Jiang, H., "Important Skills for Data Scientists in China: Two Delphi Studies", *Journal of Computer Information Systems*, 2018, 1–10. https://doi.org/10.1080/08874417.2018.1472047

Chen, H., Chiang, R. H. L. and Storey, V. C., "Business Intelligence and Analytics: From Big Data to Big Impact", *MIS Quarterly*, 36(4), 2012,1165–1188.

Chiang, R. H. L., Goes, P. and Stohr, E. A., "Business Intelligence and Analytics Education, And Program Development: A Unique Opportunity for the Information Systems Discipline", *ACM Transactions on Management Information Systems*, 3(3), 2012, 12:1–12:13, https://doi.org/10.1145/2361256.2361257

Davenport, T. H. and Patil, D. J., "Data Scientist: The Sexiest Job of the 21st Century", *Harvard Business Review*, 90(5), 2012, 70-76.

De Mauro, A., Greco, M. and Grimaldi, M., "A formal definition of Big Data based on its essential features", *Library Review*, 65(3), 2016, 122-135.

De Veaux, et al., "Curriculum Guidelines for Undergraduate Programs in Data Science", *Annual Review of Statistics and Its Application*, 4(1), 2017, 15–30. https://doi.org/10.1146/annurev-statistics-060116-053930

Dhar, V., "Data Science and Prediction", *Communications of the ACM*, 56(12), 2012, 64-73.

Donoho, D., "50 Years of Data Science", *Journal of Computational and Graphical Statistics*, 26(4), 2017, 745–766. https://doi.org/10.1080/10618600.2017.1384734

Gardiner, A., Aasheim, C., Rutner, P. and Williams, S., "Skill Requirements in Big Data: A Content Analysis of Job Advertisements", *Journal of Computer Information Systems*, 58(4), 2018, 374-384, DOI: 10.1080/08874417.2017.1289354.

Govindan, K., Cheng, T. C. E., Mishra, N. and Shukla, N., "Big Data Analytics and Application for Logistics and Supply Chain Management", *Transportation Research Part E: Logistics and Transportation Review*, 114(May), 2018, 343–349. https://doi.org/10.1016/j.tre.2018.03.011

Gupta, B., Goul, M. and Dinter, B., "Business Intelligence and Big Data in Higher Education: Status of a Multi-Year Model Curriculum Development Effort for Business School Undergraduates, MS Graduates, and MBAs", *Communications of the AIS*, 36(January), 2015, 449–476. https://doi.org/10.17705/1cais.03623

Harris, H., Murphy S. and Vaisman, M., *Analyzing the Analyzers: An Introspective Survey of Data Scientists and Their Work*. O'Reilly Media, Inc., Sebastopol, CA, 2013.

Harris, J. and Mehrotra, V., "Getting Value from Your Data Scientists", *MIT Sloan Management Review*, 56(1), 2014, 15-18.

Hassan, I. B. and Liu, J., "Data Science Academic Programs in the US", *The Journal of Computing Sciences in Colleges*, 34(7), 2019, 56-63.

Hayashi, C., What is Data Science? Fundamental Concepts and a Heuristic Example. In: Hayashi C., Yajima K., Bock HH., Ohsumi N., Tanaka Y., Baba Y. (eds), *Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization*, Springer, Tokyo, 1998, 40–51.

Hazen, B. T., Boone, C. A., Ezell, J. D. and Jones-Farmer, L. A., "Data Quality For Data Science, Predictive Analytics, And Big Data In Supply Chain Management: An Introduction To The Problem And Suggestions For Research And Applications", *International Journal of Production Economics*, 154, 2014, 72–80. https://doi.org/10.1016/j.ijpe.2014.04.018

Horton, N. J. and Hardin, J. S., "Teaching the Next Generation of Statistics Students to Think with Data: Special Issue on Statistics and the Undergraduate Curriculum", *American Statistician*, 69(4), 2015, 259–265. https://doi.org/10.1080/00031305.2015.1094283

Kundu, S. and Garg, M., "Emerging Role of Data Scientist in Big Data", *International Journal of Computer Applications*, 123(3), 2015,15–18. https://doi.org/10.5120/ijca2015905258

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Hung Byers, A., "Big Data: The Next Frontier of Innovation, Competition, and Productivity", McKinsey Global Institute; 2011. http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation (assessed October 23, 2019)

McAfee A, Brynjolfsson, E. and Davenport, T., "Big Data: The Management Revolution", *Harvard Business Review*, 90(10), 2012, 60-68.

Mercer, D., "Global Connected and IoT Device Forecast Update", May 14, 2019. https://www.strategyanalytics.com/access-services/devices/connected-home/consumer-electronics/reports/report-detail/global-connected-and-iot-device-forecast-update?slid=967984&spg=9 (accessed November 23, 2019).

Mishra, D., Gunasekaran, A., Papadopoulos, T. and Childe, S. J., "Big Data and supply chain management: a review and bibliometric analysis", *Annals of Operations Research*, 270(1–2), 2018, 313–336. https://doi.org/10.1007/s10479-016-2236-y

Mitri, M. and Palocsay, S., "Toward A Model Undergraduate Curriculum for The Emerging Business Intelligence and Analytics Discipline", *Communications of the AIS*, 37(1), 2015, 651–669. https://doi.org/10.17705/1cais.03731

Nordrum, A., *"Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated", August 18, 2016. IEEE Spectrum.* https://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated (accessed October 23, 2019).

Ortiz-Repiso, V., Greenberg, J. and Calzada-Prado, J., "A Cross-Institutional Analysis of Data-Related Curricula in Information Science Programmes: A Focused Look at the iSchools", *Journal of Information Science*, 44(6), 2018, 768–784. https://doi.org/10.1177/0165551517748149

Pattin et al., "Training the next generation of quantitative biologists in the era of big data", In Pac Symp Biocomputing. 2014;20: p. 488-492. Jan 3, Big Island, Hawaii. https://pdfs.semanticscholar.org/5348/52

0cbc6131836d98cff6aa14363aab087ee2.pdf

Phillips-Wren, G., Iyer, L., Kulkarni, U. and Ariyachandra, T., "Business Analytics in the Context of Big Data: A Roadmap for Research", *Communication of the AIS*, 34(23), 2015, 448-472.

Pranavathiyani, G., "Google trends on data science - Towards data science", June 2, 2017, https://towardsdatascience.com/google-trends-on-data-sci`ence-160146fea72a (accessed November 28, 2019).

Provost, F. and Fawcett, T., "Data Science and Its Relationship to Big Data and Data-Driven Decision Making", *Big Data*, 1(1), 2013, 51–59. https://doi.org/10.1089/big.2013.1508

Queiroz, M. M. and Telles, R., "Big Data Analytics in Supply Chain and Logistics: An Empirical Approach", *International Journal of Logistics Management*, 29(2), 2018, 767–783. https://doi.org/10.1108/IJLM-05-2017-0116

Radovilsky, Z., Hegde, V., Acharya, A. and Uma, U., "Skills Requirements of Business Data Analytics and Data Science Jobs: A Comparative Analysis", *Journal of Supply Chain and Operations Management*, 16(1), 2018, 82-101.

Schaus, P., Van Hentenryck, P. and Régin, J.C. Scalable Load Balancing in Nurse to Patient Assignment Problems. In van Hoeve WJ., Hooker J.N. (eds), Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. CPAIOR 2009. Lecture Notes in Computer Science, vol 5547. Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-01929-6_19

Schoenherr, T. and Speier-Pero, C., "Data Science, Predictive Analytics, And Big Data in Supply Chain Management: Current State and Future Potential", *Journal of Business Logistics*, 36(1), 120-132.

Song, I. Y. and Zhu, Y., "Big Data and Data Science: What Should We Teach?", *Expert Systems*, 33(4), 2015, 364–373. https://doi.org/10.1111/exsy.12130

Song, I.-Y. and Zhu, Y., "Big Data and Data Science: Opportunities and Challenges of iSchools", *Journal of Data and Information Science*, 2(3), 2017, 1–18. https://doi.org/10.1515/jdis-2017-0011

Waller, M. A. and Fawcett, S. E., "Data Science, Predictive Analytics, And Big Data: A Revolution That Will Transform Supply Chain Design and Management", *Journal of Business Logistics*, 34(2), 2013, 77–84. https://doi.org/10.1111/jbl.12010

Watson, H., Wixon, B. and Ariyachandra, T., "Insights on Hiring for BI and Analytics", *Business Intelligence Journal*, 18(2), 2013, 4-7.

Watson, H., "Tutorial: Big Data Analytics: Concepts, Technologies, and Applications", *Communication of the AIS*, 34, 2014.

Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L. and Turetken, O., "The Current State of Business Intelligence in Academia: The Arrival of Big Data", *Communications of the AIS*, 34(1), 2014, 1–13. https://doi.org/10.17705/1cais.03401

Wymbs, C., "Managing the Innovation Process: Infusing Data Analytics into the Undergraduate Business Curriculum (Lessons Learned and Next Steps)", *Journal of Information Systems Education*, 27(1), 2016, 61–74.

Zikopoulos, P., Eaton, C., Deroos, D., Deutsch, T. and Lapis, G., *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill, New York, NY, 2012.