Zhuofu Wang. Data Science Programs in U.S. Library and Information Studies Schools. A Master's Paper for the M.S. in I.S. degree. April 2021. 55 pages. Advisor: Arcot Rajasekar

This study focuses on the top 15 ranked accredited library and information schools in the United States, as ranked by U.S. News & World report. Content analysis of the websites of the top 15 information schools was conducted to measure the achievements of data science programs. The study gathers and codes data, such as course breakdown, program description, employed professors, offered specializations and certificates, and curriculum structure of each data-science-related program in information schools. The results may show that while many universities are incorporating data science into their curriculum by employing professors and adding courses, there is still significant room for improvement, particularly in the teaching of analytical skills and the specializations offered. A ranking of each information school according to their data science programs was completed and evaluated at the end of the research. The findings of the study might be beneficial to current and incoming information school faculties and students.

Headings

Data Science

Curriculum

Library and Information School

Data Visualization

Database System

Data Analysis

Career Preparation

DATA SCIENCE PROGRAMS IN U.S. LIBRARY AND INFORMATION STUDIES SCHOOLS

by
Zhuofu Wang

A Master's paper proposal submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April 2021

Approved by

_____

Arcot Rajasekar

Table of Contents

Introduction

According to the McKinsey Global Institute, there is a growing demand for data-science-related jobs. “There will be an estimated shortage of 140,000 to 190,000 people with data analysis skills” (Manyika et al., 2011). A simple search for keywords such as data scientist, data engineer, and database engineer on Linkedin.com shows a list of 2000+ job openings related to data science. However, it is still publicly unknown how universities with library and information science programs are preparing their students for a career in data science. To find out the answer, content analysis of the program websites of 15 top information schools was conducted to gather quantitative and qualitative data that can be used to measure the achievements of the programs. This study focuses on the top 15 ranked accredited library and information schools in the United States, as ranked by U.S. News & World report. The study gathers and codes data, such as course breakdown, program description, employed professors, offered specializations and certificates, and curriculum structure of each data-science-related program in information schools. The results show that while many universities are incorporating data science into their curriculum by employing professors and adding courses, there is still significant room for improvement, particularly in the teaching of analytical skills and the specializations offered. A ranking of each information school according to their data science programs was evaluated and completed at the end of the research. The findings of this study might be beneficial to current and incoming information school faculties and students.

Literature Review

Defining Data Science

Nowadays, data science has appeared as a prevalent term and topic, and the concept of data science is being used in different industries. Because data science is a broad discipline that consists of subjects such as database, business, statistics, computer programming, etc., there is no official or specific definition of data science. However, definitions developed by different researchers share some common characteristics. Wil van der Aalst (2016) defines data science as the use of "[existing] approaches [that] need to be combined to turn abundantly available data into value for individuals, organizations, and society." Dhar (2013) has a similar idea; he concludes that "data science is the study of the generalizable extraction of knowledge from data." Several researchers agree that companies or organizations who become pioneers in the use of data science are more likely to gain a crucial advantage. Kampakis (2020) believes "[it] is clear that data science is expanding both vertically and horizontally across all industries and layers of society."

For the purpose of this study, data science is defined as a concept that uses statistics, algorithms, scientific methods, data analysis techniques, and systems to extract valuable insights from different kinds of data, including unstructured and structured data.

Data Science Program in the United States

Data science is a discipline that unifies different subjects, including statistics, computer programming, computer information systems, business, and data analysis. Further, additional data science programs are progressively developed to meet business needs. Power (2012) agrees that "[faculty] entrepreneurs continue to respond to new business needs and trends in decision support technologies with revised and innovative academic programs." A typical data science program requires students to have adequate mathematical, computer science, and statistics backgrounds. According to GilPress (2012), there are more than 100 graduate programs in data science and big data analytics under different disciplines. Therefore, it is difficult to determine which university has a data science program. Tang (2016) did an exploratory content analysis of 30 randomly selected data science programs. It is interesting to discover that there are 8 different disciplines that host a data science program, which reveals that there are significant gaps in the United States data science education.

According to Sae-Lim (2016), the 8 disciplines that host data science programs are the arts and sciences, business, computer science, engineering, independent data science centers, iSchools, math and statistics, and professional studies. The current situation of data science education in the United States is confusing for students who want to pursue careers in data science, especially for students entering iSchools. Little or nearly no information about data science programs in library and information science schools was provided. Therefore, there is a need for a study that identifies iSchools that are building a data science curriculum for students who want to become data scientists. In

addition, a potential audience will be beneficial to know how these data science programs can be built up to better satisfy graduates seeking for success in data science careers.

State of Art in Data Science Education

According to usnews.com (2021), the best undergraduate data science program is the data science major offered by UC Berkeley. This program was considered as the state of art now in data science education. The other top 3 data science programs in the list ranked by usnews.com are 1) Massachusetts Institute of Technology, 2) Carnegie Mellon University, and 3) Stanford University. An analysis of these programs would be conducted later in this study. The findings from the top 4 Ranked Undergraduate Data Science Programs would be compared with the findings from the iSchools. This comparison would reveal the key elements of a successful data science program.

According to Aarshay (2016), a successful data science program should have the following characteristic: 1) Program Name. 2)Curriculum. 3) Practical Training opportunities. 4) Industry Collaborations. 5) Research Opportunities. 6)Class Profile. 7)University Reputation (Rankings). 8)Return on investment. 9)First-hand experience. The state of art in data science education, which is the data science major offered by UC Berkeley, did an outstanding job in all these 9 aspects: 1) Program Name: The program name, data science major, reveals that UC Berkeley was serious about data science. They designed a program just for data science. 2) Curriculum: UC Berkeley had a balanced curriculum with a proper ratio of elective courses and core courses. 3) Practical training opportunities, given that data science is very application-oriented, UC Berkeley provided lots of practical training opportunities in the form of capstone projects, internships,

hackathons, competitions, etc. 4) industry Collaborations: UC Berkeley partnered with lots of famous companies and hosted lots of technical talks, career fairs, research collaboration, etc. 5) Research opportunities: UC Berkeley have the best data science labs and employed lots of faculty in data science, the school also sponsored large amounts of research projects. 6)Class Profile, UC Berkeley was known as one of the most selective universities in the united states. 7)University Reputation, UC Berkeley was ranked as #1 undergraduate data science in the United States according to Usnews (2021). 8) Return on investment. UC Berkeley offered lots of financial aids opportunities, and the tuition fee of the program is relatively low compared to similar programs. There was significant demand for data analysts in the United States, and graduates had a great chance to get a job that has a high salary. 9)First-hand experience, people who are already studying are mostly feeling positive about this program in social media such as Quora, LinkedIn and Facebook.

Skills Needed in the Job Market

According to Chen et al., 2012, The types of competencies needed in the job market are listed in three different analytical components, especially in the area of business analytics. Table 1 outlines the characteristics in each analytic skills component. Employers in the job market were looking for skills specified in analytical skills 1.0 and 2.0. According to Berkeley Boot camps (2021), The top 11 data analyst skills employers want to see in 2021 are: 1) Data Visualization. 2) Python. 3) SQL/NoSQL. 4)Social Media Mining. 5)Fundamental Statistics. 6)Natural Language Processing/Machine Learning. 7)Microsoft Excel. 8)High-Level Math. 9)Teamwork. 10)Communication. 11)

Business Savvy. The skills listed above could be separated into two sections, Technical skills, and non-technical or interpersonal skills.

While employers would like to hire more data scientists to incorporate big data into their growth strategy, it does not mean that they will lower the standard of the hiring process. In this data-science-centered environment, data scientists or people who have data science skills are much more helpful than before. They are extremely helpful to business success and business growth; However, candidates would need to meet the interpersonal qualifications and technical requirements that data scientists require.

Besides specific skills that data scientists should generally have, data released by IT career finder also reveals that data science-related jobs require an advanced degree, for example, a master's degree or PH.D. They prefer candidates who have a bachelor's degree in computer science, biology, economics, engineering, Statistics, Math, and so on depends on the positions they were applying for. For example, if a candidate was applying for a data analyst job in a biology firm, the company would prefer a candidate who has a bachelor's degree in biology with a master's or PH.D. Degree in data science.

Key characteristics of analytical skills of business intelligence and analytics (Source: Chen et al., 2012)

Analytical Skills 1.0	Analytical Skills 2.0	Analytical Skills 3.0
DBMS-based, structured content	Web-based, unstructured content	Mobile and sensor-based content
– RDBMS & data warehousing	– Information retrieval and extraction	– Location-aware analysis
– ETL & OLAP	– Opinion mining	– Person-centered analysis
– Reporting, dashboards & scorecards	– Question answering	– Context-relevant analysis
– Data mining & statistical analysis	– Web analytics and web intelligence	– Mobile visualization & HCI
– Ad hoc query & search-based BI	– Social media analytics	– Mobile business intelligence
– Interactive visualization	– Social network analytics	
– Predictive modeling	– Spatial-temporal analysis	
– Column-based DBMS	– Information semantic services	
– In-memory DBMS	– Natural language question answering	
– Real-time decision		
– Data mining workbenches	– Content & text analytics	

Challenges of Data Science

According to Kampakis (2020), "[data] is being created and consumed at an unprecedented scale. Data science is progressing even faster, further speeding up the rate of data creation and consumption." As data science is moving extremely fast, some key challenges have emerged. One of the challenges, defined by Agarwal (2014), is finding better algorithms and systems. Bohlouli (2020) suggests that there is a challenge in the transition of data science from research to application. Further, Dzemyda et al. (2020) define the most critical challenges in data science as:

1) Seamless integration of technologies supporting data sciences into complex cyber-physical-social systems.
2) Further development of data-driven intelligence methods and methodologies.
3) The meaning of physical and digital worlds, within the container of

data science, artificial intelligence, and machine learning.

Why Data Science

According to a 2017 report from IBM, the Business-Higher Education Forum, and Burning Glass Technologies, a job market analytics firm, Jobs related to data science and analytics were projected to grow by 28% in the next 5 years in the United States, with projected average salaries of $80200. Another report wrote by the U.S. Bureau of Labor Statistics shows that the field of data science is expected to grow much faster than average by 16% in the next 10 years. According to Ana Echeverri, AI specialty lead at IBM, "There are literally hundreds of thousands of open jobs in this space that go

unfilled.” All of these show that the field of data science is very potential to become the best major in the future.

Data science is vital to other science such as environmental science, biology, medicine, and other disciplines such as finance, business, and social science. Especially for a business, according to the report published by Accenture (2020), “79 percent of enterprise executives agree that any organization that does not incorporate big data into their growth strategy would lose their competitive edge and potentially go out of business.” Given this, it’s not surprising that why everyone was starting to learn about data science and why data science is getting popular nowadays. The projection released by Wikibon (2018) also showed that the revenues of the global big data market increase from $42 billion to $103 billion in the next 10 years, also proved the importance of data science. Data science can not only optimize the manufacturing process but also help businesses drive better business decisions.

Besides helping the field of business, data was also important to other science, for example, environmental science. There are schools that incorporate data science into natural science programs such as UC Santa Barbara. Bren school of environmental Science & Management in UCSB has newly created a Master of Environmental Data Science degree. According to the Bren school of environmental science, students would be required to complete a capstone group project that provides real-world experience in applying data science to environmental issues. Student would need to use data science applications such as SQL, R, and python to gain valuable insights from massive amounts of data. Students would also learn how to apply data visualization skills to share beautiful narratives. Students and professionals need these data science skills to solve complicated

environmental problems that cannot be solved manually, for example, environmental modeling and spatial analysis. Besides environmental science, more and more similar academic programs were developed, such as biological data science hosted by UCSB and health care data science hosted by USC. These are perfect examples of how data science is important to other science and disciplines.

Data Science Courses Being Offered by Other Disciplines

Data science is a huge topic right now, and it is being used by lots of disciplines, including medicine, genetics, biology, environment, law, business. The process of data science involves various disciplines and skillsets for effectiveness. Different disciplines were starting to learn how and when data is being collected and how to visualize the data for presentation. Schools have created lots of data science programs. The schools that are hosting data science programs include schools, Business, Mathletics and Statistics, Arts and Sciences, Computer science, Engineering, medicine, biology, as well as independent data science research center. The schools offered academic degrees ranging from certificate, Bachelor's, Master's, to Doctoral. Firstly, These courses are designed to teach students about data science applications such as Python and R, refresh student with a background in specific fields, and how to incorporate data science skills to solve or discover the problems that they can't solve manually, including machine learning, data visualization, data modeling, data mining, and other data science skills.

The followings are examples courses being offered by other disciplines:

Healthcare:

BME 501 Advanced Topics in Biomedical Systems Units: 4

BME 566a Topics in Health, Technology and Engineering Units: 2

BME 566b Topics in Health, Technology and Engineering Units: 2

DSCI 510 Principles of Programming for Data Science Units: 4

DSCI 549 Introduction to Computational Thinking and Data Science Units: 4

DSCI 550 Data Science at Scale Units: 4

(Master of Science in Healthcare Data Science, USC Viterbi,)

Business:

Python Workshop

Data Management

Competitive Analytics

Internet Customer Analytics

Optimization (Python)**

Prescriptive Models

Operations Analytics

Entertainment Analytics

Data Visualization

Fraud Analytics

Forecasting and Time Series

(Master of Science in Business Analytics, UCLA Anderson School of Management)

Biology:

ACO 501 Database Systems and Problem Solving in Python (3)

BIO 514 Statistical Models for Biology (4)

LSC 519 Applied Learning Lab (1)

LSC 547 Wet Laboratory Experience (1)

STP 560 Experimental Statistics in Biology (3)

(Master of Science in biological data science, Arizona State University)

Environmental Science:

EDS 211 Team Science, Collaborative Analysis and Project Management

EDS 212 Essential Math for Environmental Data Science

EDS 220 Remote Sensing and Environmental Data

EDS 222 Statistics for Environmental Data Science

EDS 223 Spatial Analysis for Environmental Data Science

EDS 230 Modeling Environmental Systems

EDS 231 Text and Sentiment Analysis for Environmental Problems

EDS 232 Machine Learning in Environmental Science

EDS 240 Data Visualization and Communication

(Master of Environmental Data Science, UC Santa Barbara, Bren School of Environmental Science & Management)

Research Questions

In the previous literature review, it was noted that there are data science programs across 8 disciplines in the United States, but the literature provides little information about data science programs in library and information science schools. This study identifies the information schools that are building a data science curriculum. This study also identifies how these data science programs can be built up to better satisfy graduates seeking success in data science careers. More specifically, the purpose of this study is to address the following research questions:

1. What are the curriculum requirements for completing various data science programs in information schools?
2. Are information schools preparing their students for data science careers?
3. What are the descriptive statistics and linguistic characteristics of data science program descriptions in different information schools?

Methodology

In recent years, the demand for employees with data science skills is proliferating. Data science skills include but are not limited to data analysis, data visualization, and database design. However, it is still publicly unknown how information science schools are preparing their students for data science careers. To investigate this, content analysis was conducted on the data science program websites of the top 15 ranked accredited library and information schools in the United States, as ranked by U.S. News & World report.

According to Hay (2005), content analysis is employed as "a way of identifying and counting phrases and terms that appear in a series of documents." According to Duriau et al. (2017), content analysis is a class of research methods at the intersection of qualitative and quantitative traditions. Content analysis can be applied to quantitative and qualitative data, which perfectly fits the scale of this study. The data gathered in this study's content analysis were coded for analysis. The coded data revealed where each information school stands regarding their data science teaching methods. Different search methods were employed during the content analysis, including navigating by likely headings provided by the search bars on the schools' websites and site maps.

Positionality / Researcher Role

According to a report released by EMC (2011), there are more than 40 higher-educational institutions in the United States offering data science programs. The schools

that offer these data science programs are library and information science schools (iSchools), and specific departments include engineering, the arts and sciences, computer science, and business school.

As a current master's student at the University of North Carolina at Chapel Hill, majoring in Master of Science in Information Science (MSIS), choosing suitable data science courses is critical to my career development in data science. Thus, it is interesting to learn how information schools are preparing their students for careers in data science and to examine the effectiveness of the data science programs offered at these information schools. I believe that there are critical gaps in the data science programs offered across different disciplines. For example, in the data science program in business school, most of the courses are usually devoted to teaching data visualization and communications. In contrast, in the data science program in computer school, most of the courses are generally dedicated to teaching data programming skills. Therefore, there are significant inconsistencies in the data science programs in different disciplines. I believe the data science program in information school emphasizes data science, as most of the courses are devoted to teaching advanced data analysis skills. Nonetheless, this viewpoint needs to be supported by data.

Sample / Research Participants

The study gathers quantitative and qualitative data from the program websites of the top 15 ranked accredited library and information schools in the United States, as ranked by U.S. News & World Report. Because there is a tie between the University of Texas, Austin, and the University of Michigan, Ann Arbor, there are 16 schools in total that represent the top 15 library and information studies programs. These 16 schools are:

- University of Illinois (UL), Urbana-Champaign
- University of Washington (UW), Seattle
- University of North Carolina (UNC), Chapel Hill
- Syracuse University
- University of Michigan (UM), Ann Arbor
- University of Texas (UT), Austin
- Rutgers, The State University of New Jersey–New Brunswick
- University of Maryland (UM), College Park
- Indiana University (IU), Bloomington
- University of Pittsburgh
- Drexel University
- Florida State University
- Simmons College
- University of California (UC), Los Angeles
- University of Wisconsin (UW), Madison
- University of Wisconsin (UW), Milwaukee

Data Collection Methods

To find out the key elements of a successful data science program, The study conducted a structured content analysis in four different sections: 1) Program overview, 2) Program description, 3) Program curriculum structure and 4) Course focus.

1) Program overview

The final result of the program overview analysis is represented in a table. This table includes summarized information about a school's data science course program,

certificates, specializations, and professors. The purpose of this data collection is to summarize data science programs in each school.

2) Program description

The program description analysis was conducted on the website of four programs, as ranked by U.S. News & World Report. Data about program description, data science definitions, and program outcomes were collected and coded. Notably, this study counted the frequency of specifics words used in the program descriptions. The study also excluded many specific words (company names and university names) and stop words from the word count to avoid any possible bias. The average length of program descriptions was also calculated.

3) Program curriculum structure

The curriculum structure analysis includes the total number of credit hours, the number of electives courses and core courses, capstone project and practicum requirements, master paper requirements, and comprehensive examination requirements. If any of the universities failed to supply the total number of offered courses on their program websites, the study counted these data as missing values.

4) Course focus

The course focus analysis includes the number of professors focusing on data science from each program and the number of professors not focusing on data science. The analysis was based on whether courses in the data science programs cover 1) information, 2) mathematics and statistics, 3) visualization, 4) communication skills, 5) Programming, 6)Data structure, 7)Algorithms, 8) AI, 9) data science and other relevant words of data science. This analysis also looked for courses that cover analytical skills

1.0, 2.0, and 3.0 from Chen et al. (2012). The courses were coded based on the types of courses. In examining course descriptions and course titles, courses were coded as offering information skills with respect to information systems, information retrieval, web searching, and information organization. Courses that covered visualization tools, foundations, principles, and methods were counted as visualization skills. Communication courses covered presentation, writing, and public speaking. Mathematics and statistical methods courses were considered as related to math and statistics.

Courses that covered both data science domain knowledge and analytics skills were considered as part of the data science category. The following labels were created to identify data science courses: data science, big data, database management, data analytics, health informatics, data mining, machine learning, information analytics, data ethics, and data warehousing.

Data Analysis Methods

Quantitatively, this study documented and recorded the number of all instances of each category. This documentation includes the number of certificates, data science professors, specializations, and core and elective courses. These quantitative data were used for comparing and ranking each school based on the numbers. Qualitatively, the method of contextual coding was used to sort the gathered data. The sorting was fulfilled by reading through the specialization, course, course description, professor, and certificate that meets the requirements for data science classifications. Prospective coding labels include data science, big data, database management, data analytics, health informatics, and data mining. These broad coding labels were identified as the main focus

of different courses, professors, specializations, and certificates. Also, the labels indicated the topic covered in each section.

The data is presented mainly by graphs and charts that perfectly present the main coded results of both the qualitative and quantitative data. Different types of diagrams are used to describe the distribution of the data. Possible types of graphs include line graphs, histograms, bar graphs, and scatter plots. A ranking of data science programs in information schools was generated based on the content analysis results. To minimize human bias, the criteria were identified after the data collection process was completed. A sample criterion is if more than half of a school's professors have a data science background, then the rating on the "number of professors" section will be "Excellent." Each section is rated on a scale of poor, average, and excellent. No software was used to assist with the data analysis.

Top 4 Ranked Undergraduate Data Science Programs

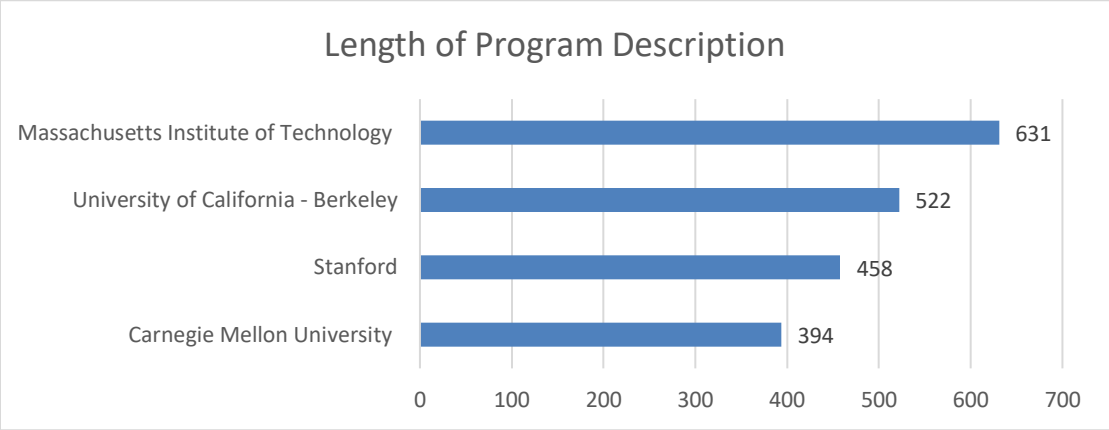
This portion of the study focuses on the top 4 ranked undergraduate data science programs in the United States, as ranked by U.S. News & World Report. They are 1) University of California, Berkeley, 2) Massachusetts Institute of Technology, 3) Carnegie Mellon University, and 4) Stanford University. These 4 programs are considered the golden standards of data science programs. The findings from the top 4 Ranked Undergraduate Data Science Programs would be compared with the findings from the iSchools. This comparison would reveal the key elements of a successful data science program.

Of the 4 data science programs information gathered, all 4 undergraduate data science programs addressed data science to some extent, whether by employing professors with a focus on data science, offering data science courses specialization and certificates, adding internship requirements to their curriculum. This table reveals that the state of arts in data science education should have all these components. To answer the research question 2, schools could show there are actually preparing their students for data science careers by showing they have all the essential elements.

Information used to create tables and graphs in this study were collected on samples' websites that hosted the programs. The majority of the data collection occurred in the official website of the programs and the school catalog. The data of description analysis was obtained on the website of sample programs, the data of course focus

analysis was obtained on the school catalog. Because not all data are available for each school from similar data, further work is needed to gather these data and provide correlation among them. See Appendix A for the entire list of programs in the sample.

Table 1. Length of program description



The longest program description of the top 4 undergraduate data science programs was MIT (L = 631), while CMU had the shortest description (L = 394). To answer the research question 3, one of the descriptive statistics of data science programs was that the length of the description was more than 300 words so that it could explain the program well.

Table 2. Number of unique words

Table 2 shows the number of unique words used in the program description. While UC- Berkeley had the highest number of unique words (N = 124), CMU had the lowest number of unique words; therefore, CMU had the lowest number of unique words as well as the length of the program description. To answer the research question 3, one of the descriptive statistics of data science programs was that the number of unique words in program description range from 40 to 100.

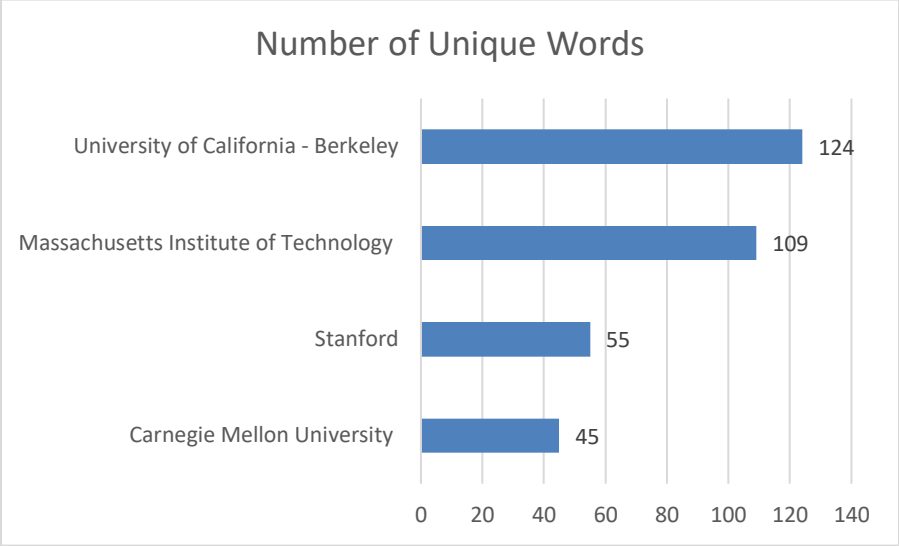
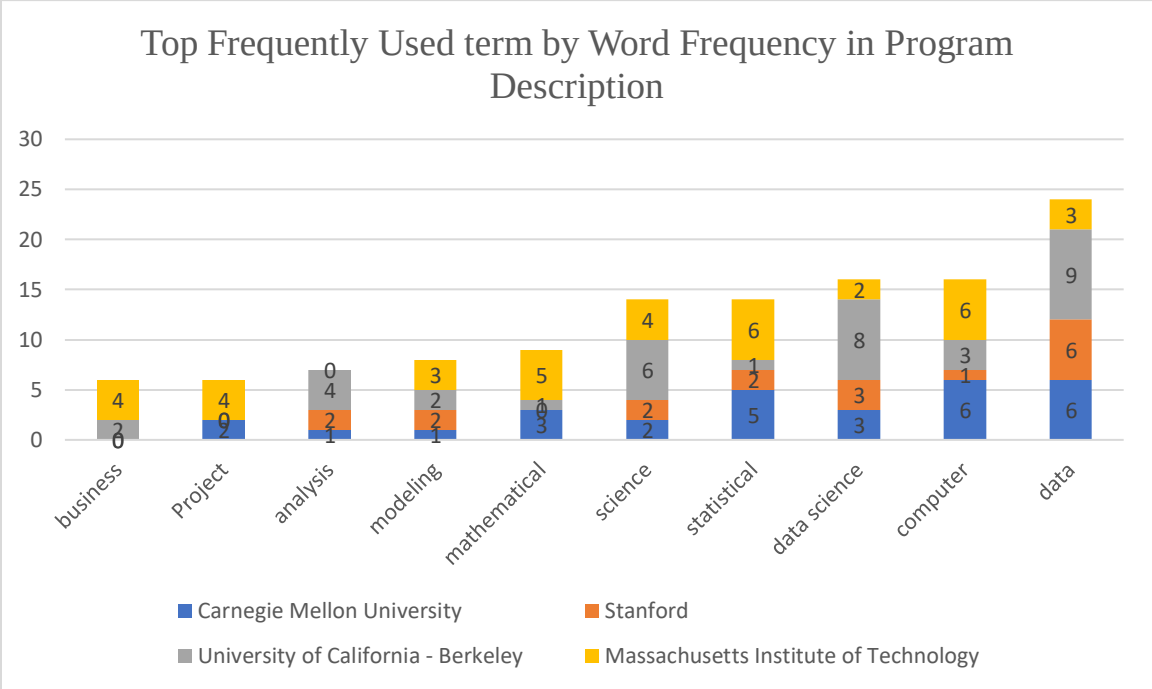


Table 3. Top frequently used term by word frequency in program description



Certain schools have used words unique to their program. For example, data science program at MIT likes to use keywords like “Computer” while data science program in UC – Berkeley tends to use keywords like “modeling” and Business.” Table 3

could confirm that the schools were actually preparing their students for data science careers and what the school cares about the most.

Table 4. Number of credit hours (Total credits vs. core vs. electives)

Table 4 summarizes the credit hours required by each university. The average number of credit hours over 4 data science programs was about 147, with about 102 in the core courses and 45 in elective courses. CMU had the highest number of core courses and total courses. The result of the table could not be used to compare with master programs in iSchools because of the lengths of different types of academic degrees. A Bachelor's degree normally takes students 4 years to complete, while a Master's degree normally takes students 2 years to complete. Bachelor's degree definitely has more credit hours and requirements than a Master's degree. Therefore, this study wouldn't compare the result of the 4 gold standards to the programs in information school directly. However, we can use the ratio/percentage in stand of numbers to discover some findings, which would be explained later in this study. To answer research question 1, the credit hours of completing various data science program ranges from 89 to 196.

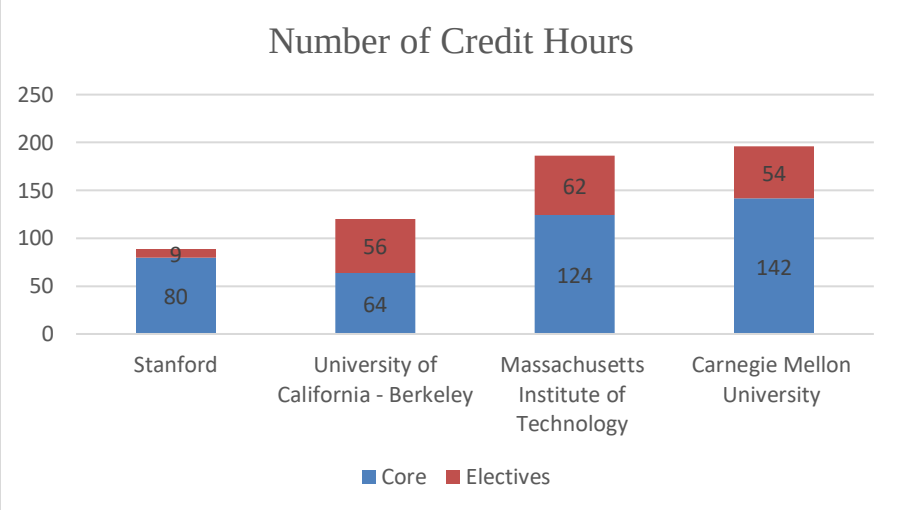
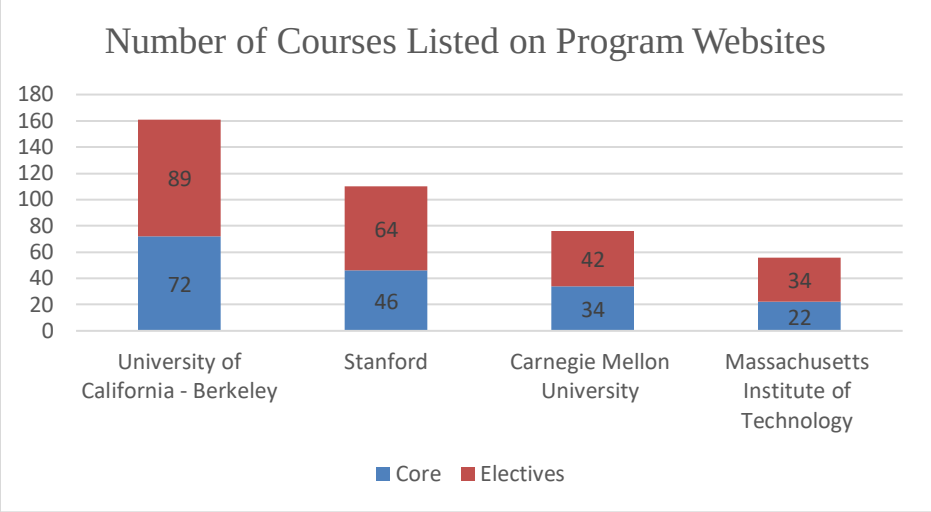


Table 5. Number of courses listed on program websites (core vs. electives)



Further analyses were conducted about the number of courses listed on program websites. UC – Berkeley had the highest number of Core courses (n = 72) and electives courses (n = 89). In contrast, MIT had the lowest number of core courses (n = 22) and electives courses (n = 34). Table 5 shows that gold standard schools have lots of data science courses for their student to choose, which confirmed these schools are serious about preparing their students for data science careers.

Table 6. Number of professors

Table 6 shows the number of professors whose research area focuses on data science or a part of data science for each program. CMU had the highest number of professors with a focus on data science (n =46), while Sandford had the lowest number of professors (n = 18). Table 6 shows that gold standard schools have lots of professors focusing on data science, which benefits students by providing quality courses and

research opportunities. Table 6 confirmed these schools are serious about preparing their students for data science careers.

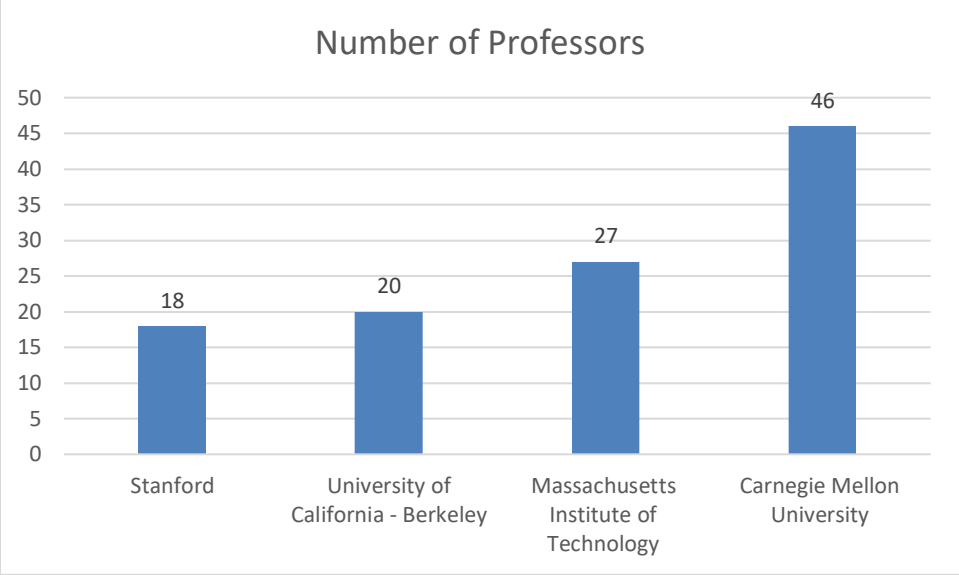


Table 7. Percentage of data science, mathematics & statistics, communication, visualization, and information skills in all courses

University	Data Science/Analytics Skills	Mathematics&Statistics	Communication	Visualizaition Skills	Information Skills
University of California, Berkeley	32.84%	23.33%	17.78%	8.25%	17.80%
Stanford	34.83%	38.94%	8.60%	7.46%	10.11%
Carnegie Mellon University	38.03%	40.82%	0%	0%	7.48%
Massachusetts Institute of Technology	35.36%	28.18%	13.33%	12.00%	4.00%

In terms of the focus of all courses, the highest percentage of courses covering data science were the programs from CMU (N = 38.03%). However, the percentages of data science in each school were really close. Table 7 refers that a great data science program should at least have 30% of courses covering data science/analytics skills in their class structure. CMU had the highest percentage of courses covering data science

and mathematics/statistics. In contrast, CMU provided none of the courses covering communication and visualization courses.

Table 8. Pie chart of the percentage of data science, mathematics & statistics, communication, visualization, and information skills in all courses

This is just another visual presentation of table 7 in pie charts. Table 8 clearly shows CMU had the highest percentage of courses covering data science and mathematics/statistics. In contrast, CMU provided none of courses covering communication and visualization courses. UC – Berkeley had the highest percentage of courses covering Information skills as well as communication skills.

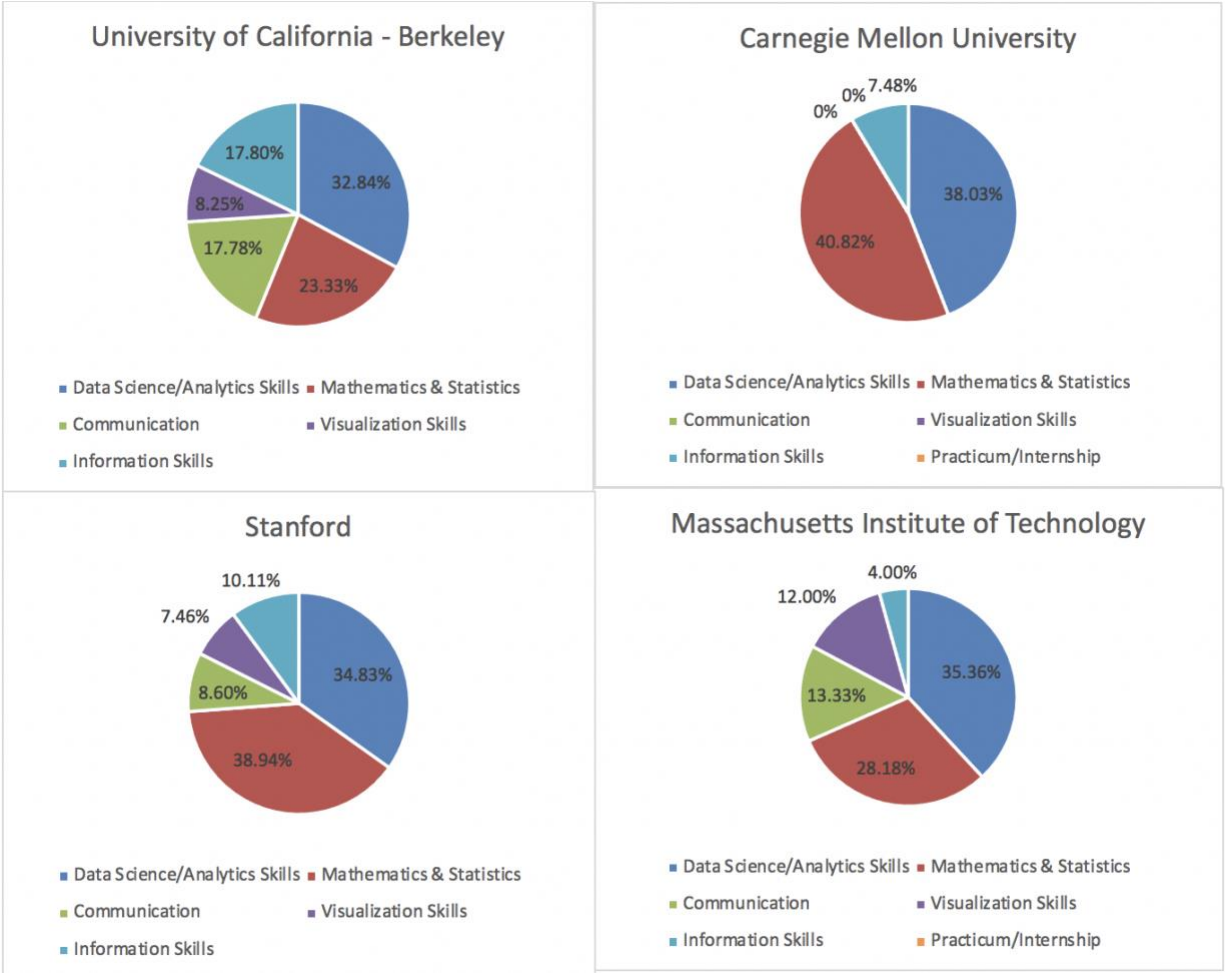
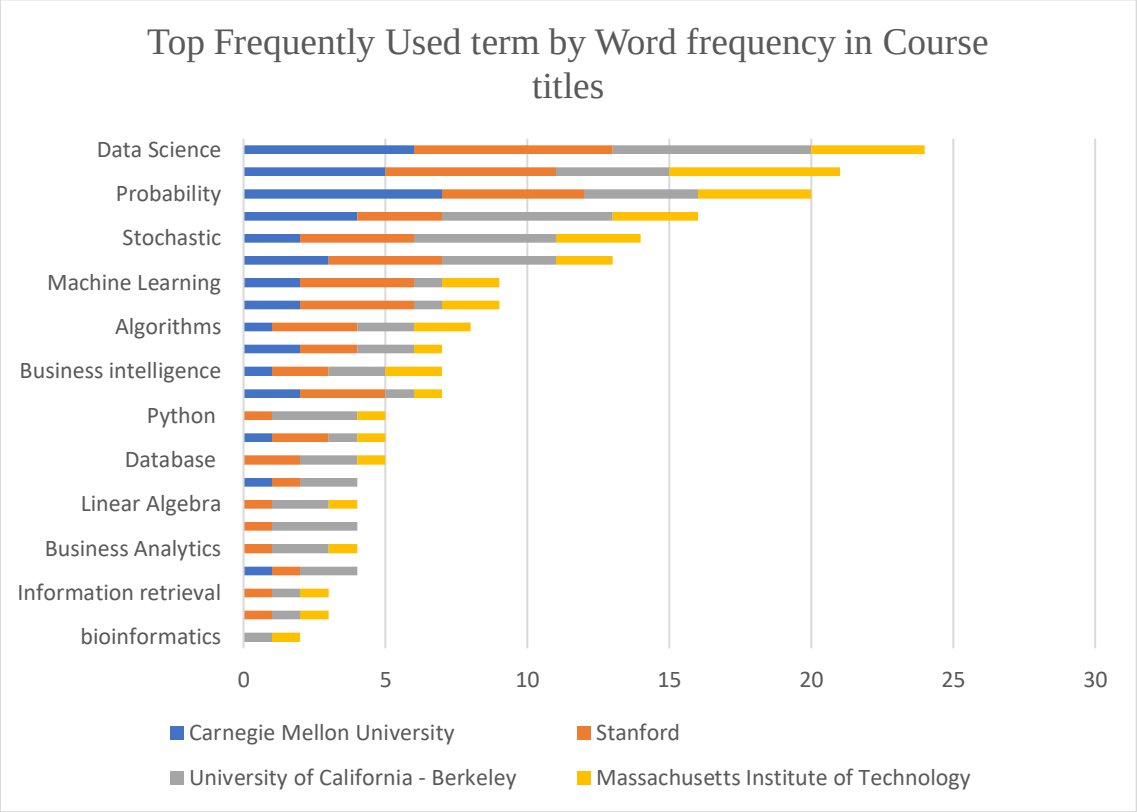


Table 9. Top frequently used term by word frequency in course titles

Table 9 reveals that are the top frequently used term by word frequency in course titles. This helps to form a set of words that the study can then see if the iSchools cover them. The top 5 words from this graph include “Data Science,” “Calculus,” “Programming,” “Stochastic,” “Probability.” Therefore, the study needed to be extra careful while noticing these words in the course titles.



Analysis of iSchools for Data Science Preparation

This portion of the study focuses on the top 15 ranked data science programs in the information schools, as ranked by U.S. News & World Report. Please see Appendix A for the entire list. The findings from the top 15 Ranked Data Science Programs in iSchools would be used to compare with the findings from the 4 gold standards. This comparison would help identifies how these data science programs can be built up to better satisfy graduates seeking for success in data science careers. More specifically, the purpose of this study is to address the following research questions:

1) What are the curriculum requirements for completing various data science programs in information schools?

2) Are information schools preparing their students for data science careers?

3) What are the literal and linguistic characteristics of data science program descriptions in different information schools?

1. Program overview

Table 10. Overview

Table 10 summarized which programs have specializations, professors, certificates, or courses focusing on data science. All information schools addressed data science to some degree by hiring professors with a background in data science or data management by implementing courses focusing on data science or offering big data or

applied data science specializations. However, Table 10 also showed that some schools did not put as much effort as others into preparing their students for data science careers. Some of them lack data science specializations such as UL- Urbana-Champaign and Syracuse University. Some of them were short of data science certificates such as Drexel University and Florida State University.

To answer the research question 2, UW(Seattle), UNC (Chapel Hill), UM (Ann Arbor), UT(Austin), IU(Bloomington), University of Pittsburgh show there were actually preparing their students for data science careers by showing they had all the essential components. These schools have similar characteristics to the 4 gold standards and possessed all features

Ranking	University	Courses	Specializations	Professors	Certificates	Capstone/Practicum requirements
1	UL, Urbana-Champaign	X		X		X
2	UW, Seattle	X	X	X	X	X
3	UNC, Chapel Hill	X	X	X	X	X
4	Syracuse University	X		X	X	X
5	UM, Ann Arbor	X	X	X	X	X
6	UT, Austin	X	X	X	X	X
7	Rutgers, The State University of New Jersey	X		X		
8	UM, College Park	X		X		X
9	IU, Bloomington	X	X	X	X	X
10	University of Pittsburgh	X	X	X	X	X
11	Drexel University	X		X		X
12	Florida State University	X		X		
13	Simmons College	X		X		X
14	UC, Los Angeles	X		X		
15	UW, Madison	X		X	X	X
15	UW, Milwaukee	X	X	X		X

2. Program description

Table 11. Length of program description

The longest program description of the top 15 data science programs in information schools was by Syracuse University (L = 353), while Rutgers had the shortest description (L = 114). The average length of program description in information schools was about 232, which is much shorter than the length in the 4 gold standards (mean = 501). In order to supply the students with enough information to understand the programs, program descriptions in information schools might need to be more detailed and comprehensive.

To answer the research question 3, one of the descriptive statistics of data science programs in information school was that the length of the description ranged from 114 to 353, and the mean was 232.

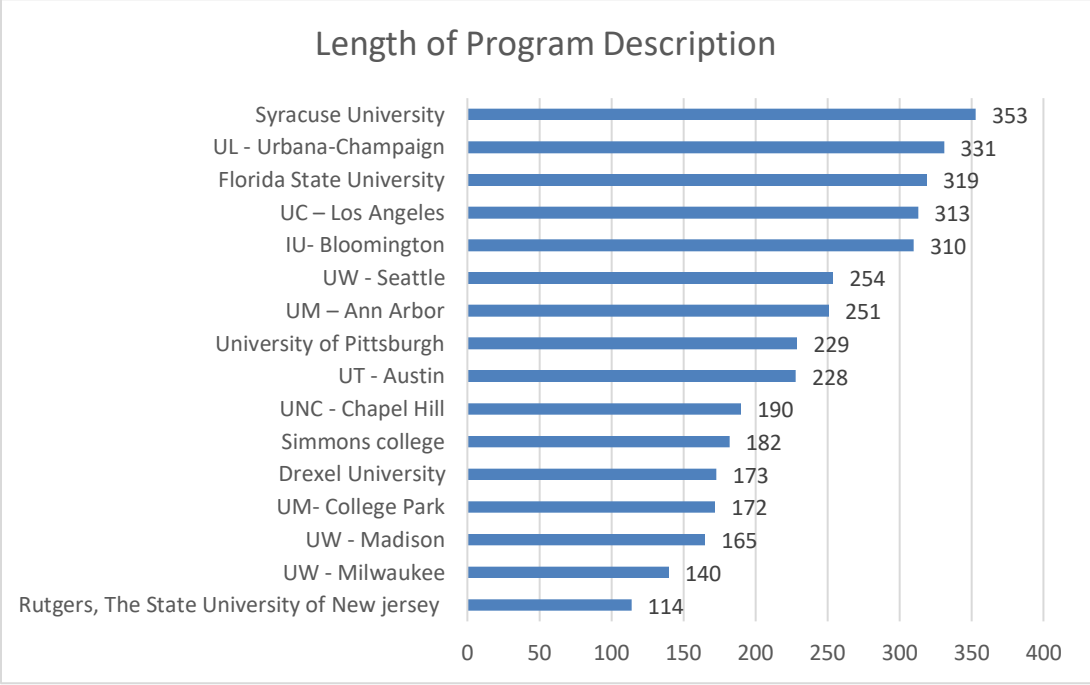


Table 12. Number of unique words

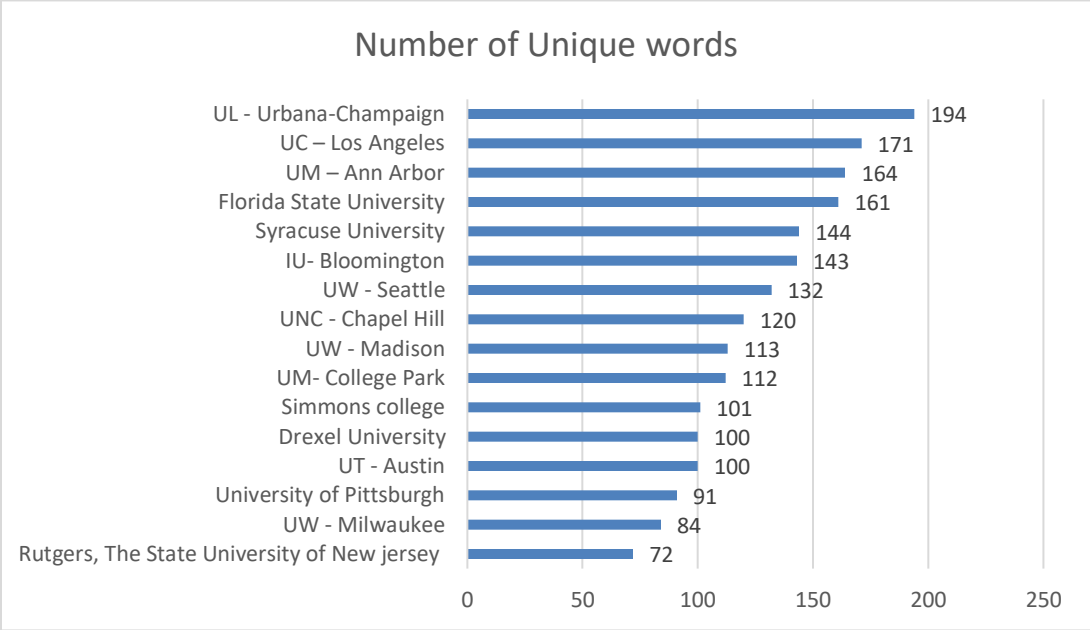
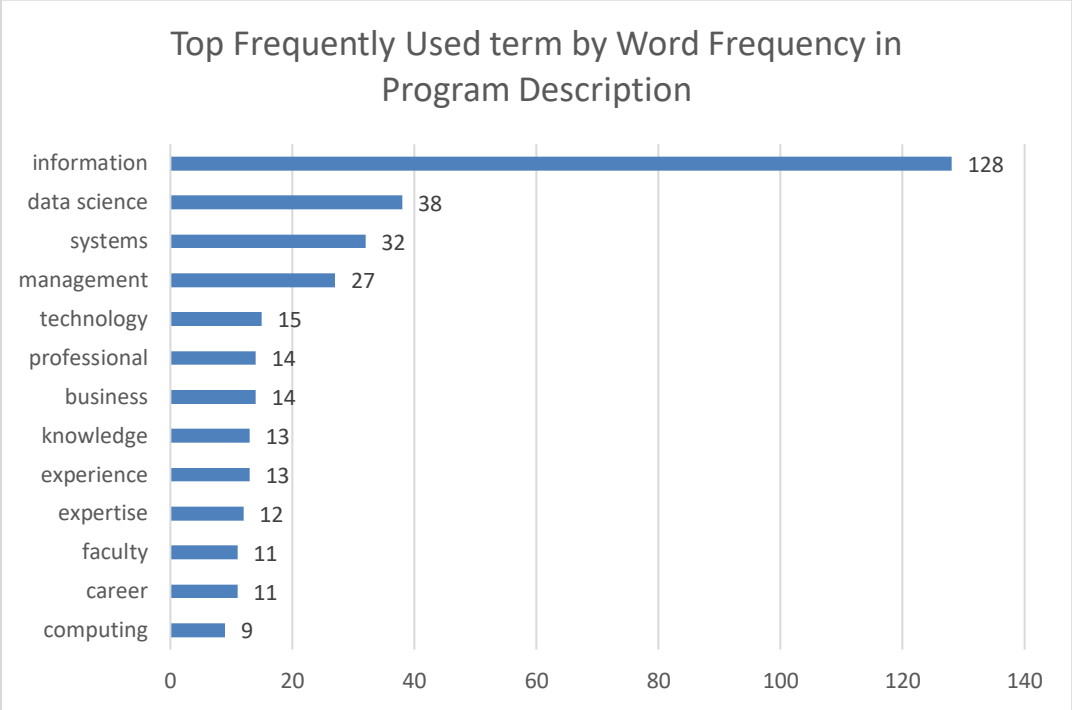


Table 12 showed the number of unique words used in program descriptions in information schools. UIUC had the highest number of unique words (N = 194), and Rutgers had the lowest number of unique words (N = 72). Therefore, Rutgers had the lowest number of unique words as well as the length of the program description. The average length of unique words in information schools was about 125, larger than the length in the 4 gold standards (mean = 83).

To answer the research question 3, one of the descriptive statistics of data science programs in information schools was that the number of unique words in program description ranged from 72 to 194, and the mean was 125.

Table 13. Top frequently used term by word frequency in program description



Unlike the study had for the 4 gold standards, Data science program description in information schools tends to use the word "information" more frequently. The other top 3 most frequent words are "data science," "systems," and "management." Table 13 could confirm the majority of information schools were actually preparing their students for data science careers and the information schools care about the most because the most frequent words are related to data science.

To answer the research question 3, one of the linguistic characteristics of data science programs in information schools was that the they tended to use words related to data science careers such as "Information," "data science," "professional."

3. Program curriculum structure

Table 14. Number of credit hours (core vs. electives)

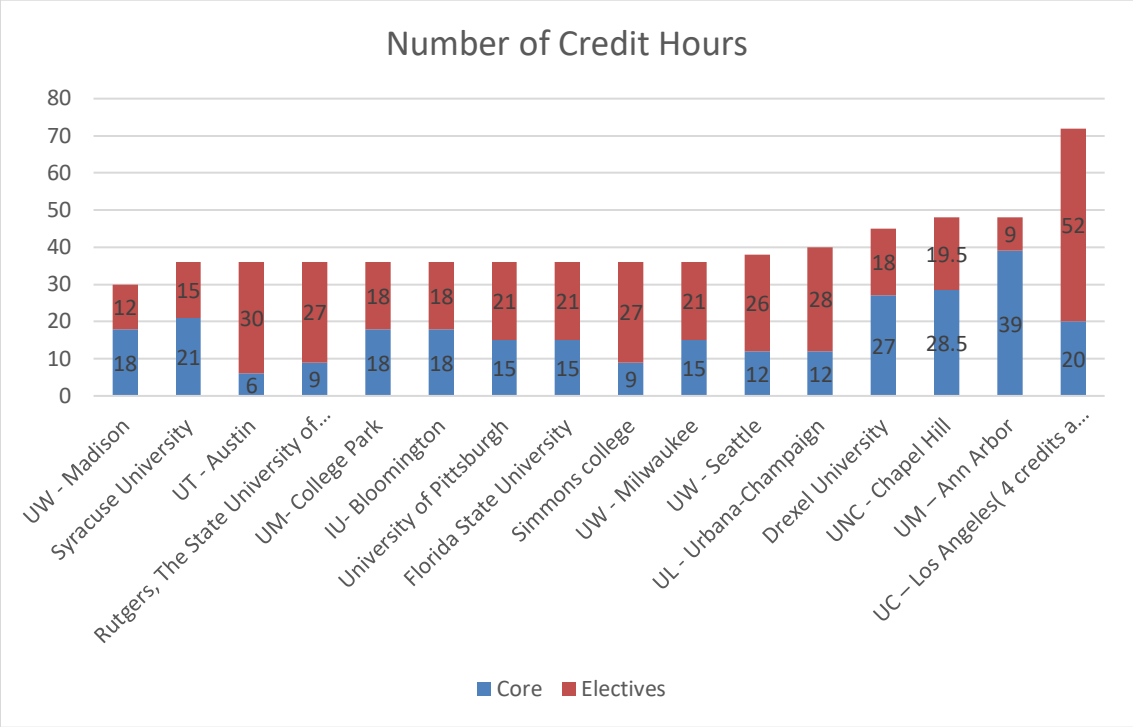


Table 16 summarizes the credit hours required by each university. The average number of credit hours over 16 data science programs was about 39 credit hours, with about 17 in the core courses and 22 in elective courses. UCLA had the highest number of core courses (n = 20) and total courses (n = 72). The result of the table could not be used to compare with master programs in iSchools because of the lengths of different types of academic degrees. Bachelor's degree normally takes students 4 years to complete, while a master's degree normally takes students 2 years to complete. However, the study could use a ratio to compare with the 4 gold standards. The ratio between core credit hours and total credit hours in information schools was 0.44 while 0.71 in the 4 gold standards. The 4 gold standards provided less flexibility in the curriculum than information schools.

To answer research question 1, the credit hours of completing various data science program in information schools ranges from 30 to 72. The ratio between core credit hours and total credit hours in information schools was 0.44, which means that

information schools provided more flexibility in program curriculum than in the 4 gold standards and allowed students to design their own career preparation plans.

Table 15. Number of courses listed on program websites (core vs. electives)

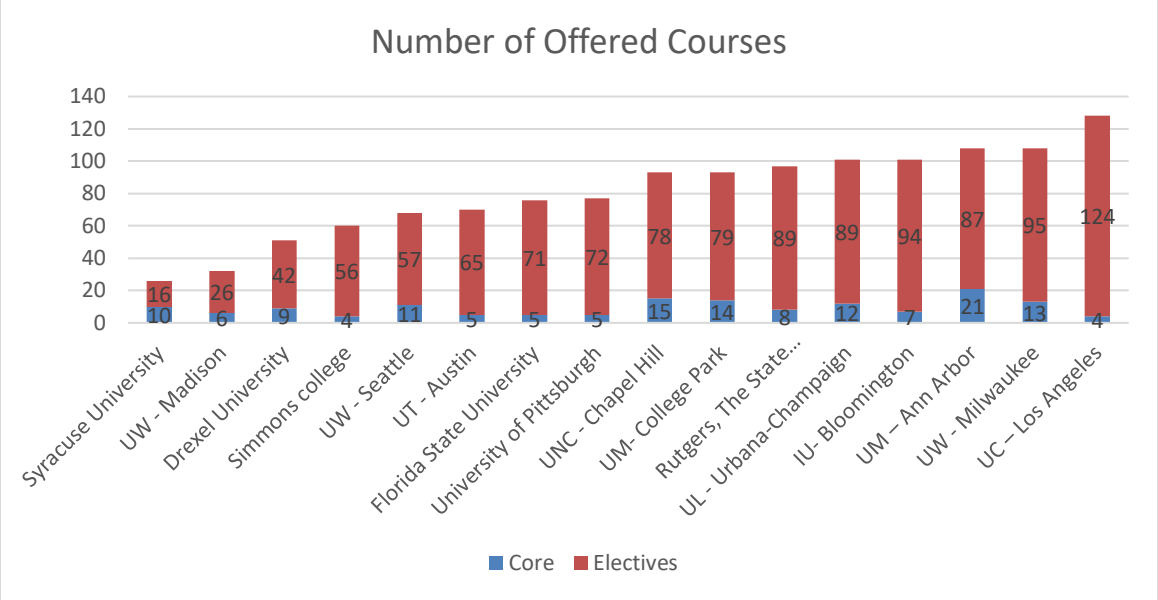


Table 15 shows the number of offered data-science-related courses. The courses listed on each program website focused on fields of data management, data science, or big data. The titles of the offered courses must contain terms such as database management, big data, and other related terms.

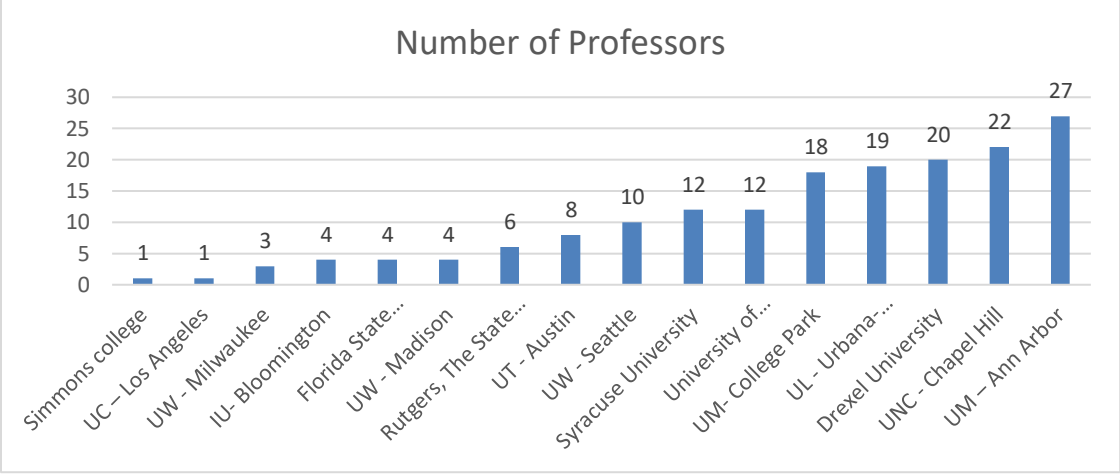
Further analyses were conducted about the number of courses listed on program websites. UCLA had the highest number of Core courses (n = 72) and electives courses (n = 89), which makes sense because UCLA required the highest number of credit hours. In contrast, Syracuse University had the lowest number of core courses (n = 10) and electives courses (n = 16). Table 15 shows that the majority of iSchools have a number of courses for their student to choose, which confirmed these schools are serious about preparing their students for data science careers.

The ratio between core courses offered and total courses offered in information schools was 0.13 while 0.43 in the 4 gold standards. The pattern in number of courses offered was similar to the pattern in credit hours, which means that information schools provided more flexibility in course selection than in the 4 gold standards and allowed students to design their own study plan.

Table 16. Number of professors

Table 16 shows the number of professors whose research area focuses on data science or a part of data science for each program. UM – Ann Arbor had the highest number of professors with a focus on data science (n = 27), while Simmons and UCLA had the lowest number of professors (n = 1).

To answer the research question 2, Table 16 showed that certain iSchools have lots of professors focusing on data science, which benefits students by providing quality courses and research opportunities. Table 16 confirmed certain information schools were serious about preparing their students for data science careers, but some weren't. For example, Simmons College and UCLA only employed 1 professor with a focus on data science.



4. Course focus

Table 17. Percentage of data science, mathematics & statistics, communication, visualization, and information skills in all courses

列1	Data Science	Mathematics	Visualization	Informat	Communic	Ethic	others
UL - Urbana-Champaign	28.71%	11.87%	8.91%	34.80%	9.54%	2.93%	3.24%
UW - Seattle	31.71%	8.84%	5.88%	39.06%	7.21%	4.45%	2.85%
UNC - Chapel Hill	29.03%	3.23%	3.23%	47.01%	4.30%	3.23%	9.98%
Syracuse University	53.85%	3.85%	3.85%	19.23%	3.85%	3.85%	11.53%
UM – Ann Arbor	32.41%	9.26%	6.48%	27.78%	8.33%	0.08%	15.66%
UT - Austin	21.43%	2.86%	8.57%	25.71%	8.57%	5.71%	27.14%
Rutgers, The State University of New	24.74%	6.19%	7.22%	28.86%	5.15%	2.06%	25.78%
UM- College Park	20.62%	4.30%	5.38%	46.24%	4.30%	5.38%	13.79%
IU- Bloomington	13.86%	3.96%	2.97%	36.63%	8.97%	4.95%	28.66%
University of Pittsburgh	38.96%	15.58%	5.19%	27.27%	3.90%	2.60%	6.51%
Drexel University	52.94%	5.88%	9.80%	11.76%	7.84%	3.92%	7.85%
Florida State University	19.74%	3.95%	0.00%	40.79%	6.58%	1.32%	27.64%
Simmons college	16.66%	0.00%	3.33%	31.66%	5.00%	1.66%	41.69%
UC – Los Angeles	11.72%	3.91%	3.91%	27.34%	12.50%	2.34%	38.29%
UW - Madison	28.13%	0.00%	3.13%	46.88%	6.25%	3.13%	12.50%
UW - Milwaukee	33.33%	1.85%	4.63%	29.63%	0.00%	3.70%	26.86%

In terms of the focus of all courses, the highest percentage of courses covering data science was the program from Syracuse University (N = 53.85%), and UNC-Chapel had the highest percentage of courses covering information skills (N = 47.01%). University of Pittsburgh had the highest percentage of courses covering Math/Stat (N = 15.58%), and Drexel University had the highest percentage of courses covering visualization skills (N = 9.8%). UCLA had the highest percentage of courses covering Communication skills. Different from the 4 gold standards, Data science programs in information schools hosted courses that talked about ethics in data science. Not surprisingly, Information Schools provided a lot more courses covering information skills than the 4 gold standards.

It was interesting to see not all schools offered every kind of course. For example, Simmons college and UW – Madison had none of the Math/Stat courses. Florida State

University had none of the Visualization skills courses, and UW – Milwaukee had none of the communication courses. On the contrary, the 4 gold standards offered all courses that covered all 6 areas, the 4 gold standards were good at hosting a boarder range in their course's offerings. To become a successful data science program, data science programs in information schools should expand their course offerings and offered all the essential courses like the 4 gold standards.

Analysis of Data Science Preparation from 2 Sample Lists

1. Comparison

By comparing to the results from the 4 gold standards, the differences were:

1) A number of data science programs in iSchools didn't have all the essential components that the 4 gold standards had.

2) Length of program description in information schools was relatively low, only ranges from 114 to 353.

3) Range of number of unique words in program description of information schools is larger, which allows the schools provided more information about what the program was about.

4) Obviously, program description in information school tends to use the word "information" more frequent. Words like "management" and "professional" were used more frequent too. This comparison concluded that master programs in information schools aim to prepare their students for big data careers.

5) Data science programs in information schools provided more electives core and fewer core courses than the 4 gold standards. This also applies to credit hours. The result of the table could not be used to compare with the 4 gold standards because of the type of academic degrees. Bachelor's degree normally takes students 4 years to complete, while a Master's degree normally takes students 2 years to complete. However, the study could use ratios to compare with the 4 gold standards. The ratio between core credit hours and

total credit hours in information schools was 0.44 while 0.71 in the 4 gold standards. The ratio between core courses offered and total courses offered in information schools was 0.13 while 0.43 in the 4 gold standards. The 4 gold standards provided less flexibility in curriculum and course selection than information schools.

6) Overall, Information schools employed fewer professors with a focus on data science than the 4 gold standards. Information Schools provided a lot more courses covering information skills than the 4 gold standards. The majority of them focused on data science and information skills only while paid a lot less attention to Mathematics/Statistic.

2. Program Overview

This part of the study concentrated on the progress made by iSchools to incorporate certificates, specializations, courses, and capstone/practicum requirements and also the efforts made by the schools to employ professors with data-science-related backgrounds.

The collected data is presented in table 12. An analysis of this table reveals that while most of the top library and information studies universities offer many classes related to data science and employ several data science professors, there is still significant room for improvement, particularly regarding certificates and specialization inside the curriculum. Schools that do not offer data science specializations and certificates should start offering them, as that can be helpful to students seeking to start careers in the data science industry.

In this study, information about the top 4 ranked undergraduate data science

programs were considered as the golden standard. Among the top 4 ranked undergraduate data science programs, all of them included a capstone project or internship requirement as part of their degree requirements. Similarly, 12 out of 16 data science programs in information schools included an internship requirement. Comparing program overviews reveals that a successful data science program should have at least a capstone project or internship requirement. According to Larry (2014), "[ninety-two] percent of the respondents attributed a great deal of importance to the capstone design course, with 59% (n = 175) reporting that it was extremely important and 33% (n = 98) reporting that it was very important." This shows that an internship or capstone project can better help students utilize the skills they learned at school and apply them to the actual workplace. Therefore, universities such as Rutgers, Florida State University, and UCLA should add a practicum requirement to their program curriculum.

3. Program Description

In this investigation, program descriptions and the definition of the term "data science" published on the program websites of each school were coded and analyzed. Particularly, the study counted the frequency of words that were used in these program descriptions.

The analysis of program descriptions' results from 16 data science programs shows that the length of the program description varies. There were also variations in terms of the number of unique words. One interesting fact was that not all the data science program descriptions defined the term "data science." Most data science definitions were found on the program pages for MSc Data Science or data science certificates. As a result, these programs had the highest number of unique terms and

longest program description. Furthermore, the majority of iSchools data science programs did not have comprehensive program descriptions as in the 4 gold standards, which didn't supply the students with enough information to understand the programs.

As indicated by O'Neil (2014), the data science program is multidisciplinary. The analysis of the frequency of unique terms used in different data science program descriptions seemed to conform to this finding. Unique and common words such as systems, management, business, and computing reflect the multidisciplinary nature of data science programs. Graduates of these programs are educated to be multiskilled.

To answer the research question 3, the descriptive statistics of data science programs in information school were: 1) The length of the description ranged from 114 to 353 and the mean was 232. 2) The number of unique words in the program description ranged from 72 to 194 and the mean was 125. One of the linguistic characteristics of data science programs in information schools was that the they tended to use words related to data science careers such as "Information," "data science," "professional."

The average length of unique words in information schools was about 125, which is larger than the length in the 4 gold standards (mean = 83). The average length of program description in information schools was about 232, which is much shorter than the length in the 4 gold standards (mean = 501). In order to supply the students with enough information to understand the programs, program descriptions in information schools might need to be more detailed and comprehensive.

4. Program Curriculum Structure

This investigation gathered data about program requirements. These data include

the total number of data science courses offered, core courses, total credit hours, and elective courses. The data also showed whether capstone courses were offered or not.

Compared to the top 4 undergraduate data science programs, it is interesting to find that data science programs in iSchools had a lower ratio between core credit hours(courses) and total credit hours(courses). Because of the limitation of the length of graduate programs, which is only 2 years, it is reasonable that graduate programs have fewer core credit hours because they assume students have certain related backgrounds. Some programs greatly emphasize their elective curriculum, as we can see that the core courses and elective courses ratio is about 1:4, including UT, Austin; Rutgers; Simmons College; UIUC; and UCLA. However, other programs devote remarkably more credit hours to their core courses. The curriculums of these programs were fixed, and they didn't offer too much flexibility for their students to take elective courses. Significantly, all four schools in the golden standard follow this pattern, and they highly emphasize their core curriculum. **Based on this observation, schools are better off forming a fixed curriculum with limited elective courses.**

The top 4 undergraduate data science programs listed more courses on their websites than the programs in information schools on average, and these undergraduate programs required a higher number of credit hours. It should be noted that the advertisements on websites of several schools are falling behind the courses they actually offer, because there are gaps between the courses they offer, and the credit hours required. Schools should update their course offering regularly to make sure prospective students can plan their schedules in advance.

All programs in information schools (100% or 16 out of 16) employed professors whose research area was within the data science industry. Of these, the University of Michigan employed the most, with 27 professors, and UNC, Chapel Hill, employed the second most, with 22 professors. A small number of schools employed only 1 or 2 professors, including Summons College and UCLA.

5. Course focus

Results from this session reveal that information schools are specializing in offering a broader selection of core courses. However, information schools lacked elective courses covering mathematics/statistics and communication skills. Summons College offers no mathematics course, and UW, Milwaukee, offers no communication course. Information schools had the lowest ratio in covering these courses. However, they are considerably good at providing data science courses and information courses. One possible reason behind that is master's programs often assume their incoming students already have a fundamental background in math or computing skills.

Overall findings

Each program in the study was assigned to a ranking: poor, average, and excellent. This ranking stands for the perceived commitment of each program to preparing students for entering the data science industry. The criteria were based on the data collected in this study. The criteria were also based on four categories: courses, specializations and certificates, professors, and capstone/practicum requirements. Each category was separated into different parts identifying an excellent, average, or poor ranking. These rankings are determined by the following conditions:

Poor: Poor in more than 1 category and average in the rest, or poor in all categories.

Average: Average in more than 1 category and poor/excellent in the rest, or average in all categories.

Excellent: Excellent in more than 1 category and average in the rest, or excellent in all categories.

Courses:

Poor: Offers less than 67 courses

Average: Offers between 67 and 97 courses

Excellent: Offers more than 97 Courses

Specializations and Certificates

Poor: Offers none of specializations or certificates

Average: Offers 1 or more certificates and 0 specializations, or offers 1 or more certificates and 0 specializations

*Excellen*t: Offers 1 or more certificates and 1 or more specializations

Professors:

Poor: Offers less than 6 Professors

Average: Offers between 6 and 12 Professors

Excellent: Offer more than 12 Professors

Capstone/Practicum requirement:

Poor: has 0 Capstone/Practicum requirements

Excellent: has 1 or more Capstone/Practicum requirements

Table 9. Overall Rankings

Ranking	University	Data Science commitment
1	UL, Urbana-Champaign	Excellent
2	UW, Seattle	Average
3	UNC, Chapel Hill	Excellent
4	Syracuse University	Average
5	UM, Ann Arbor	Excellent
6	UT, Austin	Average
7	Rutgers, The State University of New Jersey	Poor
8	UM, College Park	Average
9	IU, Bloomington	Excellent

10	University of Pittsburgh	Average
11	Drexel University	Poor
12	Florida State University	Poor
13	Simmons College	Poor
14	UC, Los Angeles	Poor
15	UW, Madison	Average
15	UW, Milwaukee	Average

Conclusion

This study observed the top 16 accredited library and information schools (iSchools) in the United States and their methods for teaching data science, with regard to courses, credit hours, curriculum, certificates, professors, specializations, and internship programs. Information schools are serious about preparing their students for data science careers. The results revealed that while schools offered plenty of data science courses and employed a few professors, it is not consistent between programs. Also, while several schools performed better in helping their students better prepare for entering the data science industry, certain information schools have decided to focus on library science only and did not put as much effort as others into data science. For schools with a clear data science focus, this study provides program coordinators with a ranking of data science programs in information schools, enabling the schools to better prepare their graduates for careers in data science. Based on the fact that most of the information schools offer plenty of data science and information courses, this study concluded that information schools are teaching the right skill sets needed for their graduates to excel in their data science careers.

These findings allow students to have the chance to learn about data science programs in different information schools, and how these schools are preparing their programs for data science from reputable sources. Besides, these findings provide graduates with the opportunity to network within the data science industry, therefore offering them advantages over their peers when starting their careers.

Impacts and Limitations

The potential stakeholders of this study are library and information science schools. For information schools that do not have a clear data science focus, this study will give them the initiative to add a data science specialization to their programs. For schools with a clear data science focus, this study provides them with a ranking of data science programs in information schools, enabling the schools to better prepare their graduates for careers in data science. For example, the data science program in information schools might devote some courses to covering advanced data analytics skills; however, the schools do not have any courses covering statistics and math courses, which are essential for the success of advanced data analytics skills. Additionally, the findings from this study will be beneficial to all students who are entering information schools. Students who want to pursue careers in data science need to know which information science schools do the best to prepare graduates for data science careers. Because the best practices for preparing students for data science careers were identified in this study, faculties in these programs can learn how to improve their courses based on the findings of this study.

Further, because this study only serves as the initial step in describing and summarizing the current stage data science program in the United States, there are some limitations and delimitations of this study: 1) The sample size is not uniform, and it may cause some biases since the sample is selected according to the U.S. News & World Report ranking. 2) This study does not include all information schools in the United

States. Other iSchools in the United States might have specializations that are different from library science or information science. 3) This study does not cover schools that are outside of the United States. There might be a global shared culture among iSchools in terms of data science focus, but universities from different countries have their own systems. Thus, it would be better to limit research to only schools in the United States, which may add credibility to the study.

This study only shows the initial step in describing the current stage of data science programs in information schools. Further comprehensive research can be done as a follow-up to this study, especially gathering qualitative and quantitative data in interviews, focus groups, and surveys from students, leaders, faculty, and graduates of the data science program. The information collected from surveys and interviews will offer more suggestions regarding the best practices of data science programs.

Bibliography

Bohlouli, M., Sadeghi, B. B., Narimani, Z., Vasighi, M., & Ansari, E. (Eds). (2020). *Data science: From research to application.* Springer.

Dhar, V. (2013). Data science and prediction. *Communications of the ACM, 56*(12), 64–73.

Dumbill, E., Liddy, E. D., Stanton, J., Mueller, K., & Farnham, S. (2013). Educating the next generation of data scientists. *Big Data, 1*(1), 21–27. https://doi.org/10.1089/big.2013.1510

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods, 10*(1), 5–34. https://doi.org/10.1177/1094428106289252

Dzemyda, G., Bernatavičienė, J., & Kacprzyk, J. (Eds). (2020). *Data science: new issues, challenges and applications.* Springer.

EMC. (2011). *Data science revealed a data-driven glimpse into the burgeoning new field*. http://www.emc.com/collateral/about/news/emc-data-science-study-wp.pdf.

Hay, I. (2005). *Qualitative research methods in human geography* (2nd ed.). Oxford University Press.

Hilbert, M., & Lopez, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science, 332*(6025), 60–65. DOI: 1126/science.1200970

Kampakis, S. (2019). *The decision maker's handbook to data science: a guide for non-executives, managers, and founders.* Apress.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011, June). *Big data: The next frontier for innovation, competition, and productivity.* McKinsey & Company. Retrieved October 10, 2020, from http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

McKenzie, L. J., Trevisan, M. S., Davis, D. C., & Beyerlein, S. W. (2004). Capstone design courses and assessment: A national study. In *Proceedings of the 2004 American society of engineering education annual conference & exposition* (pp. 1–14).

Morse, R. & Flanigan, S. (2013, September 9). *How U.S. News calculated the 2014 best colleges rankings*. U.S. News & World Report. Retrieved October 10, 2020, from http://www.usnews.com/education/bestcolleges/articles/2013/09/09/how-us-news-calculated-the-2014-best-colleges-rankings

O'Neil, M. (2014, February 3). *As data proliferate, so do data-related graduate programs.* The Chronicle of Higher Education. Retrieved April 03, 2021, from https://www.chronicle.com/article/as-data-proliferate-so-do-data-related-graduate-programs

Power, D. (n.d.). *What universities offer masters degrees in analytics and data science?* DSS Resources. http://dssresources.com/faq/index.php?action=artikel&id=250.

van der Aalst W. (2016) Data science in action. In *Process mining*. Springer. https://doi.org/10.1007/978-3-662-49851-4_1

Wilson, T. (1989). Towards an information management curriculum. *Journal of Information Science, 15*(4-5), 203–209. DOI: 10.1177/016555158901500403

GilPress, (2012). *Graduate Programs in Big Data Analytics and Data Science*. http://whatsthe bigdata.com/2012/08/09/graduate-programs-in-big-data-and-data-science/.

Tang, R., & Sae-Lim, W. (2016). Data science programs in US higher education: An exploratory content analysis of program description, curriculum structure, and course focus. *Education for Information*, *32*(3), 269-290.

Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research.

Why colleges are offering data science programs. (n.d.). Retrieved April 20, 2021, from https://www.usnews.com/education/best-colleges/articles/why-more-colleges-are-offering-data-science-programs

11 data Scientist skills employers want to see in 2021. (2021, March 24). Retrieved April 20, 2021, from https://bootcamp.berkeley.edu/blog/data-scientist-skills/

Walker, R. (2015). From big data to big profits. doi:10.1093/acprof:oso/9780199378326.001.0001

Jain, A. (2019, August 02). Https://www.analyticsvidhya.com/blog/2016/07/10-analytics-data-science-top-universities-masters-usa/. Retrieved April 20, 2021, from https://www.analyticsvidhya.com/blog/2016/07/10-analytics-data-science-top-universities-masters-usa/

Appendix A. List of Study Samples by University (Department)

- University of Illinois, Urbana-Champaign (School of Information Sciences)
- University of Washington, Seattle (Information School)
- University of North Carolina, Chapel Hill (School of Information and Library Science)
- Syracuse University (School of Information Studies)
- University of Michigan, Ann Arbor (School of Information)
- University of Texas, Austin (School of Information)
- Rutgers, The State University of New Jersey–New Brunswick (School of Communication and Information)
- University of Maryland, College Park (College of Information Studies)
- Indiana University, Bloomington (Department of Information and Library Science)
- University of Pittsburgh (School of Computing and Information)
- Drexel University (College of Computing & Informatics)
- Florida State University (College of Computing & Information)
- Simmons College (School of Library and Information Science)
- University of California, Los Angeles (Department of Information Studies)
- University of Wisconsin, Madison (The Information School)
- University of Wisconsin, Milwaukee (School of Information Studies)