Title: The Data Gold Rush in Higher Education

Authors: Jevin D. West and Jason Portenoy

Abstract: The recent explosion of interest in big data and data science has been accompanied by a rise in programs and venues to train budding data scientists in relevant skills. This chapter explores the emergent landscape of data science education in the United States. This includes formal education options—largely Master's degrees and certificate programs geared towards business and industry applications—as well as less expensive and less formalized alternative options (MOOCs, community workshops, meetups) that have appeared to meet other demands, such as opportunities in nonprofits, government, and science. The growth in training options is examined as a vitality indicator of the young field of data science. The chapter further explores the pull of thought leaders in data science and related fields away from academia and towards industry, which presents a potential challenge to the longevity of data science as a formal discipline.

Keywords: Data Science, STEM education, master's programs, Big Data, alternative education, MOOCs, eScience, software education

The Data Gold Rush in Higher Education

Jevin D. West and Jason Portenoy

Introduction

The enthusiasm for all things big data and data science is more alive now than ever. It can be seen in the frequency of big data articles published in major newspapers and in the venture capitalists betting on its economic impact (Press 2014). Governments and foundations are calling for grant proposals, and big companies are reorganizing in response to this new commodity (Gordon and Betty Moore Foundation 2013; National Science Foundation 2015). Another, often overlooked, vitality indicator of data science comes from education. Students are knocking down the doors at universities, online MOOCs, and workshops. The demand for data science skills is at an all time high, and universities are responding.

According to a widely cited 2011 report by the McKinsey Global Institute, the United States could face a shortfall of 140,000 to 180,000 people with deep analytical talent—those trained to realize the promise of big data—by 2018 (Manyika et al. 2011). Predictions like these have led to a rapid proliferation of educational programs to train this talent, popularly termed “data scientists”. In this chapter, we explore the emergent landscape of education in this field as institutions around the world race to get out ahead of the forecasted shortfall. We see this growth as an indicator of the vitality of the data science field. As students become formally trained and self-identify as data scientists, data science is more likely to establish as its own discipline.

Big data is as much a business opportunity for universities as it is for the venture capitalists or the start-ups in Silicon Valley. The McKinsey report focuses mostly on the promise of big data to provide financial value to businesses, and the educational landscape reflects this. The vast majority of new programs to train these data scientists are Master's degrees and certificates geared largely toward people going into private industry. These programs focus on giving marketable skills to students, and colleges make money by charging tuition commensurate with this promise.

The picture of formal education to train data scientists that we find is a somewhat disjointed one, the product of rapid, speculative growth. There is a good deal of variation in the departments and degree titles attached to the emerging programs. There are many elements the curricula have in common, but these too exhibit some subtle variation. The programs tend to have a strong business focus evident in their branding and coursework.

Apart from the formal and expensive education options, a number of alternative forms of education have emerged to train people in the fundamentals of data science. Many of these are free or very inexpensive. They include large-format online classes, online training websites, organized groups of individuals interested in mentoring people within their communities, and more loosely organized meetup groups. Many of these options fill roles that are under-addressed by the business-dominated degree programs.

Finally, we explore another aspect of education and data science: the pull of thought leaders in data science away from academia and toward industry. While the rise in data science training programs is a positive indicator of the vitality of the field, we recognize that there is a need for retaining some of this talent within the walls of academia for it to take hold as a true academic discipline. We find that the disparity in resources and an outdated career structure are leading many researchers to choose to leave academic research. We consider efforts underway to address this situation, and what it might mean for the future of the field.

Big Demand for Data Science Degrees

Since the late 2000s, hundreds of degree programs in data science and related fields have formed worldwide. The growth has become even more intense since 2011 when the McKinsey report was released (see Figure 1). The landscape of data science education reflects a response to the dire situation portrayed in that report: The economy desperately needs more data scientists and schools are jumping at the opportunity to train them.

The overwhelming majority of the new programs are professionally-oriented Master's degrees and certificate programs.[1] The McKinsey report is often mentioned in the promotional materials for the programs. They point out the deluge of data overwhelming businesses, the promise that it can hold, and the fevered demand for data scientists to harness it. In a 2012 article, *Harvard Business Review* identified the data scientist as "the sexiest job of the 21st century". This same article noted that there were "no university programs offering degrees in data science", although some programs in analytics were starting to alter their curricula to take advantage of the big data boom (Davenport and Patil 2012). The change in the short period since that article was published has been dramatic (see the rapid rise in "data science" programs evident in Figure 1). Budding data scientists now can choose among a host of colleges eager to supply companies with the talent they are looking for.

[Figure 1 goes here (approximately)]

[Table 1 goes here (approximately)]

[Table 2 goes here (approximately)]

---

[1] A note on methods: After searching the web for information on degree programs being offered in data science, we settled on using the data set offered by North Carolina State University at http://analytics.ncsu.edu/?page_id=4184, which claims to be a comprehensive list of "programs offering graduate degrees in Analytics or Data Science (or closely similar) at universities based in the U.S." The NCSU list includes information on duration and estimated cost of programs. We gathered information about the curricula of these data science programs by examining the websites and promotional materials of a subset of these programs. As the field matures, more exacting methods for program tracking will be possible.

The buzz of big data has not escaped the halls of academia or the offices of university presidents. The cost of completing a Master's degree is between \$20,000 and \$70,000 for a one- or two-year program (costs can be lower for state schools, especially tuition rates for residents). Tables 1 and 2 show the range in cost of these Master's programs.[2] The tuition charged for these degrees reflect the high earnings potential that the colleges expect their graduates to have in the job market. This idea is sometimes made explicit. The Frequently Asked Questions for the University of San Fransisco's M.S. in Analytics—tuition \$43,575—states,

> "We are proud to run a program that, with high probability, significantly increases the earnings power of our graduates over the long run. . . . We are confident that the return on investment associated with this particular professional program is superior to the return on investment from many other forms of professional training (law, medicine, etc.). There is a shortage of data scientists on the job market right now, and that shortage is projected to get far worse before it gets better."[3]

North Carolina State University estimates that graduates should be able to recoup the cost of tuition for its M.S. in Analytics program—\$40,800, \$23,600 for state residents—plus fees and one year of lost wages, in 19 to 26 months. They also

[2] These numbers are in rapid flux, and in fact changed substantially between versions of this chapter.

[3] http://www.usfca.edu/artsci/msan/faq/

report the average net 3-year return on investment for graduates to be $131,800 to $136,400.[4]

Many of these programs are too new to report any sort of outcomes data for their graduates. Those that do, claim very high job placement rates, often 100%. They report starting salaries in the high five or six figures, many with signing bonuses, although they generally only have these numbers for one or two cohorts. The types of positions reporting graduates tend to go into include data scientist, analyst, consultant, and manager, in large companies in industries such as finance, technology, consulting, e-commerce, and health care.

Owing to both the interdisciplinary nature of the field and the rapid proliferation of programs spurred by the urgent demand, the landscape of data science programs gives a somewhat fractured impression. The degrees are commonly named “Analytics,” “Data Science,” or “Business Analytics”—these programs all tend to offer very similar curricula, with some subtle variations. Some schools extend these titles, such as St. John's University's M.S. in Data Mining and Predictive Analytics, or Carnegie Mellon's M.S. in Computational Data Science.

In addition to the naming of the programs, the inchoate nature of data science education is reflected in the variety of academic departments that offer the degrees, and the lack of any natural academic home (Finzer 2013). Business

---

[4] http://analytics.ncsu.edu/?p=7178

schools are one common home for these programs. Many of them brand themselves as specialized business degrees—these are often the “Business Analytics” programs (Gellman 2014). Other programs are housed in departments of Statistics, Computer Science, Engineering, or Information Science. Another common practice is to have several academic departments collaborate to offer the program, for example Northeastern University's M.S. in Business Analytics, a collaboration among the Business School, the College of Computer and Information Sciences, and the College of Social Science and Humanities. At the University of Washington, there are collaborative efforts around campus departments to offer transcriptable options in data science in various fields from biology to information science. Some colleges form “institutes” out of these collaborations, such as the University of Virginia, whose Data Science Institute includes faculty from Computer Science, Statistics, and Systems and Information Engineering.

The data science/analytics curriculum reflects the skills that someone will need in this career, to the extent that that has been defined in this new field. Data scientists are expected to be able to analyze large data sets using statistical techniques, so statistics and modeling are typically among the required coursework. They must be able to find meaning in unstructured data, so classes on data mining are also usually part of the core. They often must be able to communicate their findings effectively, so courses on visualization are commonly offered as an elective. Other courses that students may take include research design, databases, parallel computing, cloud computing, computer programming, and machine learning; all of these reflect skills that an employer might expect from a data scientist. Courses in business and marketing are also common, especially among programs in business schools. Most programs also require a capstone project that gives students experience in working through real world problems in teams. The range of courses is indicative of an expectation that data scientists be strong in many different areas. Bill Howe, who teaches data science at the University of Washington, is not even sure that a curriculum to train data scientists is feasible: "It remains to be seen," he says, ". . . What employers want is someone who can do it all" (Miller 2013).

Alternative Data Science Education

The landscape of data science degree programs reflects a market dominated by private industry, and this can come at the expense of other areas of society—academia, nonprofits, and the public sector. The promise of big data extends to a wide range of applications. The value of data science is more than the money that can be made or saved by an organization. One way that this gap is being addressed is through several alternative data science training venues apart from the formal training offered through college degree programs. Lacking a solid revenue structure or robust funding, these efforts are largely grassroots and community-driven, and rely heavily on volunteers and the passion and interest of those involved. Some of these alternatives are large-format online classes called MOOCs and other forms of online instruction, organizations that run in-person workshops and meetings to offer instruction in communities, and informal meetup groups.

Massive open online courses, or MOOCs, are free or low-cost courses taught through universities or other organizations. They typically use short videos, interactive quizzes, and other assignments that can be easily disseminated to a large online audience. Courses can be offered either on a schedule or on-demand at the student's pace; in either case, the format requires a good deal of self-motivation, and completion rates tend to be low (Liyanagunawardena, Adams, and Williams 2013). Some courses offer a certificate of completion at the end. MOOCs are a very low cost alternative to formal education options, but they lack the formal recognition and networking opportunities of the more expensive degree options. Two popular MOOCs for data science are offered on the MOOC platform Coursera: a Stanford University course on Machine Learning[5] created by Andrew Ng, a computer science professor and co-founder of Coursera; and a University of Washington course titled “Introduction to Data Science.”[6]

In addition to structured MOOCs, there exist a number of options online for self-study of data science techniques and tools. Many standard data science

[5] https://www.coursera.org/course/ml
[6] https://www.coursera.org/course/datasci

tools are open source—Python, R, D3, Weka—and have active online communities. These options tend to be free or very inexpensive. To learn the basics of Python, for example, one can use the tutorials on Codecademy[7], Google's Python Class[8], Python's own tutorial[9], or a number of other options in a variety of formats and levels.

Another type of alternative education in data science attempts to meet the need for training that is more accessible than formal degree programs, and yet more social and somewhat less rigorous than MOOCs. These tend to be community-based, organized through a non-profit organization or without any formal organization at all. They consist of in-person meetings and training sessions such as workshops, typically with short time commitments. The goal of these groups is not necessarily to train data scientists, but to get the techniques and tools into the hands of a wider range of people, laying down the fundamentals and teaching people to be conversant.

Among the most formalized of these groups are Software Carpentry and its sister group, Data Carpentry. These groups focus specifically on teaching scientists the basics of computing and software development to aid in their research. They came out of the recognition that scientists—those with the highest level of domain expertise in these research endeavors—often acquire their

---

[7] https://www.codecademy.com/
[8] https://developers.google.com/edu/python/
[9] https://docs.python.org/2/tutorial/

software skills informally by self-study or through impromptu lessons from peers. This can lead to gaps in knowledge that make research far more inefficient, as the time it takes to do computational science is becoming more and more dependent on the time spent writing, testing, debugging, and using software (Hannay et al. 2009; Wilson 2014; Wilson et al. 2014).

The number of skills required of a data scientist is unrealistic. It can't be expected that all scientists working on computationally intensive research become experts in every aspect of data analytics. These research projects often require teams of people in order to cover all the necessary bases, such as statistics, software development, and domain expertise. However, having some broad training in all areas of data science techniques can go a long way in getting scientists to work more efficiently. By learning how to write and debug code, scientists can save time, communicate more readily with other members of the team, and understand more fully the methods underlying the research effort. The teamwork requirement of data science also underlies the importance for skills in version control systems (VCS). This involves the management of software, documents and data. There is a widespread movement around “reproducible science” within academia where VCS is a central focus (see chapter by Plale in this volume for more on reproducible science and reuse of data in research). This is influencing curricula in the classroom and topics on syllabi.

Software Carpentry is a nonprofit volunteer organization that runs short workshops to train scientists in basic computing skills related to programming, automation, and version control. It originated as a course taught to scientists at the Los Alamos National Laboratory in New Mexico, and developed into a network of two-day intensive workshops run all over the world. The workshops focus on teaching the basics of Unix shell commands, programming using Python and R, version control, and using databases and SQL. In 2013, they organized 91 workshops for around 4,300 scientists (Wilson 2014).

There exist several other efforts to offer informal data science training to researchers. Some of these are offered by those in charge of running supercomputer facilities at research institutions. The Princeton Institute for Computer Science and Engineering (PICSciE) and The SciNet supercomputer cluster at the University of Toronto, for instance, both offer workshops and training to researchers doing work that make use of these advanced resources. Other examples of informal training for scientists involve loose organizations such as The Hacker Within, originally started as a student group at the University of Wisconsin-Madison (and an inspiration for Software Carpentry), now a nascent association of groups offering meetings and informal bootcamps at University of

California-Berkeley, University of Wisconsin-Madison, and several others (Losh 2011).[10]

Other organizations offer data science training to different audiences. The Community Data Science Workshops (CDSW) are a series of volunteer-staffed interactive classes at the University of Washington over the course of three consecutive weekends that offer training in Python to any novices interested in using data science tools to study online communities such as Wikipedia and Twitter.[11] These types of efforts, being completely open and fueled by passion and interest, often spread—the CDSW was inspired by the Boston Python Workshop, which in turn inspired the Python Workshop for Beginners at the University of Waterloo. PyData is another community that runs conferences around the world for developers and users of Python tools; all of their net proceeds go to fund the NumFOCUS foundation, which supports and promotes open source computing tools in science.[12]

The landscape of informal data science training also includes an array of online and in-person communities that exist without any central structure or organization. The website meetup.com, which allows groups of people to organize and advertise in-person meetings, has hundreds of active groups around

[10] http://thehackerwithin.github.io/
[11] http://communitydata.cc/workshops/
[12] http://pydata.org/

the world that meet to discuss and work on issues and projects related to data science.[13]

Data Science for Science

The landscape of data science education in universities and college is one dominated by a connection to industry. This is a reflection of the enormous gains that businesses stand to make from this phenomenon. This force can also be seen in a siphoning of talent toward these businesses and away from academic science. This will be a challenge going forward for the field of data science, as the future of data science as an academic discipline depends on some of the thought leaders remaining within academia itself.

Whether learned in data science programs or in the practice of discipline-specific research, the skills many PhD graduates gain are highly valued in industry. These graduates face a job market skewed toward business. An NIH postdoctoral researcher can expect to earn a salary of around $42,000 per year (National Institutes of Health 2014). Entry level data scientists can make upwards of $100,000 per year (Dwoskin 2014). Many graduates of quantitative fields such as physics, math, and astronomy are moving into data science, where their programming and quantitative skills are highly valued. Often, they are motivated by more than just money. Many private companies offer positions that involve

[13] http://data-science.meetup.com/

working on interesting problems that are attractive to people oriented toward research. At companies like LinkedIn and Yelp, data scientists can apply statistical models and machine learning techniques in situations that will ultimately drive business, and often get the satisfaction of seeing the fruits of their work within months, as compared to the slower pace of academic research (Dwoskin 2014). Some companies also have research labs that fund scientists to do exploratory research not directly related to increasing revenue—although corporate incentives still tend to be present in these environments, and the results of research are not made available to the public or other researchers outside the company.[14] These companies are investing money in research at the same time that academic research funding is still feeling the squeeze from the recession of the late 2000s.

Exemplifying this trend is the Insight Data Science Fellows Program[15], a postdoctoral fellows program designed to help PhDs with quantitative skills make the transition from academia to industry. Fellows work on a data science project and are trained in industry standard techniques and tools, such as machine learning, version control, parallel computing, Python, and R. At the end, they are matched with industry jobs. The program boasts a 100% placement rate.

[14] Microsoft Research, the research division of Microsoft, is one notable exception to this. They tout an "open academic model" and prioritize collaboration with partners in academia, government, and industry (Microsoft 2013).

[15] http://insightdatascience.com/

Another challenge for data science in academic sciences is an incentive structure within the profession that prioritizes publications over other important work such as software development (the so-called "publish-or-perish" model). It has been estimated that, in the age of computational and data-driven research, scientists can spend 30% or more of their time developing software (Hannay et al. 2009). This being time that could be spent writing articles to publish in journals or conferences, there is a strong disincentive for scholars to spend even more time developing additional skills or writing clean and reproducible code that could serve other researchers in similar domains (Vanderplas 2013; Vanderplas 2014).

There are some efforts underway to combat the movement of talent away from scientific research. In 2013, the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation began a $38 million project to fund initiatives to use data science to advance research. The funding goes to support data science centers at three partner institutions: the Center for Data Science at New York University; the Berkeley Institute for Data Science at University of California, Berkeley; and the eScience Institute at the University of Washington. These institutes work within and across institutions to foster collaborations, develop sustainable and reusable tools for scientific research, and work toward fixing the outdated academic career structure (Gordon and Betty Moore Foundation 2013).

Similar initiatives exist in other institutions. The Stanford Data Science Initiative, the Columbia University Data Science Institute, the Massive Data

Institute at Georgetown's McCourt School of Public Policy—all seek to support the use of data science in research, and to encourage cross-disciplinary collaboration in order to do so.

While still in its infancy, there are some indications that data science is being established as an academic discipline. Data science PhD programs are beginning to appear in universities that want programs with more of a research focus. Brown University's Computer Science department, for example, offers a PhD in Big Data, and Carnegie Mellon offers a PhD in Machine Learning. Other schools—University of Washington and Penn State University—offer interdisciplinary PhD programs in Big Data or Analytics funded by the National Science Foundation's Integrative Graduate Education and Research Traineeship (IGERT). PhD programs exist to train thought leaders and innovators in research (although many will go into industry with their PhD). This first crop of PhD students in Data Science will soon be graduating.[16] The academic direction that data science takes will be influenced greatly by this set of young researchers. Will they get jobs in their home discipline or in other disciplines? Or will they go to industry? If they stay in academia, will this be enough to establish a standalone field of data science over the next 10 years or will it remain cross-disciplinary?

A Moving Target

[16] Students have been graduating with degrees in Machine Learning for some time.

The dizzying rate of change in the world of data science education gives a feeling of uncertainty, and calls into question the future of the field. Ideally, education trends should align approximately to movements in technology, but technology changes too fast for education to follow every new development or trendy whim. By chasing every new trend, educators risk diluting their brand and wasting curricula development on skills that become obsolete before the first day of class. However, for those trends in technology that have staying power, education institutions should respond, both for their own self-interest, and for the betterment of the economy and society.

It remains to be seen where data science fits into this balance—if it will turn out to be a flash in the pan, or if it will prove itself as a discipline in its own merit. Data science may diffuse to all domains and infiltrate syllabi across campus without the need for stand-alone programs. Either way, big unwieldy education bureaucracies will have a difficult time keeping up with the more flexible, adaptive education institutions when responding to trends in technology. Our findings don't indicate a dependence on the institution size, at least for data science. Big, well established universities are responding to this data revolution as fast as small institutions. We see a greater dependence at the departmental level than on overall institution size. The more interesting story developing is the impact that non-university programs (see section “Alternative Data Science Education”) are having on education and the role they will play with or without the traditional universities. The big data movement is a useful, real-time experiment for seeing which institutions and types of institutions will meet the demand (or lack of demand) most effectively.

The emergence of data science is also an opportunity to attract more students to STEM (Science, Technology, Engineering and Math) fields. The National Science Foundation has spent millions of grant dollars trying to figure out how to graduate more STEM students (National Science Foundation 2014). Data science can serve as a potent STEM attractor, given the high paying salaries (Dwoskin 2014), the industry demand (Manyika et al. 2011), the popularity of data scientists such as Nate Silver (Silver 2012), and the ostensible “sexiness” of the discipline (Davenport and Patil 2012).

Academia tends to move more slowly than industry, which can make it difficult to respond to trends appropriately. However, there are advantages to this slower pace. The data revolution has impacts well outside the consumer products with which industry is concerned. Potential benefits exist in the natural and social sciences, in government and nonprofit organizations, in health and medicine. The drawbacks of big data are important as well, such as privacy and ethical concerns that are far from resolved (see also, chapter by Cate in this volume). Academic institutions should fill the role of addressing these gaps through their curricula and activities. Some examples of this exist already. The Data Science for Social Good

Fellowship programs at the University of Chicago[17] and the University of Washington[18], as well as the Atlanta Data Science for Social Good Internship[19], mentor students to work on projects for nonprofits and other organizations that have a social mission. The field of bioinformatics is a related but separate discipline in its own right, using data science techniques to tackle issues in health and medicine. Ethics is already a component of many data science curricula. These efforts should be applauded and encouraged. Students, even those focused on a future in business, should be made to critically consider the societal impact of their work and the ethical implications that their algorithms and experiments could have (see also, chapter by DeDeo in this volume). This is where the slower pace of academia can be good.

Many of the education institutions and new programs in data science cite the McKinsey report. We would like to see more follow-ups on this kind of employment forecasting. Many decision makers are using this as their reference point. The forecasts reported may be correct; in fact, they may be underestimating the demand. However, given its influence in education board rooms, we recommend that governments, universities, and foundations sponsor more studies like this to verify and update these forecasts, particularly in fast-changing areas like data science. We would hate to see wasted effort and money at budget-

[17] http://dssg.io/
[18] http://escience.washington.edu/what-we-do/data-science-for-social-good
[19] http://dssg-atl.io/

strapped universities be passed on to students through exorbitant tuition fees with no commensurate jobs.

Conclusion

Data science is going through growing pains, and the education landscape reflects this. The enthusiasm behind big data has ignited fevered growth as institutions and organizations race to meet demand. At colleges, we see an explosion of new programs, primarily Master's degree and certificate programs with elements of business and management. Alternative forms of education have sprouted to address other sorts of demand, including online courses and community workshops. Given the demand for data scientists in business, there are incentives for leaders in this field to choose careers in industry rather than domain science research, which may slow the development of a data science discipline and its influence on non-business domains.

While the future of data science can be questioned, we see the current activity as an indication of things to come in both technology and education. The potential of big data is becoming more and more important in society, and educational institutions and organizations are beginning to form an infrastructure which can train students with the expertise to harness it. This infrastructure will be critical in sustaining the big data “gold rush” we have seen in recent years.

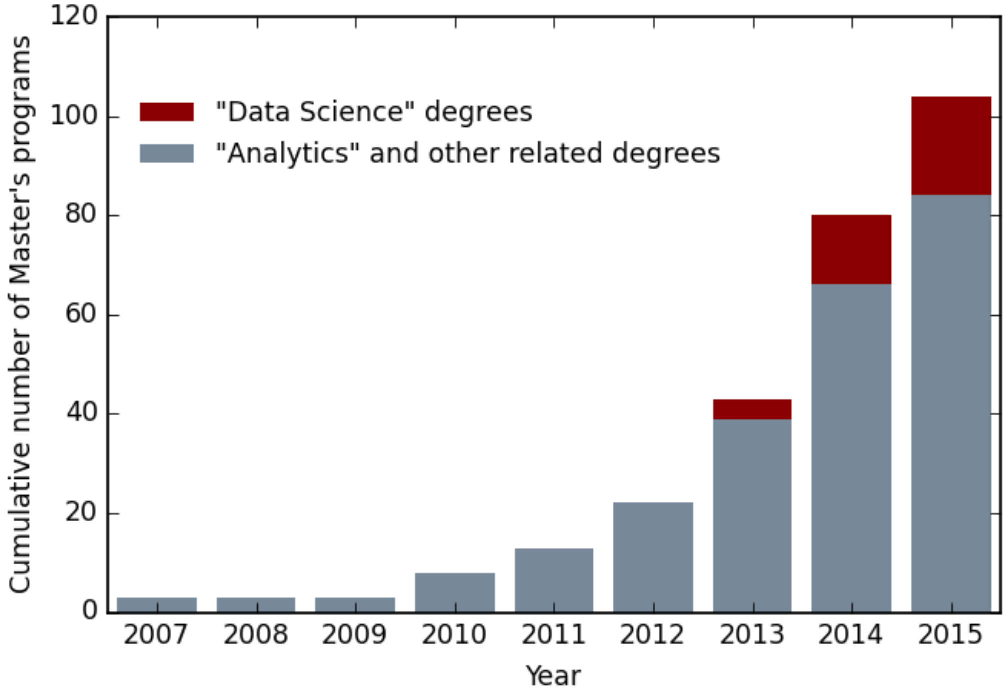

Figure 1: Growth of Data Science Master's Programs

The number of Master's degree programs in data science and analytics has risen dramatically since 2011. Data from http://analytics.ncsu.edu/?page_id=4184

Note: Defining which Master's programs belong in the category of ``data science'' is not straightforward. For example, the University of Washington offers a M.S. in Information Management with a specialization in Data Science & Analytics that has many of the elements of a Master's in Data Science; this program is not included in the data. Because of this, the growth of data science degrees reported in this figure likely far underestimates the actual growth.

Table 1: Most expensive data science programs

College	Degree	Duration	Cost (r=resident)
Univ Denver	M.S. in Business Analytics	12-36 Mo	$69,500
New York Univ	M.S. in Business Analytics	12 Mo	$67,500
Carnegie Mellon Univ	M.I.S.M. (Business Intel. track)	16 Mo	$67,200
Southern Methodist Univ	M.S. in Applied Stats and Data Analytics	18-24 Mo	$65,600
Carnegie Mellon Univ	M.S. in Computational Data Science	16 Mo	$65,000
Northwestern Univ	M.S. in Analytics	15 Mo	$64,800
Univ Rochester	M.S. in Business Analytics	10 Mo	$62,700
Univ CA, Berkeley	Master of Information and Data Science (MIDS)	12-20 Mo	$60,000
IL Inst of Technology	M.S. in Marketing Analytics and Commun.	12-24 Mo	$58,000
Univ Miami	M.S. in Business Analytics	10 Mo	$57,100

Table 2: Least expensive data science programs

College	Degree	Duration	Cost (r=resident)
Univ Connecticut	M.S. in Bus. Analytics and Project Mgmt	12+ Mo	$24,750
Fairfield Univ	M.S. in Business Analytics	12+ Mo	$24,500
Xavier Univ	M.S. in Customer Analytics	24 Mo	$24,000
Univ Alabama	M.S. in Applied Stats (Data Mining track)	9+ Mo	$24,000 ($9,1500 r)
Elmhurst College	M.S. in Data Science	24 Mo	$23,500
Southern NH Univ	M.S. in Data Analytics	20+ Mo	$22,600
Valparaiso Univ	M.S. in Analytics and Modeling	18 Mo	$21,400
Univ Iowa	M.S. In Business Analytics	12-36 Mo	$20,000
South Dakota State Univ	M.S. in Data Science	12+ Mo	$17,600 ($9,900 r)
Dakota State Univ	M.S. in Analytics	20+ Mo	$12,800 ($6,100 r)

Cited references

Davenport, Thomas H. and D.J. Patil. "Data Scientist: The Sexiest Job of the 21st Century." Harvard Business Review (October 2012). Available at https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

Dwoskin, Elizabeth. "Big Data's High-Priests of Algorithms." Wall Street Journal (August 2014). Available at http://www.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088.

Finzer, "The Data Science Education Dilemma." Technology Innovations in Statistics Education 7.2 (2013).

Gellman, Lindsay. “Big Data Gets Master Treatment at B-Schools.” Wall Street Journal (November 2014). Available at http://www.wsj.com/articles/big-data-gets-master-treatment-at-b-schools-1415226291.

Gordon and Betty Moore Foundation. “Bold New Partnership Launches to Harness Potential of Data Scientists and Big Data.” (2013). Available at http://www.moore.org/newsroom/press-releases/2013/11/12/%20bold_new_partnership_launches_to_harness_potential_of_data_scientists_and_big_data.

Hannay, Jo Erskine, Carolyn MacLeod, Janice Singer, Hans Petter Langtangen, Dietmar Pfahl, and Greg Wilson. "How do scientists develop and use scientific software?." In Proceedings of the 2009 ICSE workshop on Software Engineering for Computational Science and Engineering, pp. 1-8. IEEE Computer Society, 2009.

Liyanagunawardena, Tharindu Rekha, Andrew Alexandar Adams, and Shirley Ann Williams. "MOOCs: A systematic study of the published literature 2008-2012." The International Review of Research in Open and Distributed Learning 14, no. 3 (2013): 202-227.

Losh, Gillian. “The Hacker Within Focuses on Scientific Computing.” University of Wisconsin-Madison News (April 2011). Available at http://www.news.wisc.edu/19236.

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, and McKinsey Global Institute. "Big data: The next frontier for innovation, competition, and productivity." (2011).

Microsoft. Microsoft Research at a Glance - Microsoft Research. (February 2013) Available at http://research.microsoft.com/en-us/press/overview.aspx

Miller, Claire Cain. "Universities Offer Courses in a Hot New Field: Data Science." The New York Times (April 2013). Available at http://www.nytimes.com/2013/04/14/education/edlife/universities-offer-courses-in-a-hot-new-field-data-science.html

National Institutes of Health. "NOT-OD-14-046: Ruth L. Kirschstein National Research Service Award (NRSA) Stipends, Tuition/Fees and Other Budgetary Levels Effective for Fiscal Year 2014." (February 2014). Available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-046.html

National Science Foundation. "Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering." nsf1554 (2015). Available at http://www.nsf.gov/pubs/2015/nsf15544/nsf15544.htm.

National Science Foundation. "Investing in Science, Engineering, and Education for the Nation's Future – National Science Foundation Strategic Plan for 2014-2018. (2014). Available at http://www.nsf.gov/pubs/2014/nsf14043/nsf14043.pdf.

Press, Gil. “Top 10 Most Funded Big Data Startups (Updated).” Forbes (July 2014). Available at http://www.forbes.com/sites/gilpress/2013/10/30/top-10-most-funded-big-data-startups-updated/

Silver, Nate. *The Signal and the Noise : Why so Many Predictions Fail--but Some Don’t*. New York: Penguin Press. (2012).

Vanderplas, Jake. “The Big Data Brain Drain: Why Science Is in Trouble.” Pythonic Perambulations (2013). Available at https://jakevdp.github.io/blog/2013/10/26/big-data-brain-drain/

Vanderplas, Jake. Hacking Academia: Data Science and the University. Pythonic Perambulations (2014). Available at https://jakevdp.github.io/blog/2014/08/22/hacking-academia/

Wilson, Greg, D. A. Aruliah, C. Titus Brown, Neil P. Chue Hong, Matt Davis, Richard T. Guy, Steven HD Haddock et al. "Best practices for scientific computing." PLoS biology 12, no. 1 (2014): e1001745.

Wilson, Greg. "Software carpentry: lessons learned." F1000Research 3 (2014).