

## Designing Undergraduate Data Science Curricula: A Computer Science Perspective

**Dr. Predrag T. Tasic, University of Idaho**

Predrag Tasic is an early mid-career researcher with a unique mix of academic research, industrial and DOE lab R&D experiences. His research interests include AI, data science, machine learning, intelligent agents and multi-agent systems, cyber-physical/cyber-secure systems, distributed coordination and control, large-scale complex networks, internet-of-things/agents, and mathematical and computational models and algorithms for "smart" transportation, energy and other grids. He is interested in applying data analytics, machine learning, intelligent agent and AI techniques to emerging problems related to large-scale decentralized cyber-physical systems, critical infrastructures and "smart grids", autonomous vehicles, as well as energy, health care and other domains of major economic and societal impact.

Dr. Tasic holds a doctorate in Computer Science from the University of Illinois at Urbana-Champaign. His doctoral dissertation (2006) was on Distributed AI and large-scale Multi-Agent Systems. Most recently, at Washington State University (2015 - 2017), Dr. Tasic worked on dynamics of large-scale networks, graph pattern mining, Boolean Network models of cyber-physical systems, Internet-of-Agents, as well as AI, data analytics and knowledge engineering applied to problems in health care. While at the University of Houston (2009 – 2012), he did research in machine learning, multi-agent distributed computing and control, data mining and distributed database systems, emerging behavior in complex networks, "smart energy" and computational game theory. During his graduate studies and combined five years of non-tenure-track academic research, he has authored over 70 peer-reviewed publications. He has a versatile R&D experience spanning three different high-tech industries, with both big companies (Cisco Systems and Microsoft) and high-tech startups, as well as with a leading government research lab (Los Alamos National Laboratory). He holds three USPTO patents (IP of Cisco Systems).

In addition to a doctorate in Computer Science, Predrag Tasic holds three master's degrees, two in mathematical sciences and one in CS. Tasic has a considerable teaching and student research mentoring experience. He has enjoyed working with students of a broad variety of ethnic, cultural and socio-economic backgrounds and at different types of academic institutions. He has been actively involved with IEEE – the Palouse Section and is currently President of the Section's Computer Society. He is also an active member of ACM, ASEE and AMS.

**Dr. Julie Beeston, University of Idaho**

Dr. Julie Beeston has both a Master's degree (from Carleton University) and a PhD (from the University of Victoria) in Computer Science, and she has developed and taught over a dozen courses at the university level. Beyond her teaching experience, she also has over a decade of industry experience as a software developer.

In industry, she has a history of solving 'unsolvable' problems. She enjoys a great deal of personal satisfaction when her analytical and problem solving skills can be applied to solve complex technical problems and when she can find creative new ways to pass the things she has learned on to the next generation.

Her first teaching experience was at Ozanam Sheltered Workshop teaching adults with mental and physical disabilities. The experience gave her the opportunity to try unique teaching methods and taught her how to tailor her teaching style to a specific person's needs. That experience taught her that given enough time any student can master any concept. There is no limiting factor on an enthusiastic student's ability to learn.

Her primary mission in teaching is to get the students enthusiastic about the subject. She does this by giving real-world examples of how the subject matter she is currently teaching has helped her resolve complex problems in industry.

# Designing Undergraduate Data Science Curricula: A Computer Science Perspective

Predrag Tošić  
University of Idaho - Coeur d'Alene

## Abstract

We discuss opportunities and challenges encountered in developing new undergraduate degree programs that are inherently cross-disciplinary and require institutional and instructional support from different departments and colleges. We have recently been involved in early stages of curriculum development for an undergraduate BA/BS Data Analytics program at Washington State University (WSU), involving faculty and resources from Computer Science, Mathematics & Statistics, College of Business, and other units at the University. The development of such a new degree program required developing entirely new courses and their syllabi, identifying faculty across the existing academic units suitable and available to teach those new courses, as well as identifying key existing courses across Computer Science, Mathematics, Statistics and Business that should be incorporated into the new degree program. One challenge there was to ensure providing the solid foundations and a right balance to the future Data Analytics graduates, without making their BA/BS degree program require substantially more credit hours than the existing undergraduate programs in e.g. Statistics, Mathematics or the Data Science "track" in Computer Science.

We share some insights from this curriculum development process. Some issues that we have encountered include the following: what are the most important skills that the future BA or BS graduates in Data Analytics should acquire from computational and computer science standpoints? Is requiring those students to take the existing CS courses on Data Mining, Machine Learning or Database Systems appropriate, or should new courses (not shared with the existing CS Data Science or similar "tracks") be developed? What is the right balance between training in computational (and mathematical) foundations, vs. various application domains, such as in business or life sciences? One of the key early lessons we have learned: effective, respectful and cooperative communication across the traditional disciplinary boundaries is absolutely crucial for intrinsically cross-disciplinary higher-education undertakings such as developing a new data analytics degree program to be successful.

## Keywords:

*Industry Demand, Data Science Talent Gap, Undergraduate Curriculum Development, Data Science, Computer Science Curricula, Interdisciplinary Programs*

## 1 Introduction

An exponential growth in information and computing technologies' capabilities in terms of storing, retrieving and processing large amounts of various types of data has led to an increasing demand for experts skilled at working with data [1, 2]. Skilled Data Scientists and Data Analysts

are in high demand, from Silicon Valley to Wall Street to large retail chains to health care organizations and across many other industries and sectors. In particular, it has been observed that the demand for data science/analytics talent, unlike say 10-15 years ago, is no longer limited to high-tech giants [2]. Predictions for the near future are that this demand for data-skilled talent will continue to grow, whereas the supply (from colleges and universities training the next generation of data scientists and analysts) will not be able to keep up with that pace (for one such study see, e.g., [3]).

To fill in that gap, graduate programs at several leading research universities have created new tracks within the existing graduate programs, or in a few cases new degree programs entirely, focusing on data science and analytics [4]. These programs are led and taught by the existing faculty ranging from Mathematics & Statistics to Computer Science to Business Schools. However, while training MS and PhD level Data Scientists is important, by itself it is unlikely to suffice for bridging the aforementioned talent gap for a number of reasons, starting with the “economics of scale”: individuals with advanced degrees would fill-in some fraction of the intermediate to advanced/senior data science related positions in industry, but the unmet demand with respect to junior and entry level positions would most likely remain. That gap is partially filled by various online (degree-granting as well as non-degree/certificate-based) programs, whose quality (from this author’s own experience as well as the experiences from professional interaction with data scientists whose main or in some cases sole qualifications and training stem from such programs) tend to vary very broadly. The administrations at some of the nation’s leading research universities have decided to address this growing demand-vs-supply of talent gap in data-intensive fields, especially at the entry-/junior-level data analytics related positions, by initiating data science undergraduate programs. A prominent example of a leading university creating and successfully scaling (so far) such an undergraduate program is the Ohio State University [5]. Subsequently, a number of academic institutions have attempted creating and growing their own data science and/or analytics undergraduate degree programs, often based on the experiences and curricula at Ohio State.

According to [www.bachelorsportal.com](http://www.bachelorsportal.com), there are currently 95 data science / data analytics undergraduate programs offered by the US colleges and universities. Liberal arts / 4-year colleges and research universities alike have been developing their data science/analytics curricula in recent years. One of the reasons that the upper administration of Washington State University wanted to emulate the Ohio State’s program, were the large, public, land-granted, research-centric nature of that institution, characteristics shared with Washington State University. (Another reason was likely that Ohio State University was among the first such academic institutions to develop an undergraduate, as opposed to MS-level or certificate-based, data science curriculum and degree program; we would like to make a disclaimer, however, that these statements on likely reasons behind modeling WSU’s data analytics undergraduate

program on that at Ohio State University are our educated guesses, not a factual knowledge or any kind of official position of the Washington State University's administration.)

In this paper, we summarize our own recent experience with data science and analytics, and in particular some challenges and opportunities of training the next generation of data scientists. We share our insights from three different, mutually complementary vantage points: as someone who worked as a (PhD-level, Senior/Principal) Data Scientist in a couple of different roles and application domains in industry in recent years (2012 – 2015), as a researcher and educator who has taught (advanced undergraduate and preparatory level graduate) special topic courses focusing on certain aspects of data science and analytics, and as someone who was briefly involved in cross-disciplinary curriculum development for a new (undergraduate) Data Science BS/BA program at a leading public research university in the U.S. Pacific Northwest.

The primary purpose of this paper is to share some ideas and insights about data science/analytics training of (especially undergraduate) college students to meet the demands of the regional, national and global economies cf. in terms of supplying the data analytics talent. We hope our insights are of value, but we do not claim to have any definitive answers on what should the new degree programs on data science and/or analytics be like, or indeed what are the most effective ways of training the next generation of data scientists and analysis.

## **2 Data Science: The Demand Keeps Growing**

Companies and other organizations, large and small, are collecting and processing ever-increasing amounts of data from a variety of sources. What used to be a “specialty” of the high-tech giants and government agencies such as the NSA or DoD, is now an opportunity – and a challenge – in almost all areas of economy and society, from health care and education to many small businesses to financial sector to energy industry and many, many more. However, training human experts to effectively work with and analyze these massive amounts of various types of data has not been keeping the pace with the information technologies' capabilities of storing, retrieving, and “raw processing” of that data. Moreover, this gap continues to grow, necessitating esp. the institutions of higher learning to dedicate money, faculty and other resources to creating new programs (or, at the very least, new tracks within the established undergraduate and graduate programs) to address this mismatch between the demand for and the supply of data scientists and analysts.

Three main reasons have been identified, why (being a) Data Scientist has been in recent years, and will likely remain being for some years to come, the top or one of the very top careers in the US, across all most important dimensions for one's career choice: the demand (as measured cf. by the number of advertised job openings), the average pay, and the average job satisfaction by those individuals who already are pursuing their careers in a given field. These three reasons for

the top ranking of data science as a career choice are: i) the overall shortage of data analytics talent, ii) the ever-growing demands of “big data” on companies and other organizations, and iii) the aforementioned fact that facing the challenges of, and working with, the “big data” is no longer limited to a handful of tech giants (and government agencies) [2]. While most expert predictions and analyses project an-ever growing gap between *all* computer science and information technology related industry needs and job openings on one hand, and the produced number of college graduates with prerequisite skills and degrees on the other, such projections for data scientists and analysts are particularly telling; for instance, one analysis by the IBM suggests a 28% growth in data science demand within a three-year period 2017 – 2020 [3].

The increase of raw numbers insofar as the demand for data analytics talent is also matched by an ever-increasing variability of what it really means, to be a data scientist or analyst. In particular, specialized skill-sets related to mining and analyzing e.g. image and video data, and other types of data formats that did not exist either at all or certainly not in very large volumes a mere decade ago, will also likely grow in demand. With very rapid changes in technologies and tools, striking the right balance between solid general foundations resulting in broadly applicable, transferable skills on one hand, and focusing on mastering particular tools, methodologies and technologies on the other, is challenge to both data science practitioners and the educators of the next-generation data scientists.

So, given all these general trends that are showing no signs of slowing down, an important question arises for the educators and administrations of institution of higher learning: how should colleges and universities adjust to these new realities, and swiftly yet wisely and adequately start addressing this shortage of data-skilled talent for which the industry demand keeps growing? We provide some concrete suggestions as a partial answer to that question in the rest of this paper.

### **3 Current Industry Trends and Demands: A Real-World Data Scientist’s Skill-Set(s)**

Once the talent gap in the area of data science & analytics has been identified, and many expert predictions about the near-to-intermediate future about the continued growth in demand for the related skills, and once the institutions of higher education in the US (as well as many globally) have taken a notice, the next question is: what are the actual, concrete key skills and knowledge that the next-generation data scientists and analysts should have, in order to fill that gap most effectively? This, too, turns out to be a multi-faceted, complex question.

We will briefly summarize below our own experience, stemming cf. from i) working as a Senior/Principal Data Scientist in industry (cf. high-tech/Silicon Valley) and ii) having many conversations, formal and informal interviews, and other interactions with business leaders, companies’ CEOs and CTOs, other Data Scientists working in the private sector, as well as many high-tech, financial sector and other recruiters (those from individual companies’ HRs as well as external “head hunters”) searching for the data analytics talent. Some disclaimers are in order

here. Thoughts that follow are personal observations and insights based on personal experience (from working in and interacting with industry), and need not reflect an official position or prevalent attitudes towards how to best educate and train the next generation of Data Scientists either of the author's most recent prior employer, Washington State University (WSU), where he was involved in the undergraduate Data Analytics curriculum development, 2015 – 2016; or the author's primary academic employer at the time of the writing of this paper, University of Idaho (while the author remains an Adjunct Faculty with WSU). As of the summer 2018, University of Idaho does not have an undergraduate degree program in Data Analytics or Data Science. More about WSU's BS program in Data Analytics can be found at website [6].

When an advertised job opening at a popular careers site such as Indeed.com or LinkedIn, or in a professional publication such as IEEE or ACM newsletters or other communications, is seeking a talented or experienced *Data Scientist*, what does that actually mean in terms of the desired technical (and potentially other, such as business) skills? It turns out, it can mean a range of different things; here are some of the most common examples:

- A traditional or not-so-traditional Business Analyst;
- An expert on database technologies, esp. the most common relational database systems (DBMS), usually combined with some distributed data technologies and techniques, such as Hadoop and MapReduce;
- A research-level expert in Machine Learning (ML) and Data/Pattern Mining;
- A good Software Engineer with some expertise and experience in using ML and Data Mining tools and products, but not necessarily someone with research skills in ML;
- An Applied Statistician, usually with strong background in statistical inferencing, regression techniques and tools, some ability to prototype in and software packages (e.g., the scripting language R) and ideally some hands-on experience with the “big data”;
- An expert in working with particular types of data (e.g., images or video or time-series data), who is also comfortable with at least some of the common software tools and packages, and ideally has one or more types of domain expertise particular to the type(s) of data a given organization mostly works with (for example, certain types of image data, such as that arising in various medical sub-disciplines; or typical time-series data commonly arising in the energy sector or the financial sector, etc.);
- A well-rounded “generalist” without research-level or advanced degree-level expertise in any particular data-related area, but who “knows just enough about” inferential statistics, data visualization, most common machine learning tools, and most common DBMS/“big data” technologies – and (ideally) either already knows how to script in e.g. R or Python and how to extract the data out of a relational DBMS, or at least appears to be willing to learn how to acquire the necessary hands-on skills for a particular position and organization fairly quickly, in order to provide a good “return on investment” (which is very important in business practice, especially in an environment where good Data Scientists, just like good Software Engineers, tend to change jobs every 12-18 months!).

In fact, it is our own experience of a number of time-wasting phone interviews/pre-screens, where we would express interest in a “Data Scientist” position and it would turn out, in the first 5 or 10 minutes of a phone interview, that what a particular organization looking for a “data scientist” really needs is, for example, someone to maintain and run the operations of their Hadoop system, or a full-stack seasoned software developer with some idea about most common ML tools, and not a research-level AI and ML expert with decent scripting skills, but who is neither looking for a “Data Ops” role nor a hard-core “Software Dev” role.

Given this broad variety (and frequent misunderstandings in practice, given just how heavily the term “data science” has been overloaded in recent years), as well as the fact that many especially “big data” technologies will likely become obsolete by the time entering “data analytics” college freshmen actually get their BS or BA degree and seek to join the workforce, what should then colleges and universities do, to provide adequate training to the future entry-level and junior data scientists and analysts? What is the right balance between (mathematical and computational) foundations on one hand, and concrete tools and technologies (many of which, important and pervasive as they may be today, may well end up obsolete in 3-4 years), on the other?

Based on our own experience both from working in industry and from having taught data science related computer science and (applied) mathematics courses, we share some thoughts on which paradigms and skills we would recommend for the undergraduate programs in data analytics; we note, that the particular programming or scripting languages and other specific technologies may well change in a few years – but our recommendations are based on the relevant industry trends and practices over the past ~10 years, as well as the experts’ projections for the next 5-10 years.

- Introductory probability theory, with an emphasis on discrete probabilities;
- Introductory statistics, with an emphasis on inferential statistics (but certainly thoroughly covering foundations of descriptive statistics, as well)
- Discrete mathematics, including probabilities, combinatorics, and graphs;
- (Up to) multi-variate calculus, with some emphasis on visualization;
- Intermediate statistics, emphasizing regression techniques and data visualization, and using an appropriate scripting language such as R;
- Introductory to intermediate programming in a “digestible” yet broadly used language (with plenty of open source libraries and big community of users) such as Python;
- Fundamental data structures and algorithms (incl. some hands-on implementation, in a language such as Python);
- Introductory database systems, with an emphasis on relational databases, SQL and an overview of main “big data” and distributed data paradigms, such as MapReduce;
- Introductory to intermediate Machine Learning, focusing on core mathematical and conceptual principles, and computational techniques, across the sub-areas of Supervised Learning, Unsupervised Learning and (at least the basics of) Reinforcement Learning;

- Data/Pattern Mining, possibly fine-tuned to a general application domain (such as, for example, Life Sciences vs. Business vs. Energy Industry);
- An early exposure to Data Visualization, incl. learning how to use some of the existing simple tools, but also emphasizing the business value and transferable skills;
- An early exposure to the “business side”, including how to present and use insights from data to “tell a story”, be able to convince non-technical stakeholders and business leaders;
- Related to the “business side” and “story telling”, an early exposure to effective public speaking, both technical and especially non-technical/to general audiences.

#### **4 Developing New Data Science Undergraduate Programs**

Following the early-adopter examples such as [5], as well as the political push (and some economic incentives, in an era of an overall rapid decline of how big a fraction of the overall University’s budget actually comes from the state) by its State Legislature, the land-granted research university, Washington State University in the Pacific Northwest, has relatively recently pursued creating its own undergraduate Data Analytics program. Key academic units (and their respective faculty) involved in the development of this new program were coming from Department of Mathematics & Statistics, (the Computer Science side of) School of EECS, College of Business, and Life Sciences. This new program was designated to be directly under the Provost’s Office, meaning no individual college (specifically, College of Arts & Sciences, College of Engineering or the Business School) or the respective Deans or Department Chairs would be making the final calls on key decisions; this unique setup was due at least in part, to avoid in-fighting between the Departments and Colleges involved, over how should financial resources and faculty time / teaching allocation be done to support the new program.

Insofar as the impact of this new program on the existing academic programs and units, the general sentiment was that the new program would fill an important niche and satisfy the strategic objectives of the State legislature, without negative impact on any of the existing undergraduate degree programs. The two closest existing programs to the new Data Science/Analytics program are the undergraduate BA/BS program in Statistics (offered by the Department of Mathematics & Statistics) and the “data science track” within the undergraduate Computer Science major (offered by the School of Electrical Engineering & Computer Science). In both instances, the curriculum overlaps between the new program and the existing ones are actually fairly modest (and readily justifiable); importantly, the core objectives of the new program are rather different from those of either the existing Statistics BS/BA program, or the existing Data Science track within the Computer Science program. For these reasons, there was little concern about potential negative impact on any of the existing programs; however, several faculty shared concerns about (primarily faculty/instructional) resources for the new program. Therefore, those in charge of the curriculum development for the new program were advised, to maximize utilization of the existing courses and related resources already provided by



Department of Mathematics and Statistics, School of EECS, College of Business and other academic units participating in the new data analytics program.

In that context of maximizing the existing course re-use (within what is reasonable, given the Data Analytics program's and its students needs), several discussions took place on just how much statistics, as well as how much computer science, the students in the newly established major should get exposed to. On the computational side, for example, while Computer Science programs with data science/analytics tracks increasingly emphasize Python for data analytics purposes (in addition to training students in "standard" broadly adopted by industry programming languages such as C++ or Java), it was agreed that students in the new Data Analytics program are not supposed to be trained as computer scientists, and in particular, that Data Analytics freshmen and sophomore should get their core "hands-on" training in working with data in (the statistical scripting language) R, rather than Python or any other "more computer science-ish" programming language. In particular, the introductory 100-level course developed specifically for this new program (in whose syllabus development this author was involved during 2015-2016), giving the incoming freshmen their very first exposure to data science and analytics, introduces and uses the R language, focusing on its simple interactive data manipulation and data visualization aspects. (That is, those freshmen are basically not assumed to have any kind of programming background whatsoever. The decision on the choice of those students' first "data analytics language" involved faculty from different disciplines, including Business, Statistics and Computer Science; the decision to adopt R for the aforementioned purposes was made unanimously.)

The initial curriculum development for this new Data Analytics program had two key components:

- Identifying and incorporating into the new Data Analytics BS/BA program the most relevant existing courses from Mathematics & Statistics, Computer Science, College of Business, and "application areas" (such as Life Sciences); and
- Developing the necessary minimum of the novel "core" courses specific to the new program and data science/analytics as a new major, starting with a 100-level introductory course whose purpose was to get incoming freshmen excited about data analytics and the many career opportunities awaiting once those students complete the program.

The curriculum committee made sure that there was, in each year of the program, that is, at each of the four levels, from the 100-level non-technical popular intro to data science, to the 400-level "capstone (senior design) project", at least one novel, unique to the data analytics program course. The senior capstone project was designed to address serious practical data-intensive problem in a given application domain or industry, and to involve both academic faculty mentors, and mentors (and, whenever possible, sponsors) from the industry. Just like with existing Computer Science or Electrical Engineering senior design projects, this Data Analytics capstone project was designed with the purpose of the students putting their previously acquired

knowledge and skills to work in a very realistic setting, demonstrating their ability to successfully join the workforce upon completion of their 4-year Data Analytics degree from Washington State University.

Important practical constraints on this curriculum development were, that the total number of credit-hours should be comparable (or at least, not exceed by much) the standard 4-degree undergraduate programs at Washington State University (in the ballpark of 120 credit hours, total), and that there should be an approximate balance between the core mathematical and statistical coursework, the core and applied/hands-on computational coursework, and the data analytics “application area(s)” (such as business or life sciences). A big part of the challenge was, to keep in mind that these students aren’t Math majors – yet they need to acquire a considerable amount of mathematical foundations; and likewise, these students aren’t Computer Science majors either, yet need to acquire a considerable amount of both foundational and practical computing skills. These constraints naturally led to various compromises, including “cutting down” on several junior- and senior-level courses in both Mathematics & Statistics and Computer Science, for which (naturally!) many faculty from those respective Departments felt should be a part of the new Data Analytics curriculum.

Importantly, it was also realized early, that the future data scientists and analysts need to be well-versed in i) public speaking, ii) effective cross-disciplinary and cross-cultural communication, and iii) business ethics. While the initial committee of which we were a part of did not finalize these aspects of the Data Analytics curriculum, by the time the program had final taken off, appropriate courses (“borrowed”, for example, from the Philosophy Department or the Washington State University’s internationally renowned College of Communication) that cover these important “non-technical” skills were incorporated into the core requirements of the new BS/BA program – while still keeping the overall credit count manageable (and, in particular, comparable to other 4-year undergraduate majors). More details on the current state of this curriculum can be found at [6]. The early stages of the program’s development are nicely summarized in the blog [7].

## **5 Conclusions**

We discuss the need for and development of new academic programs designed to meet the demands of industry and educate the workforce for the 21<sup>st</sup> century’s economy – specifically, in our case, the Data Science/Analytics undergraduate programs. Development and successful implementation and growth of such programs involve close collaboration and resource- and expertise-sharing among many existing academic units and programs, including Computer Science, Mathematics, Statistics and Business.

We share our insights from the standpoints of both personally observed industry trends and needs (admittedly, with an unavoidable bias of someone who has worked as Data Scientist mainly for

high-tech, software-centric companies), and of a Computer Science educator who, in particular, has been involved in both development and delivery of Data Analytics related coursework, as well as (briefly) the curriculum development for a new *Data Analytics* undergraduate program, at a prominent public research university. In order to remain relevant (and capable of long-term financial survival!), traditional colleges and universities need to find ways of creatively and pro-actively meeting the skill and workforce demands of the 21<sup>st</sup> century global economy.

Cross-disciplinary programs such as those on Data Science/Analytics are among the best examples of how to adjust to the new realities and ensure that the current and future generations of college students are provided with the right kinds of education, skillsets and opportunities. Getting involved in and helping growth and success of such cross-disciplinary programs, rather than feared, should be embraced by the Computer Science (and other) educators and upper administrators alike. From our own experience, active involvement in both educational and research aspects of such new academic programs, while challenging, also tends to be quite professionally rewarding, as well as very much appreciated by the students.

## References

- 1 Leopold, G. "Demands, Salaries Grow for Data Scientists", Datanami, Jan. 2017  
<https://www.datanami.com/2017/01/24/demand-salaries-grow-data-scientists/>
- 2 Zhang, V. and Neimeth. C. "Changing Data Science", InfoWorld, April 2017  
<https://www.infoworld.com/article/3190008/big-data/3-reasons-why-data-scientist-remains-the-top-job-in-america.html>
- 3 Columbus, L. " IBM Predicts Demand For Data Scientists Will Soar 28% By 2020", Forbes, May 13, 2017 <https://www.forbes.com/sites/louiscolombus/2017/05/13/ibm-predicts-demand-for-data-scientists-will-soar-28-by-2020/#6fca300b7e3b>
- 4 Leopold, G. "Grad Schools Move to Fill Data Skills Gap", Datanami, April 2015
- 5 More about the Data Analytics undergraduate program at the Ohio State University:  
<https://data-analytics.osu.edu/>
- 6 More about the current state of the WSU Data Analytics B.S. program can be found at the program's website, <https://data-analytics.wsu.edu/>
- 7 Early stages of the Data Analytics B.S. program development at WSU (in which this author was involved as a Computer Science representative) can be found in this blog from 2016, written by the then-Director of the program, prof. Dave Brown:  
<https://news.wsu.edu/2016/11/22/new-degree-in-data-analytics/>