Using the ASA Curriculum Guidelines to Develop New Undergraduate Programs in Statistical Science

Stacey Hancock
University of California, Irvine, Department of Statistics, 2204 Donald Bren Hall, Irvine, CA 92697

Abstract
Statistics undergraduate programs are on the rise, and the number of students majoring in statistics has dramatically increased in recent years. The American Statistical Association (ASA) has been proactive in identifying areas in the undergraduate curriculum that need to be enriched in order to keep up with our constantly changing discipline. This paper will discuss how the newly adopted ASA undergraduate curriculum guidelines can be used to develop new undergraduate programs in statistics or data science. As a case study, we will demonstrate how the guidelines were used in developing a data science major at the University of California, Irvine, and will consider the differences and similarities between statistics and data science undergraduate majors.

Key Words: Undergraduate curriculum, statistics education, undergraduate guidelines, data science

1. Introduction

The number of bachelor's degrees in statistics, biostatistics, and related fields has grown tremendously over the past decade. From 2003 to 2013, the number of bachelor's degrees in statistics has increased by 219% (from 526 to 1678) (Pierson, 2014). Not only has there been tremendous growth in the popularity of our discipline, the discipline itself has been adapting to new data technologies that are constantly changing.

The emergence of data science as its own field has brought new challenges to the field of statistics, both as we attempt to define the relationship between statistics and data science, and to adapt our programs to incorporate data science into the curriculum. Currently, there are at least ten colleges and universities in the United States with an undergraduate major in "data science", and that number grows substantially if we count majors in "data analytics", "business analytics", "business intelligence", "big data analytics", and other analogous names. Many of these undergraduate programs are offered through business departments and some through computer science departments, but few are housed in a statistics department.

This paper discusses how the newly revised American Statistical Association (ASA) 2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science can be harnessed to develop new programs in statistics or data science. Section 2 begins by reviewing the undergraduate curriculum guidelines. We will discuss how to use these guidelines in developing new undergraduate programs in statistics or data science in Section 3. Section 4 outlines the development of the data science major at the University of California, Irvine, and Section 5 concludes with future challenges.

2. American Statistical Association Undergraduate Curriculum Guidelines

In 2013, the American Statistical Association (ASA) convened a working group to revise and build on the 2000 Curriculum Guidelines for Undergraduate Programs in Statistical Science. The working group published the new guidelines in 2014, which are “intended to provide an overview of a principled approach to ensure that undergraduate statistics majors have the appropriate skills and ability to tackle complex and important data-focused problems” (ASA, 2014). The key points expressed in the Executive Summary of the revised guidelines are:

- Increased importance of data science.
- Real applications.
- More diverse models and approaches.
- Ability to communicate.

A principal theme of the new guidelines is to provide our graduates with the ability to “‘think with data’ and to pose and answer statistical questions” (ASA, 2014).

The increased importance of data science is the most notable change to the previous guidelines, and at the heart of data science is extensive computing experience. Graduates in statistics need to be highly qualified for “data scientist” and equivalent positions in the workforce. These positions require experience with database systems, higher-level programming languages, and algorithmic problem solving.

The curriculum guidelines outline five skill areas for statistics majors: statistical methods and theory, data manipulation and computation, mathematical foundations, statistical practice, and discipline-specific knowledge. Statistical methods and theory include statistical theory, exploratory data analysis, design of studies, and statistical models. Data manipulation and computation include the use of multiple professional statistical software tools, basic programming concepts, and computationally intensive statistical methods. Mathematical foundations include calculus, linear algebra, probability, and an emphasis on connections between mathematical and statistical concepts. Statistical practice involves effective technical communication with a variety of clients and collaborators, teamwork and collaboration. The guidelines recommend “study in a substantive area of application” in order to develop skills in translating domain-specific research questions into statistical questions (ASA, 2014).

It is important to note that the curriculum guidelines purposely focus on skills rather than topics. This emphasis allows us to move beyond the passive “what courses does a student *take?*” to the more active “what can a student *do?*” (Chance & Peck, n.d.). Emphasis on skills rather than topics encourages autonomy and individuality among programs, allowing institutions to develop programs that build on their particular strengths. It also shifts the focus to the student or graduate rather than the program itself. Our discipline is continually advancing, and some topics may become out-dated, but the skills required of a graduate in statistics or data science will persist.

3. Developing an Undergraduate Curriculum

The first step to developing any undergraduate curriculum is to define clear curriculum educational goals and learning objectives or outcomes. Educational goals of a curriculum describe what the program will provide students, and learning objectives are "statements of what a student is expected to know, understand and/or be able to demonstrate after completion of a process of learning" (Kennedy, 2007). Learning objectives can be written for a single course or for an entire curriculum and provide a foundation for guiding instruction and assessment. Students can refer to learning objectives to understand what is most important, and instructors can use the learning objectives to build lesson plans, activities, and assignments.

At the program level, five to ten learning objectives are recommended (Chance et al., n.d.). Avoid words such as "know" or "understand" and use more active words such as "design", "apply" or "assess." For example, we could improve the learning objective "students will *understand* the appropriate method of data analysis" by using more active verbs: "students will *utilize* an appropriate method of data analysis and orally *communicate* the results."

A second step is to develop a set of proposed required courses for the program and use these courses and the learning objectives to build a curriculum map. A curriculum map displays how program learning objectives are developed across an entire curriculum. Required courses are listed as rows, and learning objectives are listed as columns. When creating a new program, one can first draft a curriculum map consisting of existing courses that may fit into the proposed program. The curriculum map will then help highlight gaps in the program that suggest areas for new courses. Figure 1 displays a portion of the curriculum map developed for the data science major at the University of California, Irvine.

Learning Objectives

Year 4	LO1 Understand foundational mathematical principles	LO2 Understand foundational computer science principles	LO3 Understand foundational statistical principles	LO4 Understand the principles of statistical computing	LO5 Develop solutions for real-world data analysis problems	LO6 Demonstrate effective communication skills
ICS 51		x				
Stats 115			x	x	x	
Stats 170A				x	x	x
Stats 170B				x	x	x

Required Courses

Figure 1: A portion of the curriculum map for data science majors at the University of California, Irvine. This is the start of the curriculum map for courses taken in the fourth year of the program: ICS 51: Introductory Computer Organization; Stats 115: Introduction to Bayesian Data Analysis; Stats 170A and 170B: Project in Data Science.

The curriculum map in Figure 1 only indicates whether a given course addresses and assesses a particular learning objective in some detail (indicated by an "x"). However, we can go one step further. Instead of an “x”, we can write an “I”, “D”, or “M” for Introduce, Develop, or Master, respectively. A learning objective is introduced in a course if the students learn about key ideas and concepts at a general or basic level. A course develops a learning objective when students gain additional information related to the objective and begin to synthesize key ideas and skills. Mastering a learning objective usually occurs in upper division capstone or project courses, where students demonstrate the ability to perform the objective “with a reasonably high level of independence and sophistication” (“Creating a Curriculum Map”, 2014). Each learning objective should be introduced, developed, and mastered at least once across several courses. Similarly, each course should support at least one (ideally more than one) learning objective.

Curriculum maps should be used to identify courses in which to offer signature assignments, generally in courses where mastery of the learning objective is expected. Faculty can also use curriculum maps to guide a student to appropriate courses if that student is struggling with a particular learning objective (“Creating a Curriculum Map”, 2014). Curriculum maps may also be used in guiding assessment of the entire curriculum.

4. Data Science Major at University of California, Irvine

The Department of Statistics at the University of California, Irvine was formed in 2002. We currently have ten full-time faculty members, and offer MS and PhD degrees in Statistics. We offer an undergraduate minor in statistics, and as of fall 2015, a Bachelor of Science in Data Science. An undergraduate major in statistics is included in the future plans for the department.

The UCI Department of Statistics is somewhat unique among statistics departments, housed in The Donald Bren School of Information and Computer Sciences (ICS). ICS is comprised of three departments: Statistics, Computer Science, and Informatics. (Mathematics is located in the School of Physical Sciences.) Not only are all three departments in the same school, but also we are located in the same building. This close proximity, both physically and administratively, has led to close working relationships between the three departments, particularly Computer Science and Statistics.

4.1 UCI Data Science Initiative

Started in July of 2014, the Data Science Initiative at the University of California, Irvine coordinates “the activities of researchers and students across campus involved in various aspects of data science” (“UCI Data Science Initiative”, 2014). One of the core aims of the Data Science Initiative was to develop an undergraduate major in data science.

We formed a committee of two Statistics faculty and two Computer Science faculty in the spring of 2014 to develop the major. The committee started by looking at other programs in data science and computational statistics, for example, Winona State University (BS in Data Science), Northern Kentucky University (BS in Data Science), University of California, Davis (BS in Statistics: Computational Statistics Option) and College of Charleston (BS in Data Science). The committee studied the ASA 2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science and other resources. We performed an inventory of current computer science and statistics courses and identified potential courses for the major, then identified gaps in the major that would need to be filled with new courses.

4.2 Educational Goals and Learning Objectives

The broad educational goals of our Data Science major are to produce students with foundational training in data analysis, with a dual emphasis on the principles of statistics and computing, and the ability to apply these principles to a range of data-driven problems. More specifically, the educational goals are as follows:

- Provide students with a foundation in mathematical and statistical aspects of data analysis;
- Provide students with a foundation in the general principles of computer science;
- Teach students how to utilize their knowledge of statistical and computing principles to develop algorithms and software for solving real-world data analysis problems;
- Provide students with practical experience in applying their knowledge of theories, methods, and tools, to a variety of data analysis problems;
- Teach students how to communicate effectively using data.

The major has six broad learning objectives for graduating seniors:

1. Demonstrate knowledge of foundational mathematical concepts relevant to data analysis.
2. Demonstrate knowledge of basic principles in computer science.
3. Demonstrate knowledge of foundational statistical concepts.
4. Demonstrate knowledge of basic principles in statistical computing.
5. Demonstrate the ability to take a real-world data analysis problem, formulate a conceptual approach to the problem, match aspects of the problem to previously learned theoretical and methodological tools, break down the solution into a step-by-step approach, and implement a working solution in a modern software language.
6. Demonstrate the ability to communicate effectively in data analysis projects.

Each of these learning objectives is described in more detail in the major proposal, listing specific skills and topics included in each objective.

In the development of the data science major at UCI, the committee chose an initial set of courses before writing the learning objectives. Ideally, however, the learning objectives are written before selecting courses for the program. In fact, this is often an iterative process, as was the case for us. It may take several cycles of writing and revising learning objectives and selecting courses that meet those learning objectives for the major until a final curriculum is set.

4.3 Curriculum

The BS in Data Science at UCI has a dual emphasis in statistics and computer science. It is a fairly heavy major, with 30 required 10-week courses (UCI is on the quarter system) plus three elective courses. Table 1 summarized the required courses, with newly developed courses in italics.

Mathematics: Calculus (single-variable and multivariable), linear algebra
Statistics: Introductory statistics, *statistical computing and exploratory data analysis*, statistical methods and data analysis (year-long sequence), introduction to probability and statistics (year-long sequence), introduction to Bayesian data analysis
Computer Science: Boolean algebra and logic, discrete mathematics, introduction to programming (year-long sequence), programming in C/C++, data structure, introductory computer organization, introduction to data management, design and analysis of algorithms, machine learning and data mining
Informatics: Introduction to software engineering, information visualization
Data Science: S*eminar in data science, project in data science (two-quarter sequence)*

Table 1: Required courses for the Bachelor of Science in Data Science at UCI. Newly developed courses are in italics. Students are also required to take three additional elective courses from a list of approved electives in statistics or computer science.

The ASA undergraduate curriculum guidelines are written in such a way to allow individual programs to build on their own strengths, and this is the case for UCI. Since the Statistics and Computer Science departments already have a close working relationship, most of the data science program consists of existing courses. The major only had to incorporate essentially three new courses: a first-year one-credit seminar in data science, a second-year statistical computing and exploratory data analysis course, and a senior-year two-quarter project course sequence. Current researchers and practitioners of data science introduce students to topics at the forefront of the discipline in the first-year seminar in data science, which serves as an "introducing" level for the major's learning objectives. Learning objectives are "developed" in the statistical computing and exploratory data analysis course, which has one quarter of statistics and one quarter of computer science as prerequisites. In their senior year, students "master" the learning objectives through a two-quarter team-based project course, which provides a capstone experience to the major. The project course, team-taught by statistics and computer science faculty, gives students experience with real-world problems at the intersection of statistics and computer science, allows students to practice written and oral communication and presentation skills, and exposes students to the types of problems they will encounter as a practicing data scientist.

4.4 Connecting the Guidelines to the Curriculum

Revisiting the key points of the ASA undergraduate curriculum guidelines, the entire data science curriculum was designed with "increased importance of data science" in mind. Our new course in statistical computing and exploratory data analysis will address this key point not only with our data science majors, but also with our statistics minors and future statistics majors. The second key point, "real applications as a major component available early in the curriculum" is addressed through our first-year seminar in data science and the senior year capstone project course sequence, as well as real-world applications interspersed throughout various courses in the curriculum. Bridging both computer science and statistics emphasizes both statistical and algorithmic models, addressing the third key point of “more diverse models and approaches.” The fourth key point, “ability to communicate”, is something we kept in mind throughout the entire process. Many courses within the major already require student presentations, group projects, and written reports, and our capstone project course will provide an additional outlet to build communication skills.

Each of the skill areas described in the undergraduate curriculum guidelines is also addressed in our major. Students pick up the statistical methods and theory skills from the core statistics courses, and data manipulation and computation from the core computer science courses. Mathematical foundations are included in the major (calculus, linear algebra, and probability). Statistical practice is incorporated through a required writing course specific to the School of Information and Computer Science, the capstone project course sequence, as well as projects, presentations and reports incorporated into many existing courses. The last skill area in the undergraduate curriculum guidelines is discipline-specific knowledge, and students will pick up some domain-specific knowledge in the capstone project course sequence, though the major does not require a specific area of application.

5. Conclusions and Future Challenges

In developing the data science major at UCI, partnering statistics and computer science faculty was crucial. The small committee of four faculty members, balanced between statistics and computer science, allowed for a diverse set of perspectives within a manageable sized team. Though the primary committee developed the major proposal, committee members also sought input from faculty in other disciplines that use data science, and from faculty involved in other aspects of the UCI Data Science Initiative.

A few of the ASA Curriculum Guidelines for Undergraduate Programs in Statistical Science were not implemented in our major. Due to the large number of required courses in statistics and computer science, the major does not require a field of application. Additionally, we opted for an intensive team-based project course in place of required internships in data science with local companies (though internships are still strongly encouraged). Tracks or concentrations within the data science major were also not incorporated into the major.

Though the data science major was the Statistics Department’s first major, the department plans to offer a statistics major in the future. A statistics major would consist of less computer science than the data science major and more statistics, including possible future courses such as experimental design or space-time analysis. The statistics major would provide a more solid foundation for graduate study in statistics by requiring more mathematics (e.g., real analysis) (though graduates with a data science major should also have the background for graduate study in statistics). Additionally, the statistics major may have more room to include an area of application.

With the formation of a new major in data science comes several future challenges. The fast-changing nature of data science technologies challenges us to keep pace with the progressing discipline in our courses. A large demand for the major, as well as statistics and computer science courses in general, results in larger class sizes, many over 200 or even 300 students per class. The ability to offer frequent and effective feedback and opportunities for practicing communication skills in the form of written reports or oral presentations is a major obstacle in large classes. Future plans also include a minor in data science. Many of our core data science major courses require several prerequisites, and it will take some creativity to develop a minor that offers a substantive indoctrination into the field of data science through a small number of courses. Lastly, any curriculum requires frequent review and assessment of the success of the major, providing a tool to determine if students are meeting the desired learning objectives and when we need to adjust any components of the program.

References

American Statistical Association Undergraduate Guidelines Workgroup. (2014). 2014 *curriculum guidelines for undergraduate programs in statistical science*. Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/education/curriculumguidelines.cfm

California State University Long Beach College of Education (2014). Creating a Curriculum Map. Retrieved from https://www.ced.csulb.edu/offices/assessment-office/creating-curriculum-map.

Chance, B. and Peck, R. (n.d.) *From Curriculum Guidelines to Learning Objectives: A survey of Five Statistics Programs.* Retrieved from http://www.amstat.org/education/curriculumguidelines.cfm

Kennedy, D. (2007). *Writing and using learning outcomes: A practical guide.* Quality Promotion Unit, UCC. Retrieved from: http://lo-hei.net/wp-content/uploads/2013/10/A-Learning-Outcomes-Book-D-Kennedy.pdf

Pierson, S. (2014). Categorization of Statistics Degrees through CIP Codes and Growth in Statistics-Related Degrees. Retrieved from http://community.amstat.org/blogs/steve-pierson/2014/07/28/categorization-of-statistics-degrees

UCI Data Science Initiative. (2015). Retrieved from http://datascience.uci.edu/