Undergraduate Learning Outcomes for Achieving Data Acumen

Anna Bargagliotti, Wendy Binder, Lance Blakesley, Zaki Eusufzai, Ben Fitzpatrick, Maire Ford, Karen Huchting, Suzanne Larson, Natasha Miric, Robert Rovetti, Kala Seal & Thomas Zachariah
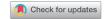
Published online: 20 Jul 2020.

Submit your article to this journal

Article views: 2494

View related articles

View Crossmark data

Citing articles: 2 View citing articles

Taylor & Francis
Taylor & Francis Group

DATA SCIENCE

Undergraduate Learning Outcomes for Achieving Data Acumen

Anna Bargagliotti[a], Wendy Binder[b], Lance Blakesley[c], Zaki Eusufzai[d], Ben Fitzpatrick[a], Maire Ford[e], Karen Huchting[f], Suzanne Larson[a], Natasha Miric[g], Robert Rovetti[a], Kala Seal[h], and Thomas Zachariah[a]

[a]Department of Mathematics, Loyola Marymount University, Los Angeles, CA; [b]Department of Biology, Loyola Marymount University, Los Angeles, CA; [c]Department of Political Science, Loyola Marymount University, Los Angeles, CA; [d]Department of Economics, Loyola Marymount University, Los Angeles, CA; [e]Department of Psychology, Loyola Marymount University, Los Angeles, CA; [f]School of Education, Loyola Marymount University, Los Angeles, CA; [g]Department of Sociology, Loyola Marymount University, Los Angeles, CA; [h]College of Business Administration, Loyola Marymount University, Los Angeles, CA

ABSTRACT

It is imperative to foster data acumen in our university student population in order to respond to an increased attention to statistics in society and in the workforce, as well as to contribute to improved career preparation for students. This article discusses 13 learning outcomes that represent achievement of undergraduate data acumen for university level students across different disciplines.

1. Introduction

Recently the value of an undergraduate degree has been challenged (Estes 2011), and demands for greater accountability in higher education have emanated from prospective students, their parents, business leaders, and politicians. The economic climate and employment considerations are central to many of these concerns. The percentage of students who report that their decision to go to college has been strongly shaped by a desire to "to get a better job" has increased in recent years, and in 2012, 88% of first-year students reported this factor as very important (Eagan et al. 2014). Colleges and universities have repeatedly been called upon to do a better job in preparing students for careers.

One area that needs to be strengthened in response to the career climate is student preparation in statistics and data science. The Chronicle of Higher Education recently listed the growth of data science programs as a key trend in higher education. However, they also noted that data science programs are being added without careful attention to what a data science curriculum should look like. Moreover, because data and statistics play an important role in all disciplines, undergraduate curricula in statistics and data science may be embedded within different disciplinary contexts. As such, there is a need for a set of comprehensive learning outcomes to help guide data learning across the disciplines. Such learning outcomes will help departments across institutions, administrators, as well as individual faculty to better understand how statistics and data courses across departments fit together to provide a coherent curriculum. Data education entails ensuring that students not only have sound computing, data analysis, and statistical skills, but also have good communication skills and the ability to work as part of a team (Zorn et al. 2014; Holdren and Lander 2012;

Davenport and Patil 2012). As noted by Horton and Hardin (2015), "the idea that an undergraduate statistics [major] develops general problem solving skills to use data to make sense of the world is powerful." This is what offerings in colleges and universities in statistics should strive to achieve—nimble computing data problem solvers (Nolan and Temple Lang 2010; Nolan and Temple Lang 2015).

Some data science education recommendation and guidelines already exist in the literature. For example, in 2018, the Two-Year College Data Science Summit published a report outlining recommendations for data science programs at community colleges (Gould et al. 2018). Recommended program outcomes are organized into four categories: computational, statistical, data management and curation, and mathematical. The program outcomes are further partitioned into foundations, thinking, and modeling outcomes (p. 16). Overall the guidelines provide community colleges housing data science programs a set of explicit learning outcomes to organize their programs around. Also in 2018, the National Academies Press put forth the Envisioning the Data Science Discipline: The Undergraduate Perspective Interim Report (National Academies of Sciences et al. 2018b). This report defined the term "data acumen" as the ability to make good judgments and decisions with data. It also notes that data acumen is "not a final state to be reached but rather a skill that data scientists develop and refine over time." (p. 12). To develop data acumen, mathematical foundations, computational thinking, statistical thinking, data management, data description and curation, data modeling, ethical problem solving, communication and reproducibility, and domain-specific considerations are needed (p. 33).

The learning outcomes presented in this article differ from those in these reports in a few ways. A main goal here is

not to provide statistics and data science guidelines for a specific program dedicated to data science but instead to present outcomes for working toward data acumen *across* university courses and *across* disciplines. Responding to the call in the National Academies report that data science requires participation from all different disciplines understanding that the degree to which different disciplines develop the components of data acumen varies (National Academies of Sciences et al. 2018a), this article presents a cross-disciplinary study that was undertaken to develop baseline learning outcomes for statistical and data learning at a university. A second National Academies report, Data Science for Undergraduates, notes the difficulty in furthering data science education across disciplines through upper division courses due to the varying topics in introductory courses in different disciplines. It also notes a need for cross-disciplinary coordination and collaboration from a wide spectrum of disciplines (p.39). While the Two-Year College report and the first National Academies Report discuss learning outcomes for specific programs in data science, this article outlines a series of common learning outcomes valid across disciplines for working toward data acumen on a university campus.

Following the recommendations of the American Statistical Association put forth in the *Curriculum Guidelines for Undergraduate Programs in Statistics* (ASA, Undergraduate Guidelines Workgroup 2014), this article discusses how statistics and data education bridge many disciplines and how the different disciplinary approaches can be integrated into one set of coherent learning outcomes for undergraduate education in statistics and data education. Overall, to fulfill the growing needs of the workforce, students graduating from college need to be prepared to tackle problems using technology, work with real data, and communicate their ideas.

2. Statistics and Data Science Education across Disciplines at Universities

Universities across the U.S. typically have many different statistics course offerings across campus. Because it is very common to have statistics courses housed in different disciplines (e.g., mathematics, computer science, psychology, economics), the ASA and Mathematical Association of America (MAA) offer guidelines for teaching introductory statistics targeted at non-statistics departments (ASA/MAA Joint Committee on Undergraduate Statistics 2014). Oftentimes these courses overlap and yet their prerequisite structures do not allow a student to move from a statistics course offered in one department to a more advanced course offered by another department. Departments often rightfully argue that the type of statistical techniques needed are discipline specific and thus necessitate the offering of a course within a specific discipline.

Although specific techniques do vary from discipline to discipline, certain basic themes of working with data should be present in all courses. Three important, fundamental, and particularly timely themes are that students need to (1) employ technology, (2) explore real datasets, and (3) practice communicating statistical ideas and results.

Scholarly articles and recommendations of professional organizations concerning undergraduate preparation in various disciplines, in addition to the ASA sponsored documents already discussed, align with these themes. For example, skills in statistics and the ability to work with data and technology are increasingly recognized as core components of an education in sociology (Wilder 2010). A 2010 report published by the American Psychological Association, recommends that psychology students complete coursework in statistics and research methods as early as possible and that the knowledge and skills gained from these courses be reinforced throughout the curriculum. A national study of undergraduate business education conducted by The Carnegie Foundation for the Advancement of Teaching concluded with recommendations that programs provide a stronger linkage between business, arts, mathematics, and science curricula and that programs promote courses that incorporate complex and ambiguous real-world issues and three essential modes of thinking: Analytical Thinking, Multiple Framing, and Reflective Exploration of Meaning (Colby et al. 2011). Statistics courses that incorporate the statistical thinking process of formulating a question, collecting appropriate data, choosing an appropriate analysis technique, and interpreting results (Franklin et al. 2007) promote these modes of thinking. Teaching statistics as an interrogative process is also stressed in both the GAISE college report (Everson et al. 2016) and the ASA Undergraduate Guidelines Workgroup (2014).

Several important reports have stated the need for students to work with real data. The Committee on the Undergraduate Program in Mathematics Curriculum Guide 2015 (Mathematical Association of American 2015) states "Working mathematicians often face quantitative problems to which analytic methods do not apply. Solutions often require data analysis, complex mathematical models, simulation, and tools from computational science." This report recommends that all mathematical sciences major programs include concepts and methods from data analysis and computing. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) college guidelines also include working with real data as one of the necessary six components of structuring an introductory statistics course (ASA **?**). In addition, the recommendations of the ASA on undergraduate programs in data science include Real Applications and Problem Solving as two of their Background and Guiding Principles. They state programs should "emphasize concepts and approaches for working with complex data and provide experiences in designing studies and analyzing real data (defined as data that have been collected to solve an authentic and relevant problem)" (ASA, Undergraduate Guidelines Workgroup 2014).

As data science has been described as an intersection of statistics with computer science, when considering undergraduate preparation, one must consider how the use of software interplays with statistics. Regardless of the discipline, technological fluency has become a must for success in the workforce. Therefore, university statistics and data science courses must incorporate heavy use of technology and computing.

The material commonly taught in introductory statistics courses often merely focuses on techniques. However, such methods are often "necessary but not sufficient" for modern data science (Hardin et al. 2015; Ridgeway 2016). Instead, an undergraduate education should focus on the unifying themes of working with technology, working with real data, and com-

municating results for *all* course offerings *across* campuses. Moreover, if a model existed for explicit learning outcome goals of an undergraduate education in statistics and data related courses, then the door may be open to creating a coherent curriculum for students seeking statistics and data education beyond what just their departments offer.

3. Undergraduate Data Pathways (UDaP) Study

The National Science Foundation (NSF)-funded project (NSF Grant No. 1712296), *Undergraduate Data Pathways* (UDaP), focused on understanding differences and similarities of statistics and data related course offerings across different disciplines. The project carried out a rigorous study to develop a set of learning outcomes for statistics and data related courses at the undergraduate level that integrated the data-related goals put forth by several different disciplines. This paper reports on that study and presents a set of learning outcomes (LOs) that work toward data acumen for university level students across different disciplines. If a student meets all of the LOs, the student will have achieved an introductory level of data acumen appropriate to the undergraduate level. The LOs not only reflect cross-disciplinary goals but they also reflect societal needs of data analysis. The study took place at Loyola Marymount University (LMU), a mid-sized comprehensive university in Los Angeles, California. Faculty from eight departments across campus carried out the study.

4. Methods

Five steps were undertaken to better understand the differences and commonalities of statistics and data education across disciplines and subsequently develop a unifying set of learning outcomes for undergraduate statistics and data education.

As a first step, a faculty working group consisting of LMU faculty from mathematics, economics, biology, psychology, sociology, business, and education was formed. While LMU does not have a department dedicated to statistics or data science, the Department of Mathematics, Department of Biology, Department of Engineering, Department of Economics, Department of Political Science, Department of Psychology, Department of Sociology, the School of Business, and the School of Education offer courses related to statistics and data analysis.

The formation of a working group of invested change agents was no easy task. The Associate Dean for Undergraduate Studies urged faculty across departments that had investment in statistics and data analysis to join the group. In addition, members of the research team personally reached out to faculty in other departments to encourage them to join the working group. A total of 10 faculty members were selected for the working group. The working group was centered around understanding the processes and support needed to implement the themes of communication, technology, and real data in statistics courses across the disciplines. Four meetings per semester were conducted over the course of two academic years. The purpose of the working group discussions was to gather qualitative data on how different disciplines articulated the importance of statistics and data analysis and to determine what all of the disciplines had in common.

A second step in the process was to develop and administer a 36 question survey to the working group. The survey asked about software platforms, data sources, types of class assignments offered (e.g., statistics investigations in the form of projects, problem sets), and the types of activities used in the classroom (e.g., students using computers in a lab setting, group work). The survey included questions from the Statistics Teaching Inventory (STI) developed by Zieffler et al. (2012) focusing on teaching practice, assessment practice, technology use, teaching, and assessment beliefs. The goal of the survey was to gather in-depth information about the statistical habits of the working group faculty across disciplines. See Appendix A for the inclusion of the entire survey.

Using the discussions and internal survey results, an initial set of learning outcomes was developed. Each disciplinary representative researched and brought forth any guiding documents that were present from their disciplines related to data education. In addition, each disciplinary representative gathered syllabi and University bulletin course descriptions for all of the courses taught within their discipline. Using a blinded exercise, the working group sorted the course descriptions by similarities—all course names were removed from the descriptions and the working group worked in pair groups to organize the descriptions into groups according to the topics covered within the courses. The names and departments of the courses were then revealed. This exercise was a catalyst for the working group to summarize common themes that were present across courses at the institution. These themes were outlined and noted. Syllabi were then reviewed to pick out how many courses highlight the themes and whether other themes were present that were not touched upon in the course descriptions. The working group members were asked to review the syllabi and then discuss the recurring themes present within and across disciplines. Courses were sorted into basic courses, introductory courses, application courses, and beyond courses. The discussions were guided by two of the PIs of the project (Bargagliotti and Larson). Bargagliotti and Larson guided the group in readings of papers in the literature discussing data acumen and readings of guidelines and reports from the different disciplines. They also presented enrollment data for specific courses at LMU (see Bargagliotti et al. 2020, for presentation of enrollment results) to help provide student context to the discussions. In addition, discussion questions focused on technology use and necessities were posed to the working group at each meeting session. This process led to the formulation by the group of explicit learning outcomes. The process was completed over the course of one year through monthly meetings and email and phone conversations in between the in-person meeting times.

To validate these learning outcomes, a third step in the development process included carrying out a larger-scale survey to the greater community, both academic and nonacademic, to garner thoughts on the necessary learning outcomes for data education at the university level. This survey gathered data on whether respondents agreed, were neutral, or disagreed that a learning outcome was important for achieving undergraduate data acumen at the undergraduate level. More specifically, the survey asked: *For each statement below, please mark whether you agree, disagree, or are neutral that the statement describes a data*

analysis skill that you believe a college graduate in today's society should have. The full survey is included in Appendix C.

A fourth step carried out by the research team was to review position statements, policy documents, and curriculum guidelines put forth by professional organizations (e.g., ASA, APA, AEA) regarding data education proficiency to understand whether there was common ground between the disciplines.

Based on the information gathered in these four steps, the culminating step of the work was to develop a final set of learning outcomes to represent appropriate data acumen at the undergraduate level across disciplines.

5. Findings

5.1. Working Group Survey Findings and Working Group Discussion

Nine members of the working group completed the internal survey. The survey was administered at the first group meeting before any discussion took place. The purpose of the survey was to gauge satisfaction with the manner in which the teaching and learning of statistics and data related topics was approached at the University as well as to gather baseline data on the typical statistical and data analytical processes used across disciplines. These data would then serve for a starting point of conversation to develop a set of learning outcomes for data education that would bridge the disciplines.

Of the nine respondents, only one reported they were happy with the course offerings and curriculum related to data, four responded they were somewhat happy, three were unable to judge, and one said they were not happy. Two main issues were identified as those keeping faculty from being satisfied or being able to change the course offerings and curriculum to their liking. The identified issues were the general feeling that the institution did not support current statistical needs – specifically with providing access to technology or materials needed to teach statistics and data analysis properly as well as the institution not providing enough faculty lines to cover the growing needs.

Respondents were asked what they would like to change about the course offerings and curriculum related to statistics if they had all of the resources needed. The responses were:

- Our students can take basic stats and they can take more advanced Biostats, though it isn't taught very frequently (once every 2 years). More gradations might be good, as well as more frequency. Also the class is co-taught with Bio and Math faculty, which is great.
- I am most familiar with the statistics requirement in the psychology dept and less aware of the offerings elsewhere.
    I know that in the psych dept in the past we didn't have enough people to cover stats and often relied on visiting professors or adjunct professors. This is changing now though. I will also say that it is difficult to get access to lab classrooms with computers for all of the stats sections that we offer.
- [The] math department has some solid courses, but it would be nice to have at least one advanced data science class.
- More computational statistics required [of students in order to graduate]

- I believe in the social sciences all students should be required to take an introductory statistics course, an empirical research methods course, and a qualitative methods course.
- Students do not have the ability to further their statistical knowledge past their own department offerings.
- Overlapping topics; no interactions between various departments

Several themes were present in the responses and these themes continued to emerge throughout subsequent discussions. These themes included: a general siloed approach to data education curriculum across departments, frustration over a lack of advanced courses, and a lack of understanding of what is happening in other departments.

Although our group of nine faculty was identified as the primary professors teaching data-related courses in their disciplines, there was repeated evidence that indicated that we had difficulty thinking of statistics past our own departments and across the university as a whole. For example, in multiple instances throughout discussions, the conversations revolved around single departments and single courses. There were many statements such as "in my department, in my course, we…" While this reaction was to be expected, there was extensive effort made to keep the cross-disciplinary goal in mind as the development of learning outcomes progressed. This cross-disciplinary focus was thus identified as a main take-away for the working group as we strove to make adjustments over the course of the next year.

Subsequent working group meetings consisted of discussions and group exercises that took a closer look at the University course descriptions of all courses offered at the University related to statistics and data analysis. A total of 29 courses offered at LMU were identified by the working group that spanned 11 different departments (see Table 1 for the list of courses, see Appendix B for brief descriptions of each course). The courses being offered cover a total of 11 different departments and therefore have a wide reach across the University. As shown in the Table 3, The College of Liberal Arts offers 12 courses related to data; the College of Science and Engineering offers 12, the College of Business Administration offers 4, and the School of Education offers 1 course. Of the 29 courses offered, 9 are lower division (shown in light gray) courses and 20 are upper division courses (shown in dark gray). Of these upper division courses, five were special reading courses offered in small settings.

Using the courses, the working group participated in an exercise where the course descriptions of all of the 29 courses were placed on 3×5 cards but without course names and titles. Each group member paired up with another member, with no pair coming from the same department. Each pair had to sort the cards by similarity of course content as well as difficulty. All pairs agreed that there appeared to be several basic and introductory statistics courses being taught across the University that had similar content. In addition, two other course types were identified – a research methods type course (where statistical methods were applied) and an advanced level course. The introductory courses could be further distinguished by those courses that covered regression and/or ANOVA versus those that did not.

Table 1. List of statistics and data-related courses at LMU.

Bellarmine College of Liberal Arts (BCLA)			
Department	Course number	Course title	Number and type of prerequisites
AFAM	208/2243	Africn Amercn Stud Rsrch Mthds	0
ECON	230/2300	Introductory Statistics	1 (calculus)
ECON	235/2350	Accel Intro Statistics	0
ECON	330/3300	Econometrics	3 (calculus, intro stats, econ)
ECON	532/5320	Advanced Econometrics	1 (linear algebra)
POLS	2100	Empirical Approaches	1 (political science)
POLS	4100	Advanced Empirical Methods	2 (political sci, empirical appr.)
POLS	440/4250	Public Policy Analysis	0
PSYC	241/243/2001	Statistical Methods for Psych	1 (gen psych)
PSYC	261/2002	Research Methods	2 (gen psych, stats for psych)
PSYC	4045	Advanced Research Methods	4 (gen psych, stats for psych, research mthds, brain /behavior)
SOCL	210/2100	Quantitative Research Methods	0

College of Business Administration			
Department	Course number	Course title	Number and type of prerequisites
AIMS	3710	Database Management Systems	2 (business, info tech in business)
AIMS	4760	Analytics & Bus Intelligence	2 (business, prod ops analysis)
AIMS	4798	*SS:Int to Big Data & Data Scnc	0
BADM	3010	Anlyticl Cncepts/Mthds fr Busn	0

School of Education			
Department	Course number	Course title	Number and type of Prerequisites
EDES	5407	Rsch Mth/Early Chld Assessment	0

Seaver College of Science and Engineering			
Department	Course number	Course title	Number and type of prerequisites
BIOL	367	▼Biological Databases	0
BIOL	398	*▼SS: Biostatistical Analysis	0
CMSI	367	▼Biological Databases	0
CMSI	485	Artificial Intelligence	2 (data structures, algorithms)
MATH	104	Elementary Statistics	0
MATH	204	Applied Statistics	1 (calculus)
MATH	360	Intro Probability/Stats	1 (integral calculus)
MATH	398	*▼SS: Biostatistical Analysis	0
MATH	560	Adv Topics in Prblty & Stats	2 (multivar calc, intro prob/stats)
MATH	598	*SS: Stats & Modeling for Tchrs	0
MATH	599	Probability & Statistics Lab	0
MECH	598	*SS: Statisticl Dsng of Exprmnt	0

Source. Data retrieved from LMU Registrar's Office, Upper Division in Dark Gray, Lower Division in light Gray, *Special Topics, ▼Co-Taught.
AFAM = African American Studies, ECON = Economics, POLS = Political Science, PSYC = Psychology, SOCL = Sociology, AIMS = Applied Information Management Systems, EDES = Education, BIOL = Biology, CSMI = Computer Science, MATH = Mathematics, MECH = Mechanical Engineering.
Some courses listed in the table have multiple course numbers due to cross-listings or changing courses numbers during the study period.

Table 2 shows the established descriptions for the different types of courses found.

Examples of Basic Introductory courses were MATH 104, a non-calculus based introductory statistics course aimed at general education requirements. In contrast, MATH 204 was considered to be a Basic Introductory Statistics + Regression/ANOVA course. Similarly, SOCL 2100 and PSYC 2001 were also in this category of courses. Beyond Basic courses were defined by the inclusion of topics that are past basic inference and linear regression. For example, Biological Databases and PSYC 2002 both go beyond the introductory courses by including topics specific to data in biology and database structures and research methods in psychology. Special Topics courses included Machine Learning and Deep Learning as well as courses like Econometrics.

The classification exercise led to an attempt by the working group to create a set of interchangeable courses – interchangeable intro level statistics courses, interchangeable research

Table 2. Course-level types.

Course type	Description
Basic Introductory Statistics	Basic introductory courses that cover descriptive statistics, distributions, and some inferential statistics
Basic Introductory Statistics + Regression/ANOVA	Basic introductory courses that cover descriptive statistics, distributions, inferential statistics, and regression/ANOVA
Beyond Basic	Second course beyond level 1 course. Level two courses cover statistical content past basic inferential statistics and linear regression
Application	Application courses where students use their level 1 and/or level 2 skills in an applied setting
Special Topics	Special topics courses that are meant for deeper understanding of statistics

Table 3. Proposed Undergraduate Data Pathways (UDaP) learning outcomes.

1. Student should be able to formulate questions about univariate data, collect univariate data/consider univariate data, analyze univariate data, and interpret results
2. Student should be able to understand, calculate and interpret descriptive measures for quantitative and categorical variables to describe characteristics of the data
3. Students should be able to create and interpret basic data visualizations for quantitative and categorical variables
4. Students should be able to understand, carry out and interpret basic inferential procedures for one or two samples.
5. Students should be able to understand, carry out and interpret statistical procedures for predicting future data (predictive inference)
6. Students should carry out and communicate results from an extensive data-driven project that is related to a real-life problem (extensive means that the project takes more than two weeks to complete and is worth at least 25% of the final grade)
7. Students should be able to communicate their analyses and the interpretations of their results in a manner that is appropriate to their discipline in the context of the data
8. Students should be able to select appropriate methods for data analysis and explain limitations of their analyses and interpretations
9. Student should be able to formulate questions about multivariate data, collect multivariate data/consider multivariate data, analyzing multivariate data, and interpret results
10. Student should be able to use current statistical software, or statistical packages appropriate to the discipline and context beyond basic Excel or a calculator
11. Student should be able to write a program (using a programing language) to analyze data
12. Students should study at least one type of advanced data-analytic methods such as (not limited to) generalized linear models, Bayesian analysis, advanced probability theory and stochastic processes, non-linear models, machine learning, advanced study-design, big data analysis, econometrics, or statistical computing

methods courses, and interchangeable advanced courses. For example, if two courses within a group were deemed interchangeable, then those courses would fulfill the same requirements. The idea of creating some type of interchangeability course map was grounded in the belief that students might then be provided more ways to reach advanced content. It was also at this third meeting that the working group determined that a set of learning outcomes for a complete data pathway explicitly needed to be defined. All meetings that followed focused solely on the purpose of determining these learning outcomes.

To guide the creation of the learning outcomes, further analyses were done on the survey discussed above in order to identify the types of statistical techniques that were frequently used in each discipline. The results-were categorized into five groups:

- Descriptive statistics
- Visualization
- Inferential
- Predictive
- Application

Based on these results, the working group defined a set of 12 learning outcomes with the idea that certain learning outcomes might be developed within a category of interchangeable courses. Table 3 presents the initial 12 learning outcomes put forth by the working group.

The learning outcomes highlighted in yellow describe the Descriptive bullet, the blue describes the Visualization, the purple describes Inferential, and the green describes Predictive. Several outcomes, highlighted in orange, focused on Application. The remaining outcomes characterized data processes. Using these 12 learning outcomes as a guide, an external community survey was administered.

5.2. Community Survey

To validate the 12 learning outcomes, a community survey was administered online. The goal of this survey was to assess whether peers at other universities and in industry would also view this set of outcomes to adequately represent the skills that a university graduate should have today. The online survey was sent out by members of the working group to peers and

Table 4. Distribution of background of survey respondents.

Background of survey participants	Frequency	Relative frequency
College/University Administrator	9	0.03
College/University Faculty	235	0.82
Government Scientist/Researcher/Consultant	9	0.03
Industry Scientist/Researcher/Consultant	18	0.06
Other (please specify on next question)	14	0.05

Table 5. Distribution of disciplines of survey respondents.

Discipline	Number of people who took survey
Statistics	133
Mathematics	44
Psychology	25
Business	22
Other	14
Education	9
Biology	8
Sociology	8
Economics	7
Data Science	7
Computer Science	3
Chemistry	2
Physics	1
Political Science	1

to listserves for several disciplines (specifically sent out by the American Statistical Association and CAUSEWeb). It was also posted on several listserve forums (e.g., isostat). A total of 367 people opened the survey and 287 people completed the survey within the allotted time frame of one week. Table 4 shows the distributions of backgrounds of people who completed the survey. The survey was largely dominated by College and University Faculty with 82% of the total respondents. Industry scientists, researchers or consultants made up the next largest category at approximately 6% of the respondents.

A total of 14 disciplinary backgrounds were represented in the respondents as noted in Table 5. Statisticians were the largest group of respondents, with mathematicians and psychologists being the second and third largest.

Because the American Statistical Association (ASA) helped in the distribution of the survey, it was expected that statistics would have a large response rate. The working group team all sent out to their personal contacts, however, due to feasibility, the only large organization to actively post and distribute the

Table 6. Percentages of agree, neutral, and disagreements.

LO Summary	LO #	Agree (%)	Neutral (%)	Disagree (%)
Univariate data	1	91	6.6	2.4
Descriptives	2	93.7	5.2	1.0
Visualizations/Graphs	3	94.8	1.4	3.8
Inferential procedures	4	73.2	22.3	4.5
Predictive	5	54	34.1	11.8
Project	6	47.7	16.7	35.5
Communicate in context	7	95.8	1	3.1
Limitations	8	75.3	18.4	6.2
Multivariate data	9	45.6	37.3	16.1
Use software	10	57.8	27.9	14.3
Write program	11	21.9	31.7	46.3
Advanced methods	12	18.5	36.9	44.6

survey was the ASA. Despite the imbalance in discipline representation of the survey respondents, the responses were still varied.

Table 6 shows the percentage of survey respondents that agreed, were neutral, or disagreed with the statement that the learning outcome was an important skill that a university student must acquire.

Respondents grouped the LOs into roughly three categories. Of the 12 learning outcomes, four seemed especially important, as 90% of respondents agreed that they are important skills that a university student should acquire. These included univariate statistics, descriptive statistics, graphs and visualizations, and communicating in context. Five other learning outcomes had a large majority of respondents stated that they agreed or were neutral. This category included inferential statistics, predictive statistics, discussion of limitations, multivariate statistics, and use of software. Only three learning outcomes had large disagreements with the statements. This third category included having a large project, writing a program to analyze data from scratch, and studying advanced statistical methods.

Based on these data, the working group agreed that a student meeting all 12 learning outcomes would be deemed to have undergraduate data acumen. Due to the disagreements on three learning outcomes, subsequent levels of data acumen were then defined (see Bargagliotti et al. 2020 for a description of the categorizations of levels of acumen).

5.3. Policy Documents

To further validate the learning outcomes, nine curriculum guidelines from various professional organizations were reviewed by the working group. These guidelines specifically discussed students' necessary data acumen skills for a given discipline. The working group identified different disciplines that had position statements or curriculum guidelines that mentioned statistics or data education explicitly. Those disciplines represented in the policy documents were: mathematics, statistics, psychology, economics, sociology, science, engineering and medicine. The policy documents reviewed were:

- American Statistical Association, Curriculum Guidelines for Undergraduate Programs in Statistical Science
- Mathematical Association of American CUPM-MAA's Committee on the Undergraduate Program in Mathematics

- American Statistical Association, Guidelines for Assessment and Instruction in Statistics Education (GAISE) in Statistics Education (GAISE) College Report College Report2016.
- American Psychology Association, APA Guidelines for the Undergraduate Psychology Major
- American Economic Association, Recommended Mathematical Training to Prepare for Graduate School in Economics.
- American Sociological Association, The Sociology Major in the Changing Landscape of Higher Education: Curriculum, Careers, and Online Learning, A Report of the ASA Task Force on Liberal Learning and the Sociology Major
- National Academies of Sciences et al. (2018a), Data Science for Undergraduates: Opportunities and Options.
- National Academies of Sciences et al. (2018b), Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim Report.
- American Statistical Association, The Two-Year College Data Science Summit

Table 7 illustrates that seven of the 12 learning outcomes were discussed in *all* of the policy documents. The remaining learning outcomes were supported by most of the documents.

Interestingly, the policy documents all mentioned a learning outcome that was not included in the hypothesized 12 outcomes. That is:

Students should become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted.

Due to its inclusion in all of the policy documents from the various disciplines, the working group felt that it should be added to the 12 developed learning outcomes. This LO aligned with growing societal needs of being able to merely ingest the news and participate in the information age. Because it was included in all eight of the policy documents, the project team opted to include it as an explicit LO. A total of 13 LOs were then proposed.

5.4. Final Learning Outcomes

Table 8 presents the final 13 Undergraduate Data Pathways (UDaP) learning outcomes that were established as important for students to meet at the university level currently today. Several edits to the original learning outcomes were undertaken. They were:

- LO6 emphasizes that the project must count for a large portion of the final grade but does not specify an arbitrary percentage
- LO11 articulates that the use of software be used to manipulate, extract information, and carry out statistical analyses from data
- LO12 is rewritten to better reflect the data tasks a student would undertake using a software program

These adjustments were made based on the open comments received in the community survey, feedback from reviewers of this article during the revision process, feedback from audiences

Table 7. Policy document alignment with LOs.

Subject/Association	Learning Outcomes											
	1	2	3	4	5	6	7	8	9	10	11	12
Statistical Science/ASA	x	x	x	x	X	x	x	x	x	x	x	x
Math, Applied Math/MAA	x	x	x	x	X	x	x	x	x	x	x	x
Stat Ed/GAISE	x	x	x	x	X		x	x	x	x		
Psychology/APA	x	x	x	x		x	x		x	x		
Econ/Amer Econ Assoc	x	x	x	x	X	x	x	x	x	x	x	x
Sociology/ASA	x	x	x	x		x	x	x	x	x		
Data Science/National Academies 2018a	x	x	x	x	X	x	x	x	x	x	x	x
Data Science/National Academies 2018b	x	x	x	x	X	x	x	x	x	x	x	x
Data Science/2-Year Colleges	x	x	x	x	X	x	x	x	x	x	x	x

Table 8. Final UDaP learning outcomes.

	Learning Outcomes
1	Students formulate and/or address questions about univariate data, collect/consider univariate data, analyze univariate data, and interpret results
2	Students understand, calculate, and interpret descriptive measures for quantitative and/or categorical variables to describe characteristics of the data
3	Students create and interpret basic data visualizations for quantitative and categorical variables
4	Students understand, carry out, and interpret basic inferential statistical procedures for one or two samples
5	Students understand, carry out, and interpret results from estimating statistical models for bivariate data (e.g., linear regression, interpolation, extrapolation, predictive inference)
6	Students carry out and communicate results from extensive data-driven project(s) related to a real-life problem (Extensive means that a single project takes more than two weeks to complete or a series of projects take more than two weeks to complete and are worth a large percentage of the final grade.)
7	Students communicate their analyses and the interpretations of their results in a manner that is appropriate to their discipline in the context of the data (e.g., communication could be emphasized with presentations, oral explanations of results, oral/written answers for in-class work, written explanation of results)
8	Students understand the implications of study design, can select appropriate statistical methods for data analysis, and can explain limitations of their analyses and interpretations
9	Students become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted
10	Students formulate and/or address questions about multivariate data, collect/consider multivariate data, analyze multivariate data, and interpret results
11	Students use current statistical software or statistical packages that are appropriate to the discipline and context beyond basic Excel or a calculator to manipulate data, extract information from data, and carry out statistical analyses with data
12	Students write a program (using a programing language) to manage and curate data by finding, manipulating, analyzing data or extracting information from the data
13	Students study at least one type of advanced data-analytic methods such as (but not limited to): generalized linear models, Bayesian analysis, advanced probability theory and stochastic processes, non-linear models, machine learning, advanced study-design, big data analysis, econometrics, or statistical computing

when the paper was presented in three different settings, discussions among the PIs on how to incorporate the comments, and approval from the working group members in writing the final LOs. Students meeting these 13 learning outcomes are deemed to have undergraduate data acumen.

The UDaP learning outcomes span both content and process. The important themes of using real data, communication with data, and technology are well-represented within the learning outcomes as well. These outcomes are meant to be broad and cross-disciplinary so they can serve as benchmarks *across all* disciplines offering statistics and data education courses on a university campus. These learning outcomes stemmed from two-years of discussions within the working group as well as the review of the policy documents and the community survey.

6. Discussion and Future Research

While there has been a large increase in data science and statistics majors and minors across the US over the past several years (Pierson 2018), explicit learning outcomes to govern such programs are relatively new (see Gould et al., 2018; National Academies of Sciences et al. 2018a and 2018b). Furthermore, while there is consensus that data education reaches across disciplines the wide reach and wide importance of data across disciplines makes it difficult to put forth coordinated efforts

for student learning. In the cross-disciplinary context across departments and disciplines, no set of coordinated learning outcomes exist as a bridge to data education. There is an important need to acknowledge that data education is not taught solely in statistics, computer science departments, or within a single data science program but instead working with data is present in most disciplines and is often intertwined with disciplinary content. Therefore, although guidelines exist that specify recommendations on how to teach statistics courses (GAISE, **?**) and guide data science specific programs (Gould et al., 2018; National Academies of Sciences et al. 2018a, 2018b), these guidelines are not designed to bridge the interdisciplinary context.

Several challenges emerge as data education is conceptualized across disciplines. Perhaps a first step in advancing this conceptualization is an agreement of some basic content and process outcomes that students should acquire. The implementation of such outcomes necessitates departmental agreements and a concerted effort to create opportunities for students to advance their data acumen despite potential departmental limited offerings.

To develop goals for data education at the undergraduate level, the UDaP project explicitly considered the cross disciplinary nature of coursework related to data as well as the overall learning goals for students driven by current workforce

and societal needs. As society pushes toward being more data-driven, it is important to understand and characterize what education should be doing as a response. Moreover, cross-disciplinary demands are more and more emerging in society with data being embedded in policy and discussions across all subjects. As such, how we conceptualize data acumen at the undergraduate level must be flexible enough to bridge many contexts and students with diverse academic backgrounds. This is different than the way the literature has conceptualized data science as being a three circle Venn diagram with computer science, statistics, and context; instead undergraduate data acumen aims to be flexible and broad to span disciplines. In other words, a sociology major must have the opportunity to gain data acumen just as much as a computer science major.

The UDaP learning outcomes presented in this paper can be used by colleges and universities that plan to assess their capacity across disciplines to produce undergraduates with data acumen by matching existing course offerings with the learning outcomes presented here. This could provide insight about the accessibility, quantity, and difficulty of existing pathways to achieving data acumen and guide the resource-efficient development of new pathways using cross-disciplinary badges, concentrations, minors, or majors. The UDaP learning outcomes can form a basis for ongoing assessment of data-related concentrations, minors, or majors. Moreover, they could form a basis for assessment of the role of co-curricular learning through internships, campus jobs, etc., toward students earning badges around data acumen. Universities with specific statistics departments or data science programs can lead such efforts by ensuring that their offerings can meet the learning outcomes without many prerequisite costs to students. Efforts for general data education courses (much like writing requirements) required by all students at a university could fulfill such a need.

Through a rigorous process, UDaP developed a set of 13 learning outcomes for undergraduate data acumen at the university level. The learning outcomes focused around three important themes of working with real data, communicating data driven results, and working with technology. Of the developed learning outcomes, five focus on process and communication while eight focus specifically on content. This breakdown reflects the changing needs of statistics education today.

This paper offers the important initial step in finding common ground across disciplines. The creation of a working group of "change agents" from different disciplines on a university campus that are invested in furthering data acumen in students has been an invaluable asset to the project. Next steps for research could include the design of assessments and curriculum that could bridge disciplinary contexts as well as the development of curriculum and projects that foster collaboration among students and embody the learning outcomes (e.g., *https://ww2.amstat.org/education/datafest/*, *https://www.causeweb.org/usproc/*).

The authors hope that this study will persuade readers to consider doing something similar on their own campuses. The manuscript provides an example of how to have cross-disciplinary discussions which can be invaluable to creating opportunities for students to achieve data acumen.

Appendix A

A.1. Internal Working Group Faculty Survey

Welcome to Project Undergraduate Data Pathways (UDaP). As an initial step in our research, we would like to gather some feedback about your opinions and thoughts about data analysis and statistics at LMU and beyond. We greatly appreciate you taking the time to answer the questions below. This survey should take approximately 10 minutes to complete.

A.2. Background Information

1. Please enter your name: _____
2. What is your home department?

   Business
   Educational Psychology/Educational Statistics
   Mathematics
   Mathematics Education
   Psychology
   Economics
   Biology
   Sociology
   Other, please specify: _____

3. Please classify your position:

   a. Adjunct Faculty/Instructor (part time)
   b. Adjunct Faculty/Instructor (full time)
   c. Faculty (tenure track)
   d. Faculty (tenured)
   e. Other, please specify: _____

4. How many years have you been teaching statistics or data analysis courses at LMU or somewhere else? Please specify how many years at LMU and how many years at other places.
5. What courses do you teach related to statistics/data analysis at LMU? List the name of the course and the course number.
6. Please rate the frequency with which you analyze data outside of your coursework in statistics (e.g., research, consulting, etc.)

   a. On a daily basis
   b. On a weekly basis
   c. On a monthly basis
   d. Once per semester
   e. Once per year
   f. Less than once per year but occasionally
   g. Never

7. Select all the types of data analyses you perform outside of your coursework (e.g., research, consulting)

   a. Descriptive
   b. Confidence intervals
   c. T-tests
   d. ANOVA
   e. Non-parametric tests
   f. Regression
   g. Linear models
   h. Visualization
   i. Parameter estimation
   j. Factor analysis
   k. SVD
   l. PCA
   m. Non-linear regression
   n. Generalized linear models
   o. Logistic regression
   p. Classification

q. Clustering
r. Supervised learning
s. Unsupervised learning
t. Longitudinal models
u. Time series
v. Other, please specify: _____

8. What type of statistical procedures do you do a lot of? For example, do you do factor analysis over and over again or do you do a variety of analyses depending on the type of data you have in front of you?

9. How often do you use the same set of statistical procedure?

a. 0-25% of the time
b. 26-50% of the time
c. 51-75% of the time
d. 76-100% of the time

10. Circle all the types of data you use outside of your coursework (e.g., research, consulting)

a. Observational data with no time or spatial component
b. Longitudinal/Repeated Measures/Panel
c. Time series
d. Big data (e.g., image, voice, spatial), please specify: _____
e. Experimental
f. Other, please specify: _____

11. Has anything changed with the way your discipline makes use of statistics or gives importance to statistics over the past 5, 10, 20 years? Please explain.

12. In your discipline, do you deal with Big Data and if so, how is Big Data defined?

A.3. Teaching Information

Write in the table the percent of class time that you spend for each of the items below for each of your classes. In the table, write the course number for each item. For example, I teach math 104, 204, and 360 and on the first item, I would fill in the table in the following way:

	0-25% of the class time	26–50% of the class time	51–75% of the class time	51-75% of the class time	75-100% of the class time
13. Teacher presentations (e.g., lectures, demonstrations, etc.) are used to help students learn statistics	104	204, 360			

Fill in the table below.

	0-25% of the class time	26–50% of the class time	51–75% of the class time	51-75% of the class time	76-100% of the class time
12. Teacher presentations (e.g., lectures, demonstrations, etc.) are used to help students learn statistics					
13. Activities are used to help students learn statistics					
14. Students use computing tools to analyze data					
15. Students use computing tools to help them understand statistics					
16. Real datasets are used during instruction					
17. Students use graphical displays to help visualize data					
18. Students write-up reports of their findings					
19. Students use distribution tables to compute p-values and probabilities					
20. Students are assigned project that require them to analyze data					
21. Students present their findings to the class					

For questions #22–#25, if you teach several classes, write in the course number for the class if you answer varies per course. *For example, I teach math 104, 204, and 360 and I would answer question 20 in the following way:*

Indicate the type of data that you believe helps students learn statistics best.

a. *All constructed data*
b. *Mostly constructed data*
c. *Equal amounts of constructed and real data (360)*
d. *Mostly real data (204)*
e. *All real data (104)*

22. Indicate the type of data that you use helps students learn statistics best.

a. All constructed data
b. Mostly constructed data
c. Equal amounts of constructed and real data
d. Mostly real data
e. All real data

23. Indicate the method of computing numerical solutions to problems that you believe helps students learn statistics best.

a. All solutions computed by hand
b. Most solutions computed by hand
c. Equal amounts of computing solutions by hand using technology tools
d. Most solutions computed using technology tools
e. All solutions computed using technology tools

24. What computing resources do you have available in the classrooms you teach in?

a. I teach in a computer lab
b. Students are required to bring their laptops
c. Students can bring their laptops but it is not required
d. Students have calculators
e. Students have no technology in the classroom

25. What software do students use in your class?

a. SPSS
b. STATA
c. SAS
d. MATLAB
e. R
f. Python
g. Point and click statistical software (e.g., Minitab, Tinkerplots, Fathom, JMP, StatCrunch)
h. Excel
i. TI Calculators
j. None
k. Other, please specify: _____

Overarching Information

26. Are you happy with the way statistics is currently taught at LMU?

   a. Yes
   b. No
   c. Somewhat

27. If not, how do you believe the teaching of statistics at LMU needs to change?

28. Identify any institutional constraints that keep you from being satisfied with the manner in which statistics is currently taught at LMU (Check all that apply).

   a. Hiring issues (not enough, not the correct people, etc.)
   b. Institution does not understand current statistical needs and what statistics is
   c. Institution does not provide access to the required technology, materials, etc. to teach statistics well
   d. Student body does not have an interest in statistics or does not have the intellectual ability for statistics
   e. Other, please specify: _____

29. Identify any constraints that keep you from making any changes that you would like to implement to improve your statistics courses (Check all that apply).

   a. Your personal time constraints
   b. Institutional constraints (e.g., choice of textbook, class size, mandated curriculum, etc.) If you circle this, please specify: _____
   c. Technology constraints (e.g., lack of computer lab, cost of software)
   d. Characteristics of students (ability, interest, etc.)
   e. Other, please specify: _____

30. Does your discipline organization (e.g., Mathematics Association of America, APA) have guidelines or other resources for statistics teaching and learning?

   a. Yes
   b. No
   c. I don't know

31. If yes, do you follow those? How closely? What is your involvement with them? How do you use the guidelines directly in your classroom?

Appendix B

Course Descriptions for Statistics and Data-Related Courses at LMU

Course	Course description
AFAM 208/2243 African American Studies Research Methods (4 semester hours)	An introduction to the methods used to acquire and disseminate knowledge about the systems and policies that affect the African American Community.
ECON 230/2300 Introductory Statistics (4 semester hours)	An introduction to the modern methods of analyzing sample data. Topics include descriptive statistics, probability theory, binomial and normal distributions, estimation, hypothesis testing, and simple regression analysis. Recommended: MATH 112 (Business Calculus) or MATH 131 (Calculus I) or concurrent enrollment.
ECON 235/2350 Accelerated Intro Statistics (4 semester hours)	Accelerated introduction to statistics with applications to economics. Topics include descriptive statistics, probability theory, binomial and normal distributions, estimation, hypothesis testing, and regression analysis. This course also involves exercises in applying theoretical concepts to real-world empirical problems, e.g., for policy analysis. This course can be taken in lieu of ECON 2300 and substitutes for ECON 2300 wherever it is stated as a prerequisite. Recommended: MATH 112 (Business Calculus) or MATH 131 (Calculus I) or concurrent enrollment.
ECON 330/3300 Econometrics (4 semester hours)	Analysis of the linear regression model and its practical applications in economics, finance, marketing, and other areas of business. Material covered will be the two variable model, hypothesis testing, forecasting, functional forms of regression models, regression using dummy explanatory variables, multiple regression, autocorrelation, heteroscedasticity and multicollinearity. Emphasis is placed on the application of the techniques covered in the course to the solution of real-world problems. Prerequisites: ECON 1050 (Introductory Economics) and ECON 2300 (Introductory Statistics).
ECON 532/5320 Advanced Econometrics (4 semester hours)	This branch of economics uses mathematical and statistical tools to analyze economic phenomena. Mathematical formulation, establishment of hypotheses, model construction, data collection, and statistical estimation and inference. Prerequisite: ECON 5300 (Math for Economics) or MATH 250 (Linear Algebra).
POLS 2100 Empirical Approaches (4 semester hours)	This course introduces the assumptions of the scientific approach in the study of politics, the process of concept formation, and research design. Includes data analysis laboratory sessions. Prerequisite: POLS 1200 (U.S. Politics), POLS 1400 (Comparative Politics), or POLS 1600 (International Relations).
POLS 4100 Advanced Empirical Approaches (4 semester hours)	This course introduces advanced techniques in the empirical study of politics. Prerequisite: POLS 2100 (Empirical Approaches).

Course	Course description
POLS 440/4250 Public Policy Analysis (4 semester hours)	An examination of the processes by which public policy is formulated, implemented, and evaluated. Emphasis will be placed on policy planning and evaluation competencies.
PSYC 241/243/2001 Statistical Methods for Psychology (4 semester hours)	Statistical concepts and methods related to psychological testing and research, including measures of central tendency, variability, hypothesis testing, analysis of variance, correlation, regression, non-parametric tests, and use of statistical software programs. Prerequisite: Grade of C (2.0) or higher in PSYC 1000 (General Psychology).
PSYC 261/2002 Research Methods (4 semester hours)	Introduces the basic principles of common psychology research methods and designs. Provides students with fundamental background for planning, conducting, and critiquing research in psychology. Emphasizes scientific writing, including APA style, and data interpretation using descriptive and inferential statistics. Topics include: hypothesis testing, searching scientific literature, variables and measurement, reliability and validity, and ethics in the conduct of empirical research. Prerequisites: Grade of C (2.0) or higher in PSYC 1000 (General Psychology) and PSYC 2001 (Statistical Methods for Psychology).
PSYC 4045 Advanced Research Methods (4 semester hours)	Building on lower division statistics and research methods Psychology requirements, this course further investigates research designs and statistical analyses psychologists use to understand, predict, and influence human behavior. Students will conduct an empirical research study in a topical area of psychology and present results in APA-style written, oral, and/or poster presentations. Highly recommended for students considering graduate study in psychology. Prerequisites: Grade of C (2.0) or higher in PSYC 1000 (General Psychology), PSYC 2001 (Statistical Methods for Psychology), PSYC 2002 (Research Methods), and PSYC 2003 (Brain and Behavior).
SOCL 210/2100 Quantitative Research Methods (4 semester hours)	An introduction to basic inferential and descriptive statistics commonly used in the social sciences. Among the topics covered are: table construction; central tendency; variation; probability, sampling distributions, and the normal curve; hypothesis testing; and measures of association.
AIMS 3710 Database Management Systems (3 semester hours)	This course is intended for the student who wishes to become more proficient at developing and managing database applications. It is designed to provide an introduction to the conceptual foundations underlying database management systems, with an emphasis on its applications in business and organizations. The course begins with an introduction to the fundamental principles of database design-from data modeling to the actual implementation of a business application. Particular emphasis will be placed on the careful planning and analysis of business needs, which will lead to the appropriate development of an Entity-Relationship Model. Using these principles, each student will design and implement a database application using Access. This part of the course will employ lectures describing database theory, as well as hands-on tutorials demonstrating database concepts using Access. The second part of the course will further investigate the relational model, which is the basis for the most popular DBMS products on the marketplace today (i.e., Oracle, SQL Server, MS Access, Sybase). Topics to be studied include relational algebra, Structured Query Language (SQL), and maintaining data integrity in a relational design. In addition, important managerial concerns will be covered including database administration and the management of multi-user databases. Prerequisites: ACCT 3140 (Accounting Information Systems) or AIMS 2710 (Management Information Systems); BADM 1030 (Business Perspectives –Information Technology in Organizations) with a grade of C (2.0) or better
AIMS 4760 Analytics & Business Intelligence (3 semester hours)	The course introduces students to the scientific process of understanding, displaying, and transforming data into insight in order to help managerial decision makers do their job effectively and make better, more informed, decisions. The nature of data/information used in the decision making process and the role of information technology in that process is discussed. The course focuses on data preparation and transformation, descriptive and predictive analytics, data mining, and data visualization and dashboards. An overview of prescriptive analytics is presented as well as the role of business analytics in the context of business intelligence. Hands-on learning is an important feature of the course. For each topic, a case analysis will require the use of Excel and/or other specialized data mining and analytics software to reinforce the underlying theoretical concepts. Prerequisites: ACCT 3140 (Accounting Information Systems) or AIMS 2710 (Management Information Systems); AIMS 3770 (Production Operations Analysis).
AIMS 4798 SS: Intro to Big Data & Data Science (3 semester hours)	*Taught as a special topics courses so description varies depending on instructor
BADM 3010 Analytical Concepts/Methods for Business (3 semester hours)	The course is meant as an introductory basis for students seeking a minor in Business Administration. Topics will include fractions, decimals vs. percentages, Excel spreadsheets vs. calculators, future and present value, interest rates, compounding, credit terms in personal finance, probabilities, weights, mean or expected value, statistics, variance, standard deviation, sample vs. population statistics, correlation and covariance, and linear estimations.

Course	Course description
EDES 5407 Research Methods/Early Childhood Assessment (3 semester hours)	Students will gain knowledge in planning and conducting research as well as further advance their written communication skills. Students will critically evaluate published research. Students will use and apply various observation techniques such as narrative records, running records, time sampling, and event sampling to the understanding of child behavior and developmental processes. Students will demonstrate data analysis skills. Students will gain knowledge in the assessment of both typical and atypical development. Students will explore issues of professional ethics related to working with parents and teachers when special needs in children are identified and require intervention. Field experience will be required.
BIOL 367 Biological Databases (3 semester hours)	Interdisciplinary course at the interface between biology and computer science focusing on how biological information is encoded in the genome of a cell and represented as data in a database. Biological concepts include DNA structure and function, the central dogma of molecular biology, and regulation of gene expression. Computer science concepts and skills include command line interaction, the structure and functions of a database, and the management of data ranging from individual files to a full relational database management system. Emphasis on science and engineering best practices, such as maintaining journals and notebooks, managing files and code, and critically evaluating scientific and technical information. Course culminates with team projects to create new gene databases.
BIOL 398 SS: Biostatistical Analysis (3 semester hours)	This course is designed to teach students how to analyze and interpret quantitative data. It will demonstrate practical applications in addition to basic theory. The emphasis will be how and when to use (or not use) each method. We will apply these methods to actual data from biological, ecological, and public health applications. This course will also include the use of computer programs (SPSS, R) to apply tests to datasets. By the end of the course the student should have a good understanding of basic parametric and nonparametric statistical methods, their assumptions and applications, and how and when to apply them to different types of data. Prerequisite: One year calculus and one year biology or consent of instructor.
CMSI 367 Biological Databases (3 semester hours)	The representation, storage, and transformation of biological data. The central dogma of molecular biology, the genetic code, the Human Genome Project, sequence databases, formats and conversion, searching and regular expressions. Document and relational databases in biology. Students build and potentially release an open source gene database for a new species at the end of the course.
CMSI 485 Artificial Intelligence (3 semester hours)	Introduction to the fundamental concepts underlying the implementation of human-level intelligence (and beyond) in computer systems. Topics include agent architectures, problem-solving methods, knowledge representation, symbolic reasoning, Bayesian inference, Markov and Hidden Markov models, utilities and preferences, fundamentals of machine and deep learning, and contemporary applications. Prerequisites: CMSI 281 (Data Structures) and CMSI 282 (Algorithms).
MATH 104 Elementary Statistics (3 semester hours)	Introduction to methods of inferential statistics, histograms, elementary probability, and random variables and distributions.
MATH 204 Applied Statistics (3 semester hours)	This course provides an introduction to statistics emphasizing data analysis and applications to life sciences. Topics include: descriptive statistics, elementary probability, various discrete and continuous distributions, confidence intervals and hypothesis tests for means and proportions, correlation and linear regression, as well as analysis of variance. This course will also include the use of computer programs to analyze datasets. Prerequisite: MATH 122 (Calculus for the Life Sciences I) or MATH 131 (Calculus I).
MATH 360 Intro Probability/Stats (3 semester hours)	Descriptive statistics, probability, discrete and continuous random variables, limit theorems, sampling distributions, estimations of parameters, nonparametric methods, hypothesis testing, linear regression. Prerequisite: MATH 123 (Calculus for the Life Sciences II) or MATH 132 (Calculus II).
MATH 398 Biostatistical Analysis (3 semester hours)	This course is designed to teach students how to analyze and interpret quantitative data. It will demonstrate practical applications in addition to basic theory. The emphasis will be how and when to use (or not use) each method. We will apply these methods to actual data from biological, ecological, and public health applications. This course will also include the use of computer programs (SPSS, R) to apply tests to datasets. By the end of the course the student should have a good understanding of basic parametric and nonparametric statistical methods, their assumptions and applications, and how and when to apply them to different types of data. Prerequisite: One year calculus and one year biology or consent of instructor.
MATH 560 Adv Topics in Probability & Stats (3 semester hours)	Material to be covered will be determined by the instructor. Consult with the instructor for the specific topics in probability and statistics that will be covered in any given semester. Prerequisites: MATH 234 (Calculus III) and MATH 360 (Intro to Probability and Statistics).
MATH 598 SS: Stats & Modeling for Teachers (3 semester hours)	This course is specifically designed for secondary teachers. It covers descriptive statistics, basic data visualizations, data seen in the news, study design, introduction to inference, and liner regression

Course	Course description
MATH 599 Probability & Statistics Lab (3 semester hours)	This course is meant to accompany MATH 360 as an additional lab. The lab focuses on programing in R and formulating probabilistic models to be simulated in R.
MECH 598 SS: Statistical Design of Experiments (3 semester hours)	Basic concepts of probability and statistics that are fundamental to Design of Experiments (DOE). The key topics will include sampling, hypothesis testing [t-statistic, f-statistic, analysis of variance (ANOVA), p-value], experimental design matrices, full-factorial and fractional factorial designs, normal probability plots, factor level interactions, regression modeling. Case studies and a design project will be used to illustrate the methodology. Prerequisite: Undergraduate Calculus I and II.

Some courses listed in the table have multiple course numbers due to cross-listings or changing courses numbers during the study period.

Appendix C

Community Survey. The following questions were asked in the external community survey.

For each statement below, please mark whether you agree, disagree, or are neutral that the statement described a data analysis skill that you believe a college graduate in today's society should have

	Agree	Neutral	Disagree
1. Student should be able to formulate questions about univariate data, collect univariate data/consider univariate data, analyze univariate data, and interpret results			
2. Student should be able to understand, calculate and interpret descriptive measures for quantitative and categorical variables to describe characteristics of the data			
3. Students should be able to create and interpret basic data visualizations for quantitative and categorical variables			
4. Students should be able to understand, carry out and interpret basic inferential procedures for one or two samples.			
5. Students should be able to understand, carry out and interpret statistical procedures for predicting future data (predictive inference)			
6. Students should carry out and communicate results from an extensive data-driven project that is related to a real-life problem (extensive means that the project takes more than two weeks to complete and is worth at least 25% of the final grade)			
7. Students should be able to communicate their analyses and the interpretations of their results in a manner that is appropriate to their discipline in the context of the data			
8. Students should be able to select appropriate methods for data analysis and explain limitations of their analyses and interpretations			
9. Student should be able to formulate questions about multivariate data, collect multivariate data/consider multivariate data, analyzing multivariate data, and interpret results			
10. Student should be able to use current statistical software, or statistical packages appropriate to the discipline and context beyond basic Excel or a calculator			
11. Student should be able to write a program (using a programing language) to analyses data			
12. Students should study at least one type of advanced data-analytic methods such as (not limited to) generalized linear models, Bayesian analysis, advanced probability theory and stochastic processes, non-linear models, machine learning, advanced study-design, big data analysis, econometrics, or statistical computing			

What is your primary professional role? If other, please specify.

- College/University Faculty
- College/University Administrator
- Industry Scientist/Researcher/Consultant
- Government Scientist/Researcher/Consultant
- Other

What field of study do you feel most close to in your work? This could be the field of your highest degree or the field that relates most closely to your current work. If other, please specify.

- Statistics
- Data Science
- Mathematics
- Computer Science
- Sociology
- Psychology
- Business
- Education
- Biology
- Chemistry
- Economics
- Political Science
- Other

Funding

References

American Economic Association. "Recommended Mathematical Training to Prepare for Graduate School in Economics." Available at *https://www.aeaweb.org/resources/students/grad-prep/math-training*

American Psychological Association. "Guidelines for the Undergraduate Psychology Major." Available at *https://www.apa.org/ed/precollege/psn/2014/09/career-counselor*

American Sociological Association. "The Sociology Major in the Changing Landscape of Higher Education: Curriculum, Careers, and Online Learning," A Report of the ASA Task Force on Liberal Learning and the Sociology Major. Available at *http://www.asanet.org/sites/default/files/asa-booklet-2017.pdf*

American Statistical Association (2016), "Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report". Available at *https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf*

American Statistical Association, Undergraduate Guidelines Workgroup (2014). *Curriculum Guidelines for Undergraduate Programs in Statistical Science*, Available at *http://www.amstat.org/education/curriculumguidelines.cfm*.

ASA/MAA Joint Committee on Undergraduate Statistics. (2014), *Qualifications for Teaching an Introductory Statistics Course*. Available at *https://www.maa.org/sites/default/files/JointASA-MAAStatementQualificationsTeachCollegeIntroStatCourse.pdf*

Bargagliotti, A., Rovetti, R., Larson, S., Zachariah, T., and Fitzpatrick, B. (2020), "Are universities adequately preparing students for our data-driven world?," Under submission.

Colby, A., Ehrlich, T., Sullivan, W. M., and Dolle, J. R. (2011), *Rethinking Undergraduate Business Education: Liberal Learning for the Profession* (Vol. 20). New Jersey: John Wiley & Sons.

Committee on the Undergraduate Program in Mathematics (CUPM) (2004), *Undergraduate Programs and Courses in the Mathematical Sciences: CUPM Curriculum Guide 2004*. Mathematical Association of America. Available at www.maa.org/cupm/.

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

Eagan, K., Stolzenberg, E. B., Ramirez, J. J., Aragon, M. C., Suchard, M. R., and Hurtado, S. (2014), *The American Freshman: National Norms Fall 2014*. Los Angeles, CA: Higher Education Research Institute, UCLA, 36.

Everson, M., Mocko, M., Carver, R., Gabrosek, J., Horton, N., Lock, R., Rossman, A., Holmes Rowell, G., Velleman, P., Witmer, J., and Wood, B. (2016), *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report*. Available at *https://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf*

Estes, A. (2011), *Peter Theil Bets $2.4 Million Against the Value of a College Degree*. Available at *https://www.theatlantic.com/national/archive/2011/05/peter-thiel-bets-million-against-value-college-degree/351076/*

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., and Scheaffer, R. (2007), *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A pre-K–12 Curriculum Framework*. Alexandria, VA: American Statistical Association. Available at *www.amstat.org*.

Gould, R., Peck, R., Hanson, J., Horton, N., Kotz, B., Kubo, K., Malyn-Smith, J., Rudis, M., Thompson, B., Ward, M., and Wong, R. (2018), *The Two-Year College Data Science Summit*. Available at *https://www.amstat.org/asa/files/pdfs/2018TYCDS-Final-Report.pdf*.

Hardin, J., Hoerl, R., Norton, N., and Nolan, D. (2015), Data Science in Statistics Curricula: Preparing Students to "Think with Data," working paper. Available at *http://arxiv.org/ftp/arxiv/papers/1410/1410.3127.pdf*

Holdren, J. P., and Lander, E. (2012), "Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics." Available at *https://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf*.

Horton, N., and Hardin, J. (2015), "Teaching the Next Generation of Statistics Students to 'think with Data': Special Issue on Statistics and the Undergraduate Curriculum," *The American Statistician*, 69, 259–265.

Mathematical Association of American (2015), *2015 CUPM Curriculum Guide to Majors in the Mathematical Sciences*. Washington, DC: MAA. Available at *http://www.maa.org/sites/default/files/pdf/CUPM/pdf/CUPMguiderint.pdf*

Mathematical Association of American (CUPM -MAA's Committee on the Undergraduate Program in Mathematics). Available at *https://www.maa.org/programs/faculty-and-departments/curriculum-department-guidelines-recommendations/cupm/2015-cupm-curriculum-guide*

National Academies of Sciences, Engineering, and Medicine (2018a), *Data Science for Undergraduates: Opportunities and Options*. Washington, DC: The National Academies Press. Available at *https://doi.org/10.17226/25104*.

——— (2018b), *Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim Report*. Washington, DC: The National Academies Press. *https://doi.org/10.17226/24886*.

Nolan, D., and Temple Lang, D. (2010), "Computing in the Statistics Curricula," *The American Statistician*, 64, 97–107.

——— (2015), "Data Science in R: A Case Studies Approval to Computational Reason and Problem Solving," *The American Statistician*, forthcoming November issue.

Pierson, S. (2018). "Highlights from 2017 Degree Release: Bachelor's Numbers Close in on Master's." *AMSTAT News*. August.

Ridgway, J. (2016), "Implications of the Data Revolution for Statistics Education," *International Statistical Review*, 84, 528–549.

Wilder, E. I. (2010), "A Qualitative Assessment of Efforts to Integrate Data Analysis throughout the Sociology Curriculum Feedback from Students, Faculty, and Alumni," *Teaching Sociology*, 38, 226–246.

Zieffler, A., Park, J., Garfield, J., Delmas, R., and Bjornsdottir, A. (2012), "The Statistics Teaching Inventory: A Survey on Statistics Teachers' Classroom Practices and Beliefs," *Journal of Statistics Education*, 20.

Zorn, P., Bailer, J., Braddy, L., Carpenter, J., Jaco, W., and Turner, P. (2014), "The INGenIOuS Project: Mathematics, Statistics, and Preparing the 21st Century." Available at *http://www.maa.org/sites/default/files/pdf/ingenious/INGenIOuS-report.pdf*.