Master of Science in Data Science Program
2016-2017 Assessment Report

1 Identifying Information

Name of Program: Data Science (formerly Analytics)
Type of Program: Graduate Program
College of Arts and Sciences Division: Sciences
Submitter: David Uminsky, Program Director
Submitter Email Address: duminsky@usfca.edu

Additional feedback concerning this report can be directed to Kirsten Keihl (Program Manager), Mindi Mysliwiec (Director of Operations, Data Institute), or Jeff Hamrick (Assessment Liaison).

2 Program Vision and Mission

Program Vision. Our vision is to become a national leader in training the next generation of technically-competent and career-ready professionals who are fully engaged in, and continuously advance, the data science revolution in the Bay Area and beyond.

Program Mission. The mission of our program is to produce graduates who possess a theoretical and practical understanding of many classical and modern statistical modeling and machine learning techniques; who use contemporary programming languages and technologies to scrape, clean, organize, query, summarize, visualize, and model large volumes and varieties of data; and who use their knowledge and skills to successfully solve real-world data-driven business problems and to communicate those solutions effectively.

Notes. The MSDS program's vision and mission statements were ratified by its faculty during a January 2016 vote over email.

3 Program Goals

Faculty members affiliated with the MSDS program have identified a number of long-term goals for the program. These goals include, but are are broader than, student satisfaction or student learning outcomes. Some of these goals are tangible; others are not.

1. **Student satisfaction:** In exit surveys, a super-majority of graduates will indicate a high degree of satisfaction with quality of the faculty's teaching, the practicum component of the MSDS curriculum, and the program's value proposition.
2. **A strong value proposition:** In a recent exit survey, a graduate of the program indicated that "It was worth it, quitting my job and making [this] investment." Another said "I learned things that I didn't know about or how to do. I got a job in the career

that I wanted." Still another said "What I know now compared to a year ago—the change is insane." We seek to transform people with a wide variety of backgrounds and aptitudes into highly-skilled and in-demand data scientists within 12 months.

3. **Practitioner engagement and recognition:** Each year, the program will meaningfully engage with data-driven Bay Area companies. This engagement will extend to a number of companies greater than to one-third of the program's current enrollment. It will also be measured by practitioner participation in the Data Science Seminar Series, which is jointly sponsored by the MSDS program and the Data Institute.
4. **Increased national and international visibility:** Over time, the program's faculty will increasingly be well-known for making theoretical and applied contributions to fields such as computer science, statistics, and business. The program will routinely be identified with other top-tier analytics programs (e.g., Columbia University, Georgia Tech, Northwestern University, Georgetown University, Johns Hopkins University, New York University, etc.).
5. **A community of scholars:** Faculty members will actively and regularly pursue scholarly activities, including the publication of high quality research papers and books, grant-writing and submission, and dissemination of their work at important conferences and colloquia in their respective fields. The faculty will share their interdisciplinary expertise with one another and build a culture of research collaboration with each other, as well as with external collaborators.
6. **A community of teachers:** Faculty members will establish a culture of helping one another to become more successful teachers (e.g., open and constructive and voluntary discussion of BLUE results, voluntary classroom visits to secure feedback and suggestions for improvement, pedagogical innovations, etc.).
7. **A culture of service to the program:** Faculty members will participate in recruitment and yield events, conduct technical admissions interviews, support the Data Institute, and vigorously engage in deliberations related to the program's curriculum.

Notes: These program goals were ratified by the program's faculty during a May 2016 vote over email.

4 Program Learning Outcomes (PLOs)

Upon successfully completing the Master of Science in Data Science (MSDS) program, our graduates will:

(1) Possess a theoretical understanding of classical statistical models (e.g., generalized linear models, linear time series models, etc.), as well as the ability to apply those models effectively;

(2) Possess a theoretical understanding of machine learning techniques (e.g., random forests, neutral networks, naive Bayes, k-means, etc.), as well as the ability to apply those techniques effectively;

(3) Effectively use modern programming languages (e.g., R, Python, SQL, etc.) and technologies (AWS, Hive, Spark, Hadoop, etc.) to scrape, clean, organize, query, summarize,

visualize, and model large volumes and varieties of data;

(4) Be prepared for careers as data scientists by solving real-world data-driven business problems with other data scientists; and

(5) Develop professional communication skills (e.g., presentations, interviews, email etiquette, etc.), and begin integrating with the Bay Area data science community.

Notes: Please refer to the attached curriculum map relating program learning outcomes to courses in the MSDS curriculum. The latest version of the MSDS program learning outcomes was ratified by a vote of faculty members over email on March 23, 2018.

5 Academic Program Review

The Master of Science in Data Science program was founded in academic year 2012-2013 and has, therefore, not yet had its first formal program review. Its first academic program review is currently planned for academic year 2019-2020, with the external review team visiting in fall 2019. The program's self-study will be completed in the latter half of academic year 2018-2019 and be available between three and six months prior to the on-campus visit.

6 2016-2017 Assessment Plan

For academic year 2016-2017, the program's faculty decided to review the statistical component of the curriculum and address the following sorts of questions: Is there evidence that our students walk away from the program having satisfied our first PLO? In aggregate, do the learning outcomes of the program's statistics courses (MSDS 502, MSDS 504, MSDS 602, MSDS 604, MSDS 623) reflect the body of knowledge, skills, and experiences that a data scientist at the beginning of his or her career should possess? Are the courses organized in the right way with respect to the rest of the curriculum? Are the courses scheduled in the right modules?

6.1 The Methodologies.

The statistics faculty met throughout the fall semester and early spring semester to plan, review, and edit the instrument devised to support a review of the statistical component of the curriculum. This instrument is located in appendix A. It was administered to graduating students towards the end of the program, i.e., in the final module of the cohort year. The statistics faculty wrote approximately two dozen possible quiz questions from four major subareas of the statistics curriculum: basic probability and statistics, linear regression analysis, time series analysis, and computational statistics. The quiz was administered through a specially-developed Canvas module. Canvas supports random selection of quiz questions from subgroups identified by an instructor. It can also automatically grade any set of multiple-choice questions. Three questions from each topic were chosen at random, generating random quizzes with a total of 12 questions. Students were given one hour to

complete the examination, or about five minutes per question. This length of time is relatively standard for questions that might involve some number of selected computations, e.g., the 1/P examination for the Society of Actuaries uses questions requiring about six minutes, on average. A total of 54 students took the statistics assessment examination. Sixty students graduated from the program's fifth cohort; the participation rate of 90% was very high.

We also undertook a second investigation as part of the 2016-2017 assessment plan. As noted in Appendix D, the MSDS faculty added a supplemental section to the annual technology survey (which is typically conducted by Nick Ross). In this supplemental section, we ask our students questions about the statistical concepts, methods, or models they see during their practicum experiences. Students were asked to identify how important it is that they master each concept, method, or model. Additionally, students were asked to identify whether or not they had seen, or had not seen, the concept, method, or model during their practicum experience.

6.2 The Results.

Faculty members met for two hours on August 14, 2017 to discuss the results of the direct assessment instrument, which are summarized in Appendix B. In studying the question-by-question results of the assessment instrument, faculty made the following observations.

(a) Linear regression question #7 was probably not optimally phrased. Using the phrase "Statistical inference on the coefficients of a linear regression model..." would probably help more of our students to answer this question correctly.
(b) For computational statistics question #3: Faculty feel like this question should have a correct answer rate of 100% since the logit function is central to logistic regression which, in turn, is fairly central to predictive modeling and classification. It perhaps remains unclear who is responsible for introducing the logistic regression model in the program. James Wilson is putting it ("owning it") in his version of Linear Regression Analysis (MSAN 601), so this problem is being addressed.
(c) For linear regression question #6: The question should be rephrased so that it says "Which of the following **is** false...?" It is possible that the current phrasing created some confusion among students.
(d) For basic probability and statistics question #3: More students should have gotten this answer correct. The primary responsible instructor (Jeff Hamrick) indicated that he will make stronger efforts to emphasize the formula for $\text{Var}(X+Y)$, or the variance of the sum of random variables, and would also provide more opportunities for students to do associated calculations. This calculation is already reinforced in the time series class. It clearly gets used throughout the program and is clearly of fundamental importance to a practicing data scientist.
(e) For basic probability and statistics question #6: The faculty suspect that, in general, students have a hard time remembering the formal statements of theorems (and certainly, their proofs). The Central Limit Theorem is perhaps the most important theorem in all of probability and statistics and is worthy of exceptional treatment. To improve on a question like this one, our students could perhaps use additional practice writing down formal models (including all assumptions). This might reinforce better

recall in other contexts (e.g., theorem-writing) in which writing down assumptions is important. Hamrick agreed to take up this suggestion concerning key theorems in MSDS 504, and the other faculty agreed to take up the suggestion concerning key models in their respective courses.

(f) For computational statistics question #2: The MSDS program should probably emphasize multiple comparisons procedures throughout the curriculum more. Faculty agree that in each (or most) of the statistics classes, this will be emphasized more. Jeff Hamrick will attempt to carve out a piece of a lecture in MSDS 504 to introduce the most elementary aspects of this issue, e.g., the Bonferroni correction.

(g) In general, the faculty felt like the computational statistics questions ended up being a little jargon-laden. James Wilson agreed to adapt some of these questions for the future so that they focus more on high-level concepts.

(h) The faculty agreed that the students did exceptionally well on time series question #1 and time series question #2.

(i) The faculty agreed that time series question #4 is a hard and compound question that is also a calculation. However, students get extensive experience with answering this sort of question in the time series course, MSDS 604. Ideally, students would have performed better on this question.

(j) For time series question #5, faculty agreed that phrasing (or terminology) might have gotten in the way. The students probably understand what needs to be done but might—by the end of the program—have forgotten appropriate language/terminology learned since the time series course (which is in the second half of the fall semester).

(k) For time series question #7, the students picked — all of them — two options that are related to multivariate time series analysis. However, a minority of students ignored the hint about working with exogenous variables, which caused them to be "tricked" by the first answer choice.

(l) Basic probability and statistics questions #2 and #4 provide strong evidence that students that students really do learn (and have a strong foundation in) basic probability theory by the end of the program. The binomial random variable is emphasized throughout the program and is well-understood and easily identified by students.

(m) Additionally, Bayesian reasoning is strongly emphasized throughout many areas of the program—and so it is not surprising that students would do well on a Bayes' Theorem question. This is true for the fifth basic probability and statistics question, which is clearly re-emphasized and underlined in the Design of Experiments elective.

The faculty also had some overall suggestions concerning the direct assessment instrument. First, the faculty should place a greater emphasis on reviewing concepts throughout the program. Doing so would require better coordination among statistics instructors, as well as between faculty who primarily teach statistics courses and faculty who primarily teach computer science courses. Reviewing concepts is not a bad thing in a program that is so intense — though, because the program has such a short (12 months) and intense format, some faculty feel like we should minimize the duplication of topics.

Second, the statistics faculty agreed that there should be a separate annual meeting with the computer science instructors to define and to better understand all of the relationships between the concepts in "our" courses and in the machine learning courses. A graph show-

ing the dependencies between the course learning outcomes could be developed, and better clarify who is supposed to be teaching which material.

Finally, the statistics faculty agreed that this test, in and of itself, should not excessively inform current conversations about possible revisions to the statistical component of the MSDS curriculum. It is only one tool, and other tools (e.g., the exit survey, our conversations with students, the lived experiences of our faculty members as they teach, formal and informal data collection efforts over longer periods of time, etc.) should receive weight as well.

With respect to the supplemental section of the practicum technology survey, faculty noticed that the feedback mostly validates what we are currently doing—throughout the curriculum, and certainly in the statistical component of the curriculum. The faculty did notice, however, a theme: "We saw this in the program's curriculum, but it isn't as useful as you might think in a real-world context." Items related to this theme, and the faculty's responses, were as follows:

(a) **Linear discriminant analysis and quadratic discriminant analysis.** These two topics are classical among statisticians interested in classification problems. However, they are heavily parametric, i.e., they are sensitive to a fairly unrealistic assumption of a multivariate Gaussian distribution (or many multivariate Gaussian distributions) of features. We agreed to drop these topics altogether from the course now called Computational Statistics (MSDS 628).

(b) **Properties of the multivariate Gaussian distribution.** For a variety of reasons, the statisticians cannot simply excise this topic from the curriculum. It is important to understand, for example, how certain statistical modeling techniques or machine learning techniques do, or do not, depend upon an approximately multivariate Gaussian distribution. Instead, the faculty opted to de-emphasize this topic in the course now called Computational Statistics (MSDS 628) and to move this topic (in large part, though not completely) into the probability and statistics boot camp (MSDS 504).

(c) **Spectral clustering and the Davis-Kahan Theorem.** This topic is an unfortunate case in which the companies we work with on practicum projects are not aware of the power, or the utility, or certain more cutting-edge clustering techniques. The faculty agreed that it would be appropriate to drop this topic from the course now called Computational Statistics (MSDS 628) and move the topic to the elective course in network analytics. The faculty also agreed to undertake greater efforts to educate practicum partner companies about these, and related, techniques.

(d) **ARCH and GARCH models.** Nobody sees these models in practicum projects—basically ever. The primary instructor responsible for time series analysis (MSDS 604) has agreed to drop these models as a course learning outcome in the 2017-2018 academic year. He will replace this topic with a brief foray into time series clusting techniques, which should prove more relevant and more cutting-edge.

Faculty also made the following observations about the results of the supplemental section of the practicum technology survey:

(a) Student see penalized regression enough in the practicum that it needs to be more heavily emphasized in the linear regression course. **Action item:** James Wilson will

make this change to Linear Regression Analysis (MSDS 601) in the 2017-2018 academic year.

(b) A few years ago, Poisson regression was added to Linear Regression Analysis (MSDS 601) but students seem to think it is not important and we can find no evidence that it is ever used in a practicum project. **Action item:** James Wilson agreed to drop this topic from the very crowded set of course learning outcomes for Linear Regression Analysis (MSDS 601).

(c) The faculty acknowledge that principal component analysis (PCA) does not get used very much in a practicum context. However, there was strong discomfort with the notion of eliminating such a classical and central dimension reduction technique. As with logistic regression, PCA is not a topic that has a clearly identifiable home in the MSDS curriculum. **Action item:** Clarify the primary course that "owns" this topic but do not delete or de-emphasize it.

(d) Topics that are centrally featured in our curriculum, that we teach effectively, and that are easy to find in many practicum projects: issues of robustness, outliers, how to wrangle data, the development of hypotheses for testing and for investigation, exploratory data analysis, model construction and selection, regression, time series analysis, measurements of goodness of fit, construction of data sets for training and for validation, etc. **Action item:** We should continue to build upon these strengths.

(e) With some topics in this survey, there was probably conflation with instructor teaching success. We have had difficulty consistently delivering our introductory machine learning course well across different instructors. **Action item:** Conduct an annual curriculum meeting with the computer science instructors, and consider developing a map of the interactions between the course learning outcomes for the statistically-oriented courses and the courses more typically taught by our computer scientists.

(f) Suggestions for this survey when we implement it again in three years: ask about more general concepts (e.g., Bayesian modeling) and less about specific concepts (e.g., the prior distribution). Students have more pathways into the survey item if they do not have to worry about what very specific names (or jargon) means.

The complete results of the annual technology survey, as well as the supplemental section included in academic year 2016-2017, are available upon request. In summary, very few students are deploying techniques or tools in the practicum projects that we do not cover in the curriculum. The open-ended questions would seem to confirm these results. There are obviously some instances in which the practicum requires a student to make an extra effort to pick up a statistical technique or statistical concept that was omitted from the curriculum. But this phenomenon is rare.

6.3 Closing the Loop.

In addition to the commitments ("action steps") made by faculty members and recorded in section 6.2, the statistics faculty endorsed the transformation of an old course, Multivariate Statistical Analysis (MSDS 623), into a new course, Computational Statistics (MSDS 628). Sample syllabi for both courses are included in appendix E. This course was approved by the MSDS faculty by a vote over email on November 3, 2016. The New/Change in Course Proposal process was completed on February 17, 20187. The new course description reads as

follows: "This course covers advanced statistical and computational techniques for estimation, imputation, simulation, and hypothesis testing. Topics include numerical integration, multivariate analysis, Bayesian inference, Markov Chain Monte Carlo, the E-M algorithm, graphical models, and multiple testing."

As a result of the unfolding annual assessment exercise for academic year 2016-2017, the faculty essentially felt that MSDS 623 was destined for a sufficient number of changes (e.g., moving the topic of spectral clustering to a network analytics elective, eliminating the topics of linear and quadratic discriminant analysis, etc.) to justify the introduction of a new course. This course, MSDS 628, focuses on advanced statistical computational techniques related to predictive modeling, approximation, estimation, imputation, sampling, and multiple testing from both the Bayesian and frequentist perspectives. Some elements of MSDS 623 made their way into MSDS 628 (e.g., multivariate probability distributions and densities). Topics from other courses (e.g., Bayesian inference) were carefully selected for repetition and reinforcement in MSDS 628. Still other topics (e.g., Gibbs sampling and the Metropolis-Hastings algorithm) are effectively new to the program.

This new course was informally taught for the first time in academic year 2016-2017 and officially incorporated into the curriculum for academic year 2017-2018. It is an example of how the MSDS faculty use an ongoing process of assessment to continuously transform and to steadily improve the MSDS curriculum.

A Direct Assessment Instrument

Appendix A contains both the instructions for, and questions that were sampled at random (in the stratified sense) by Canvas to produce, the direct assessment instrument.

Instructions. Momentarily, you will be given 12 multiple choice questions. These questions cover topics from the statistical component of the Master of Science in Analytics curriculum, i.e., basic probability and statistics, regression, time series, and computational statistics. We believe that each question will take an average of 5 minutes to complete. Hence, you should expect to spend one hour on this assessment examination.

You may not consult with any outside materials (e.g., notes, web sites, R documentation, etc.) as you work through this examination. You also may not consult with any other individuals, including the faculty and/or staff members proctoring the examination. You have been given scratch paper and you are free to use it.

Notation. Please be advised of the following conventions as you take this examination:

1. We use ε to denote a residual, typically in a regression or time series model.
2. We use e to denote a *fitted* residual, typically from a regression or time series model.
3. We use Greek letters (sometimes with subscripts) such as β, α, θ_1, ϕ_1, etc., to denote parameters that control or characterize regression or time series models.
4. For regression models, we use b_1 to describe the fitted coefficient associated to the coefficient β_1 in a regression model.
5. We use $\widehat{Y}$ to describe the predicted values of Y from a regression model.
6. We use notation like $\{X_t\}$ to describe a time series.

Basic Probability and Statistics.

1. Suppose $P(A) = 0.60$, $P(B) = 0.47$ and $P(A \cap B) = 0.19$. Calculate $P(A|B^C)$.

(a) 0.36
(b) 0.41
(c) 0.68
(d) 0.77
(e) 0.87

2. Suppose a store has 100 light bulbs in stock. Assume 40 light bulbs are from Distributor A and the remainder of the light bulbs are from Distributor B. Assume 5.0% of the light bulbs from Distributor A are defective and 10.0% are defective from Distributor B.

If a consumer purchases 3 light bulbs, what is the probability that exactly 2 of the light bulbs are defective? Choose the correct expression.

(a) $\binom{2}{3}0.08^3(1-0.08)^2$
(b) $0.08^2(1-0.08)$
(c) $\binom{3}{2}0.08^2(1-0.08)$
(d) 0.08
(e) $0.08(1-0.08)$

3. You are given two random variables X and Y such that $\mathbb{E}(X) = 0$, $\mathbb{E}(Y) = -1$, $\text{Var}(X) = 1$, $\text{Var}(Y) = 4$, and $\text{Var}(X + Y) = 9$. Find a value of a parameter a such that $X + Y$ and $X + aY$ are uncorrelated.

(a) $-\frac{1}{3}$
(b) $-\frac{1}{2}$
(c) 0
(d) $\frac{1}{3}$
(e) $\frac{1}{6}$

4. Suppose that you know that exactly 10% of all emails sent to you are spam. Moreover, you know that 80% of the emails that you have received in the past that were not spam contained the word "analytics" and that 40% of the emails that you received that were spam contained that word. Suppose that you receive a new email that contains the word "analytics." What is the probability that the email is spam?

(a) 0.40
(b) 0.10
(c) 0.053
(d) 0.056

5. Which of the following statements about the type 2 error rate is false?

(a) It is the probability that you reject the null hypothesis when the null hypothesis is, in fact, true.
(b) It is the probability that you do not reject the null hypothesis when the null hypothesis is, in fact, false.
(c) It is equal to one minus the power of the statistical test.
(d) It decreases as the sample size increases.
(e) It decreases as the effect size increases.

6. The classical statement of the Central Limit Theorem requires which of the following assumptions about the members of a sequence of random variables $X_1, X_2, ..., X_n$?

(a) The X_i must be uncorrelated with one another.
(b) Each X_i must be such that $\mathbb{E}[X^4] < \infty$.
(c) $\mathbb{E}[X_i] = 0$ for every i.
(d) The means of the various X_i must be finite but need not be equal to one another.
(e) X_i must have the same distribution as X_j for every i and j.

Regression.

1. When working with a classical multiple linear regression model, we often interpret the fitted coefficient b_j in the following fashion: "Given a one-unit increase in the variable X_j, the dependent variable will increase by b_j on average and all else being equal." Which problem with a classical multiple linear regression model will prove the most problematic for the "all else being equal" part of this interpretation?

(a) serially correlated errors
(b) non-normal errors
(c) multicollinearity
(d) endogeneity
(e) heteroscedasticity

2. In a multiple linear regression model, the coefficient β_1 represents the

(a) expected value of the response variable Y when the explanatory variable $X_1 = 0$
(b) average change in the response variable Y per unit change in the explanatory variable X_1
(c) predicted value of the response variable Y
(d) variation around the regression line
(e) none of the above

3. In multiple regression, a diagnostic measure used to study the possible existence of heteroskedasticity is

(a) Variance inflation factors (VIFs)
(b) Mallow's C_p
(c) standardized residuals
(d) the Durbin-Watson test statistic
(e) all of the above

4. In linear regression analysis, a residual plot is

(a) $\widehat{Y}$ against X
(b) $\widehat{Y}$ against Y
(c) $(Y - \widehat{Y})$ against X
(d) $(Y - \widehat{Y})$ against Y
(e) none of the above

5. In a linear regression equation

(a) the Y_i are random
(b) the X_i are random
(c) the ε_i are random
(d) (a) and (c)
(e) none of the above

6. Which of the following are false with respect to the properties of the fitted regression line?

(a) $\sum e_i = 0$
(b) $\sum e_i^2 = \sigma^2$
(c) $s_{xe} = s_{\hat{Y}e} = 0$
(d) $(\overline{X}, \overline{Y})$ is always on the ordinary least squares (OLS) regression line

(e) none of the above

7. For statistical inference to be valid for a linear regression model, what is the most important assumption that we need to impose on the error term ε?

(a) The expected value of ε_i must be zero for every i.
(b) The variance of ε_i must be σ^2 for every i.
(c) The ε_i must be independent of one another.
(d) The ε_i must be normally distributed.
(e) The ε_i must be non-degenerate.

Time Series.

1. Which of the following is an AR(1) time series, i.e., an autoregressive time series of order 1?

(a) $X_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$
(b) $X_t = \mu + \phi_1 X_{t-1} + \varepsilon_t$
(c) $X_t - \phi_1 X_{t-1} = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$
(d) $\varepsilon_t = \sigma_t Z_t$, with $Z_t \sim N(0, \sigma_t^2)$ and $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$
(e) $\varepsilon_t = \sigma_t Z_t$, with $Z_t \sim N(0, \sigma_t^2)$ and $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

2. The time series $\{X_t\}$ is said to be *weakly stationary* if

(a) $\mathbb{E}(X_t) = 0$ for all t
(b) $\mathbb{E}(X_t)$ is independent of t
(c) $\mathrm{Cov}(X_t, X_{t+h})$ is independent of t for all h
(d) (a) and (c)
(e) (b) and (c)

3. Unit root tests are used to check whether an observed time series is stationary. One such unit root test is the:

(a) Bartlett test
(b) Augmented Dickey-Fuller test
(c) Ljung-Box test
(d) Shapiro-Wilk test
(e) Wilcoxon signed-rank test

4. Consider the following ARMA(1,1) process: $X_t + 0.6X_{t-1} = \varepsilon_t + 1.2\varepsilon_{t-1}$, $\{\varepsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$. By finding the roots of the autoregressive and moving average generating functions, determine which of the following statements is true:

(a) $\{X_t\}$ is both stationary and invertible
(b) $\{X_t\}$ is not stationary but is invertible
(c) $\{X_t\}$ is stationary but not invertible
(d) $\{X_t\}$ is neither stationary nor invertible

5. Suppose that the time series $\{X_t\}$ exhibits monthly seasonality. Which of the following modeling approaches would be *most* suitable?

(a) ARIMA
(b) Double exponential smoothing
(c) Triple exponential smoothing
(d) Vector autoregression
(e) GARCH

6. Suppose you have have fit the following two models to an observed time series: SARIMA$(2,1,1)\times(1,1,1)_{12}$ and SARIMA$(1,1,0)\times(1,1,1)_{12}$. If you would like to choose the model that fits the data best but also protect yourself from over-fitting, the goodness-of-fit metric that is *most* appropriate is:

(a) $\widehat{\sigma}^2$
(b) R^2
(c) Adjusted R^2 (or R_a^2)
(d) AIC
(e) maximized log-likelihood

7. Suppose that a farmer's crop yield is recorded monthly, and predicting future month's yields is of interest. Crop yield is dependent on a myriad of factors such as temperature, for example, whose influence may be accounted for by a multivariate time series approach. Suppose that you wish to model crop yield treating temperature as an *exogenous* variable. The most appropriate multivariate time series approach is

(a) Vector autoregression
(b) ARIMAX
(c) Multiple regression
(d) Convolutional neural networks
(e) All of the above

Computational Statistics.

1. Suppose that you possess a data set of size n with p features $X_1, ..., X_p$, all of which are quantitative (or continuous) variables. These features have a $p \times p$ variance-covariance matrix called Σ. Which of the following statement is true?

(a) Changing the means of the features may change the first principal component of Σ.
(b) The sum of the k largest eigenvalues of Σ is equal to the sum of the k largest variances among the $X_1, ..., X_p$.
(c) The principal components of Σ are unique.
(d) The trace of Σ is equal to the sum of the eigenvalues of Σ.
(e) Any given principal component of Σ can be expressed as a vector, each component of which is definitely positive.

2. Let $H_1, \ldots, H_m$ be a family of m null hypotheses for which the p-values $p_1, \ldots, p_m$ are calculated. Suppose that $1 \leq m_o \leq m$ of the null hypotheses are actually true and that the p-values are mutually independent. For all $j = 1, \ldots, m$, you decide to reject hypothesis j if $p_j \leq \alpha/m$. What is the familywise error rate of these decisions (i.e., the probability of rejecting at least one true hypothesis)?

(a) α
(b) m_0/m
(c) $m_0\alpha/m$
(d) $m_o\alpha$
(e) none of the above

3. Let $p \in (0, 1)$. What is the logit(p)?

(a) $\log(p)$
(b) $\log\left(\dfrac{p}{1-p}\right)$
(c) $\dfrac{\log(p)}{\log(1-p)}$
(d) $\dfrac{e^p}{1+e^p}$
(e) none of the above

4. Suppose that Y and θ are jointly distributed random variables, where $Y \mid \theta \sim N(\theta, \theta^2)$ and $\theta \sim N(0, \sigma^2)$ and $\sigma^2 > 0$ is fixed. What is Var(Y)? (**Hint:** Use the Law of Total Variance.)

(a) σ^2
(b) θ^2
(c) $2\sigma^2$
(d) $2\theta^2$
(e) none of the above

5. Which of the following is (or are) true about the EM (i.e., expectation maximization) algorithm when estimating an unknown parameter from a latent variable model?

(a) It converges to a global maximum for any likelihood function.
(b) It converges to a global maximum for exponential family likelihood functions.
(c) It requires knowing the first derivative of the likelihood function with respect to the unknown parameter.
(d) Both (a) and (b).
(e) All of (a), (b), and (c).

6. Suppose that we would like to find the parameter θ that maximizes the following log posterior density:

$$L(\theta) = \log(p(\theta \mid y))$$

Computationally, one way to do this is using the following algorithm:

1. Choose a starting value θ^0 and threshold $\epsilon > 0$
2. For $t = 2, 3, \ldots,$
   (a) **Compute** the first and second derivatives $L'(\theta^{t-1})$, $L''(\theta^{t-1})$
   (b) **Update** θ^t with
   $$\theta^t = \theta^{t-1} - [L''(\theta^{t-1})]^{-1} L'(\theta^{t-1})$$
   (c) **Stop** when $|\theta^t - \theta^{t-1}| < \epsilon$

What is the name of this algorithm, and what key approximation is the basis of the **Update** rule?

(a) quasi-Newton, linear Taylor series approximation
(b) quasi-Newton, quadratic Taylor series approximation
(c) Newton-Raphson, linear Taylor series approximation
(d) Newton-Raphson, quadratic Taylor series approximation
(e) none of the above

B Scoring the Direct Assessment Instrument

The results of the direct assessment of statistical knowledge referenced in section 6.1 and appendix A are shown in tabular format here. The acronym "BPS" means "basic probability and statistics." The four topic areas that were examined were: basic probability and statistics, linear regression analysis, time series analysis, and computational statistics.

Question ID	A	B	C	D	E	No Answer	Correct Choice	Percent Correct
BPS 1	2	7	1	18	0	1	D	62.07%
BPS 2	1	1	21	0	0	1	C	87.50%
BPS 3	4	11	10	3	2	1	B	35.48%
BPS 4	1	0	24	2	0	0	C	88.89%
BPS 5	22	4	0	0	0	1	A	81.48%
BPS 6	6	1	1	6	12	0	E	46.15%
Regression 1	2	0	23	4	1	1	C	74.19%
Regression 2	0	16	0	1	3	0	E	80.00%
Regression 3	3	1	9	3	4	0	C	45.00%
Regression 4	0	1	12	12	0	1	C	46.15%
Regression 5	2	1	7	13	0	1	D	54.17%
Regression 6	2	6	6	0	6	0	B	30.00%
Regression 7	8	0	9	6	1	0	D	25.00%
Time Series 1	7	12	0	1	0	0	B	60.00%
Time Series 2	0	0	2	3	21	1	E	77.78%
Time Series 3	3	2	17	10	1	0	B	56.67%
Time Series 4	4	6	9	1	0	0	C	45.00%
Time Series 5	7	7	8	1	0	0	C	33.33%
Time Series 6	0	0	2	15	3	0	D	75.00%
Time Series 7	5	16	0	0	1	0	B	69.57%
Comp. Statistics 1	0	7	6	16	2	1	D	50.00%
Comp. Statistics 2	8	1	13	3	6	1	C	40.63%
Comp. Statistics 3	1	21	1	7	0	0	B	70.00%
Comp. Statistics 4	4	3	21	5	3	0	C	58.33%
Comp. Statistics 5	1	5	15	2	11	1	B	14.29%

C Exit Survey from Cohort Five

Since its inception, the MSAN program has issued exit surveys to all of its graduates. These surveys have played a critical role in informing faculty decisions about the curriculum. The program's most recent cohort of graduates were asked an open-ended question: "What class or topic would you remove from the curriculum?" The tabulation of the results is as follows:

Course or Topic	Count
Web, or Google, Analytics (MSDS 695)	23
Application Development (MSDS 698)	12
Communications for Analytics (MSDS 610)	10
Interviewing Skills (MSDS 696)	10
Business Strategies for Big Data (MSDS 603)	9
Data Vizualization (MSDS 694)	3
Introduction to Machine Learning (MSDS 621)	1
Computational Statistics (MSDS 628)	1
Bayesian statistics (a topic)	1

The standout courses here are MSDS 695, MSDS 698, MSDS 610, MSDS 696, and MSDS 603. Changes to each of these courses rolled out contemporaneously, i.e., in academic year 2016-2017, or were subsequently discussed and voted upon by the MSDS faculty during academic year 2017-2018 for implementation in that academic year or in academic year 2018-2019. For example, the faculty voted to retire the Web Analytics (MSDS 695) course during academic year 2017-2018. Started in academic year 2018-2019, elements of MDS 698 and MSDS 603 courses are being combined into MSDS 603, with MSDS 698 itself eliminated. Interviewing Skills (MSDS 696) is being eliminated in lieu of a series of mandatory interview- and resume-preparation workshops.

Notably, during academic year 2016-2017, James Wilson effectively converted (vis-a-vis ad hoc substitution of topics) the old Multivariate Statistics course (MSDS 623) into the new Computational Statistics course (MSDS 628). In prior instances of the exit survey, students had placed the Multivariate Statistics course onto this list. However, only one memebr of the program's fifth cohort suggested eliminating the new course (MSDS 628). We consider this an indication that the new course was was well-received.

The exit survey was used to inform other important decisions as well. Approximately 75% of respondents indicated that the Design of Experiments course, a special topics (elective) course offered at the very end of the program by Nathaniel Stevens, was so excellent—and potentially useful in their careers as data scientists—that they asked the faculty to consider making it a mandatory component of the curriculum. In academic year 2017-2018, and effective for academic year 2018-2019, the MSDS faculty agreed to this concept.

Results of the exit survey are available in spreadsheet format upon request.

D Technology Survey

Nicholas Ross conducts an annual survey concerning the technologies used by our students during their practicum experiences. The results of this survey are used to have a general discussion (typically, at the faculty's annual winter meeting) about the state of the program's curriculum with respect to cutting-edge programming languages and analytical technologies. An aside: the current year's results suggest that while students may argue about depth of coverage of various technologies in the program's curriculum, the breadth of what they see in the classroom is what they are seeing in the field (i.e., in their practicum assignments). In academic year 2016-2017, we supplemented this survey with questions about the statistical concepts and methodologies seen by students during their practicum experiences. The presentation that Nick Ross made to the faculty is included below.

Technology Survey 2017

Caveats / Introduction / Takeaways

- Students were asked to choose among the following options:
  - I used the technology frequently in my practicum
  - I used the technology occasionally in my practicum
  - The technology was used by the company, but I did not use it for practicum
  - N/A
- Charts are grouped with “I used” and “Any use”
- Responses: 47/60. (Last year: 27/42) (E.g. 1% ~ 2 Students)
- Changes from last year = not perfectly comparable
- For next year: Fix where Hive and Tensorflow appear in the survey. Students were confused.
- Key take-away: A lot of noise in the #’s, but there do not seem to be any unknown holes in the curriculum (IMO). While students may argue about depth of coverage, the breadth of what we teach is what they are seeing

SQL

Others Mentioned: Hive (4), Impala (2), Big Query (2) and MSSQL (2)

	2017			2016			Change	
	I Used	Any Use		I Used	Any Use		I Used	Any Use
RedShift	15%	23%	RedShift	33%	19%		-18%	5%
Postgres	34%	45%	Postgres	44%	41%		-10%	4%
MySQL	13%	32%	MySQL	37%	19%		-24%	13%
Spark SQL	11%	15%	Spark SQL	22%	11%		-12%	4%
MongoDB	4%	15%	MongoDB	15%	7%		-11%	7%
Other NoSQL/S	19%	36%						

Deep Learning

Others Mentioned: TensorFlow (2) and Keras (2)

	2017			2016			Change	
	I Used	Any Use		I Used	Any Use		I Used	Any Use
LSTM	2%	2%	LSTM	11%	4%		-9%	-2%
Theano	6%	11%	Theano	15%	7%		-8%	3%
Other Deep Le:	9%	17%	Other Deep Le:	15%	7%		-6%	10%

Statistical Software

	2017			2016			Change	
	I Used	Any Use		I Used	Any Use		I Used	Any Use
R	60%	79%	R	78%	70%		-18%	8%
Spark R	0%	6%	Spark R	15%	4%		-15%	3%
Sci-kit (Python	72%	81%	Sci-kit (Python	70%	67%		2%	14%
NumPy (Pythor	81%	89%	NumPy (Pythor	74%	74%		7%	15%
Pandas (Pythc	81%	87%						
Other Statistic	23%	32%						

General Computing

Others Mentioned: Ruby (3), Java (2), Golang (2)

	2017			2016			Change	
	I Used	Any Use		I Used	Any Use		I Used	Any Use
Python	85%	94%	Python	81%	78%		4%	16%
Scala	2%	17%	Scala	15%	4%		-13%	13%
Java	0%	38%	Java	19%	4%		-19%	35%
C/C++	2%	23%	C/C++	15%	4%		-13%	20%
Shell	64%	74%						
Javascript	13%	40%						
Other General I	4%	19%						

Visualization

Others: Plotly (3), Chartio (2)

	2017			2016			Change	
	I Used	Any Use		I Used	Any Use		I Used	Any Use
Tableau	13%	45%	Tableau	41%	11%		-28%	34%
Jupyter	70%	72%	Jupyter	67%	67%		4%	6%
D3	6%	21%	D3	37%	26%		-31%	-5%
Shiny (R Librar	11%	21%	Shiny (R Librar	15%	7%		-4%	14%
matplotlib	77%	79%	matplotlib	63%	63%		14%	16%
ggplot	47%	66%						
Other Visualiza	11%	21%						

Other

	2017			2016			Change	
	I Used	Any Use		I Used	Any Use		I Used	Any Use
AWS: EC2	32%	53%	AWS: EC2	44%	26%		-13%	27%
AWS: S3 stora	21%	53%	AWS: S3 stora	48%	26%		-27%	27%
AWS: Elastic M	6%	21%	AWS: Elastic M	19%	7%		-12%	14%
Google Cloud C	2%	9%						
Google Cloud S	4%	15%						
Hadoop	11%	26%	Hadoop	26%	11%		-15%	14%
Hive	9%	19%	Hive	37%	22%		-29%	-3%
Pig	0%	4%	Pig	22%	11%		-22%	-7%
Spark	13%	28%	Spark	30%	19%		-17%	9%
Apache Parque	11%	11%	Apache Parque	15%	7%		-4%	3%
TensorFlow	2%	15%	TensorFlow	15%	4%		-13%	11%
CloudML	0%	0%	CloudML	11%	4%		-11%	-4%

E Syllabi for MSDS 623 and MSDS 628

As mentioned in section 6.3, the faculty voted to retire the Multivariate Statistics (MSDS 623) course and replace it with a related, but in many ways different, course: Computational Statistics (MSDS 628). Sample syllabi for both the old version of the course and the new alternative are provided on the following pages.

MSAN 623 - 01 Multivariate Statistical Analysis Spring 2016

Class Time and Location: T, TH 9:00 - 11:00 AM; Howard 529

Recitation Time and Location: T 11:00 AM - 1:00 PM; Howard 529

Instructor: James D. Wilson **Office:** 203B Harney Science Building **Email:** jdwilson4@usfca.edu

Office Hours: Tuesdays by appointment (in Howard on the 5th floor)

Grader: Me

Course Website: Canvas website

Textbooks: Much of this course will be based on the following text book:

- *An Introduction to Multivariate Statistical Analysis, 3rd ed.* by T.W. Anderson

Learning Outcomes: In this class we will cover topics generally about the statistical inference of multivariate data. Specifically, we will cover:

- Multivariate probability distributions and densities
- The multivariate normal distribution
- Measures of dependence:
  - correlation coefficient
  - partial correlation
  - multiple correlation coefficient
  - canonical correlations
  - covariance
- The generalized T^2 statistic
- The distribution of the covariance matrix estimates
- Multivariate ananlysis of variance (MANOVA)
- Hypothesis testing involving:
  - general linear hypotheses
  - independence of variables
  - equality of mean vectors and covariance matrices
- Graphical models and patterns of dependence
- Multivariate time series (if time permits)

What you should bring to Class:

A pencil or pen, paper, and a sunny disposition :)

Homework

- Throughout the course, I will assign small homework assignments that "fill in the gaps" of some of the theory presented in class. In total, these will be worth 60% of your final grade.

Exams

- There will be a final exam, which will be worth 40% of your grade. Missed exams will receive a grade of zero.
- Exams are closed-book, closed-note unless otherwise specified.

Exam Date: Tuesday, May 17th: 1:30 - 3:30 PM

Grading Rubric: I will round your final grade to the nearest hundredth and assign grades according to

A+	97 - 100
A	93 - 96
A-	90 - 92
B+	87 - 89
B	83 - 86
B-	80 - 82
C+	77 - 79
C	73 - 76
C-	70 - 72
D+	67 - 69
D	63 - 66
D-	60 - 62
F	< 60

Academic Integrity

As a Jesuit institution committed to cura personalis - the care and education of the whole person - USF has an obligation to embody and foster the values of honesty and integrity. USF upholds the standards of honesty and integrity from all members of the academic community. All students are expected to know and adhere to the University's Honor Code. You can find the full text of the code online at `www.usfca.edu/academic_integrity`. The policy covers:

- Plagiarism: intentionally or unintentionally representing the words or ideas of another person as your own; failure to properly cite references; manufacturing references.
- Working with another person when independent work is required.
- Submission of the same paper in more than one course without the specific permission of each instructor.
- Submitting a paper written by another person or obtained from the internet.
- The penalties for violation of the policy may include a failing grade on the assignment, a failing grade in the course, and/or a referral to the Academic Integrity Committee.

Students with Disabilities

If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF Student Disability Services (SDS) at 415 422-2613 within the first week of class, or immediately upon onset of disability, to speak with a disability specialist. If you are determined eligible for reasonable accommodations, please meet with your disability specialist so they can arrange to have your accommodation letter sent to me, and we will discuss your needs for this course. For more information, please visit: http://www.usfca.edu/sds or call (415) 422-2613.

Behavioral Expectations

All students are expected to behave in accordance with the Student Conduct Code and other University policies (see http://www.usfca.edu/fogcutter/). Open discussion and disagreement is encouraged when done respectfully and in the spirit of academic discourse. There are also a variety of behaviors that, while not against a specific University policy, may create disruption in this course. Students whose behavior is disruptive or who fail to comply with the instructor may be dismissed from the class for the remainder of the class period and may need to meet with the instructor or Dean prior to returning to the next class period. If necessary, referrals may also be made to the Student Conduct process for violations of the Student Conduct Code.

Learning & Writing Center

The Learning & Writing Center provides assistance to all USF students in pursuit of academic success. Peer tutors provide regular review and practice of course materials in the subjects of Math, Science, Business, Economics, Nursing and Languages. `https://tutortrac.usfca.edu`. Students may also take advantage of writing support provided by Rhetoric and Language Department instructors and academic study skills support provided by Learning Center professional staff. For more information about these services contact the Learning & Writing Center at (415) 422-6713, email: `lwc@usfca.edu` or stop by our office in Cowell 215. Information can also be found on our website at `www.usfca.edu/lwc`.

Counseling and Psychological Services

Our diverse staff offers brief individual, couple, and group counseling to student members of our community. CAPS services are confidential and free of charge. Call 415-422-6352 for an initial consultation appointment. Having a crisis at 3 AM? We are still here for you. Telephone consultation through CAPS After Hours is available between the hours of 5:00 PM to 8:30 AM; call the above number and press 2.

Confidentiality, Mandatory Reporting, and Sexual Assault

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I also have a mandatory reporting responsibility related to my role as a faculty member. I am required to share information regarding sexual misconduct or information about a crime that may have occurred on USFs campus with the University. Here are other resources:

- To report any sexual misconduct, students may visit Anna Bartkowski (UC 5th floor) or see many other options by visiting our website: `www.usfca.edu/student_life/safer`.
- Students may speak to someone confidentially, or report a sexual assault confidentially by contacting Counseling and Psychological Services at 415-422-6352.
- To find out more about reporting a sexual assault at USF, visit USF's Callisto website at: `www.usfca.callistocampus.org`.
- For an off-campus resource, contact San Francisco Women Against Rape (SFWAR) (415) 647-7273 (`www.sfwar.org`).

Student Accounts - Last day to withdraw with tuition reversal

Students who wish to have the tuition charges reversed on their student account should withdraw from the course(s) by the end of the business day on the last day to withdraw with tuition credit (census date) for the applicable course(s) in which the student is enrolled. Please note that the last day to withdraw with tuition credit may vary by course. The last day to withdraw with tuition credit (census date) listed in the Academic Calendar is applicable only to courses which meet for the standard 15-week semester. To find what the last day to withdraw with tuition credit is for a specific course, please visit the Online Class Schedule at `www.usfca.edu/schedules`.

MSAN 628 Computational Statistics Spring 2017

Class Time and Location:

- Section 1: W 11:05 - 12:55; F 10:00 - 12:00 Howard 529
- Section 2: W 3:15 - 5:05; F 2:30 - 4:30 Howard 529

Instructor: James D. Wilson **Office:** 203B Harney Science Building **Email:** jdwilson4@usfca.edu

Office Hours: Wednesdays 2:00 - 3:00, Fridays: 9:00 - 10:00 (in Howard on the 5th floor)

Grader: Me

Course Website: Canvas website

Textbooks: This course will include a survey of important statistical computational methods. As such, lectures will be drawn from many sources. No books are required, but some recommended books include

- *Bayesian Data Analysis, 3rd Ed.* by Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin
- *Statistical Inference, 2nd Ed.* by Casella and Berger
- *Machine Learning: A Probabilistic Perspective* by Kevin Murphy

Course Overview: This course provides an in-depth look into advanced statistical computational techniques for estimation, imputation, simulation, and hypothesis testing. These skills are incredibly important for modern practitioners in data science and analytics. By the end of the semester, students will be comfortable with aspects of wrangling and modeling data using state-of-the-art techniques in estimation, hypothesis testing, and imputation. Students will implement all computational techniques using the R programming language, and will have familiarity with the following key aspects of statistical analysis:

- multivariate probability distributions and densities
- the likelihood paradigm for statistical inference and prediction
- introduction to Bayesian inference
- computational techniques for estimation, simulation, and approximation including Markov Chain Monte Carlo simulation via Gibbs and Metropolis-Hastings, numerical integration, importance sampling, and rejection sampling
- imputation models and techniques including the expectation-maximization algorithm and its extensions, variational inference, and expectation propogation
- graphical models including hidden Markov models and Bayesian networks
- multiple hypothesis testing techniques including Bonferroni correction and the Benjamini-Hochberg step-up procedure

Course Learning Outcomes: By the end of the course, students will be able to use the R programming language to

- proficiently wrangle and model data with missing values
- estimate point and maximum a posteriori estimators for Bayesian and frequentist predictive models
- approximate otherwise computationally intractable functions using numerical integration and resampling methods
- communicate results using R
- fit a wide array of predictive models using advanced computational techniques

What you should bring to Class:

A pencil or pen, paper, a computer, any notes that have been posted to Canvas, and a sunny disposition :)

Attendance: Attendance is required every day and will be recorded and worth 20% of your final grade. It is your responsibility to catch up on any lecture material, homework, or programming lesson that you miss due to an absence.

Topics Covered: We will cover the following topics in accordance with the schedule below.

Week	Topic	Description
1	**Multivariate Models**	review of multivariate probability distributions, Bayesian inference, multi-parameter models and interpretation, the multivariate normal, conditional expectation
2	**Point Estimation and Simulation**	maximum likelihood estimators, maximum a posteriori (MAP) estimators, Markov Chains, an overview of simulation
3	**Estimation via Bayesian Computation**	importance and rejection sampling, Metropolis algorithm, Metropolis-Hastings the Gibbs Sampler, numerical integration
4	**Models for Missing Data and E-M**	joint probability models for missing data at random and completely at random, expectation maximization algorithm and its extensions for multiple imputation
5	**Other Methods for Missing Data**	k-Nearest Neighbors, variational inference, expectation propogation, and multiple imputation
6	**Multiple Testing**	the multiple testing problem, Bonferroni adjustment, false discovery rate, Benjamini-Hochberg step-up procedure, hidden Markov models
7	**Bayes Nets**	models, dependence, and estimation

Assessment:

- **Attendance** (20%): Attendance will be recorded every class. You will lose 2% of this grade for every class that you miss, unless previously discussed.
- **Assignments** (50%): For each assignment, you will be required to upload a .pdf file to the Canvas site that contains clear demonstrations of R code, any analyses, and any visualization used to answer the questions on the assignment. Assignments will provide case studies that emphasize the methodology learned in class that week. These must be submitted before the deadline set on Canvas, else you will receive **10** points off every day that it is late.
- **Final Exam** (30%): The final exam is a cumulative exam that will assess the concepts learned throughout this course. This will be an in-class written exam given on the scheduled final exam date provided by the University of San Francisco final exam schedule.

Grading Procedure: At the end of the semester, your grade will be calculated according to the following rubric:

A	90 - 100
B	75 - 89
C	60 - 74
F	≤ 59

There will be no curve implemented in this course. Late assignments will not be accepted and will automatically receive a grade of 0.

Important Dates:

- Wednesday, March 23rd - First day of class!
- Monday, April 10th - Last day to withdraw
- Friday, April 14th - Easter Holiday (**no class**)
- Wednesday, May 10th - Last day of class!
- Final Exam: this will be scheduled according to university standards and will be held in class. Scheduling information is available at `https://myusf.usfca.edu/onestop/registration/class-schedule-final-exams`.

Academic Integrity

As a Jesuit institution committed to cura personalis - the care and education of the whole person - USF has an obligation to embody and foster the values of honesty and integrity. USF upholds the standards of honesty and integrity from all members of the academic community. All students are expected to know and adhere to the University's Honor Code. You can find the full text of the code online at `www.usfca.edu/academic_integrity`. The policy covers:

- Plagiarism: intentionally or unintentionally representing the words or ideas of another person as your own; failure to properly cite references; manufacturing references.
- Working with another person when independent work is required.
- Submission of the same paper in more than one course without the specific permission of each instructor.
- Submitting a paper written by another person or obtained from the internet.
- The penalties for violation of the policy may include a failing grade on the assignment, a failing grade in the course, and/or a referral to the Academic Integrity Committee.

Students with Disabilities

If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF Student Disability Services (SDS) at 415 422-2613 within the first week of class, or immediately upon onset of disability, to speak with a disability specialist. If you are determined eligible for reasonable accommodations, please meet with your disability specialist so they can arrange to have your accommodation letter sent to me, and we will discuss your needs for this course. For more information, please visit: http://www.usfca.edu/sds or call (415) 422-2613.

Behavioral Expectations

All students are expected to behave in accordance with the Student Conduct Code and other University policies (see http://www.usfca.edu/fogcutter/). Open discussion and disagreement is encouraged when done respectfully and in the spirit of academic discourse. There are also a variety of behaviors that, while not against a specific University policy, may create disruption in this course. Students whose behavior is disruptive or who fail to comply with the instructor may be dismissed from the class for the remainder of the class period and may need to meet with the instructor or Dean prior to returning to the next class period. If necessary, referrals may also be made to the Student Conduct process for violations of the Student Conduct Code.

Learning & Writing Center

The Learning & Writing Center provides assistance to all USF students in pursuit of academic success. Peer tutors provide regular review and practice of course materials in the subjects of Math, Science, Business, Economics, Nursing and Languages. `https://tutortrac.usfca.edu`. Students may also take advantage of writing support provided by Rhetoric and Language Department instructors and academic study skills support provided by Learning Center professional staff. For more information about these services contact the Learning & Writing Center at (415) 422-6713, email: `lwc@usfca.edu` or stop by our office in Cowell 215. Information can also be found on our website at `www.usfca.edu/lwc`.

Counseling and Psychological Services

Our diverse staff offers brief individual, couple, and group counseling to student members of our community. CAPS services are confidential and free of charge. Call 415-422-6352 for an initial consultation appointment. Having a crisis at 3 AM? We are still here for you. Telephone consultation through CAPS After Hours is available between the hours of 5:00 PM to 8:30 AM; call the above number and press 2.

Confidentiality, Mandatory Reporting, and Sexual Assault

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I also have a mandatory reporting responsibility related to my role as a faculty member. I am required to share information regarding sexual misconduct or information about a crime that may have occurred on USFs campus with the University. Here are other resources:

- To report any sexual misconduct, students may visit Anna Bartkowski (UC 5th floor) or see many other options by visiting our website: `www.usfca.edu/student_life/safer`.
- Students may speak to someone confidentially, or report a sexual assault confidentially by contacting Counseling and Psychological Services at 415-422-6352.
- To find out more about reporting a sexual assault at USF, visit USF's Callisto website at: `www.usfca.callistocampus.org`.
- For an off-campus resource, contact San Francisco Women Against Rape (SFWAR) (415) 647-7273 (`www.sfwar.org`).

Student Accounts - Last day to withdraw with tuition reversal

Students who wish to have the tuition charges reversed on their student account should withdraw from the course(s) by the end of the business day on the last day to withdraw with tuition credit (census date) for the applicable course(s) in which the student is enrolled. Please note that the last day to withdraw with tuition credit may vary by course. The last day to withdraw with tuition credit (census date) listed in the Academic Calendar is applicable only to courses which meet for the standard 15-week semester. To find what the last day to withdraw with tuition credit is for a specific course, please visit the Online Class Schedule at `www.usfca.edu/schedules`.

F Minutes of MSAN Faculty Meetings

The MSAN faculty typically meets between two and four times per year. At nearly every one of these meetings, the program's curriculum is discussed. We attach, as evidence of our ongoing and vigorous discussions of the curriculum, minutes from our meetings during academic year 2016-2017.

Faculty Meeting Minutes 9/26/16

Agenda:
1. Review last year - so please brainstorm the highs and lows from your POV
2. Curriculum discussion - including proposed changes and tweaks, I expect this to be the longest part.
3. Briefly review and vote on program goals.
4. Practicum update
5. DI update
5. Your suggested items!

1. Review of Last Year
How did the discussion model go?
-Some instructors lectured through their discussion time
-Some did quizzes
-Some did lab time

Question: What is the correct format for quizzes?
Beginning of class? End of class? 30 minutes before class? Do students all take quizzes together at the same time?

1.A. Side topic: Admissions
- What are we looking for in a candidate? There is a massive difference between the students with CS background students and those with none. We gave the hardest technical interview this year, but still accepted people with weaker coding skills.
- Are we accepting more people with better communication skills vs others with better experience in prereqs but weaker comm skills? Note: The majority of interviews are with applicants that are strong in 1 or 2 areas, but not all 3.
-We don't want the reputation that 1 in 10 people aren't making it through the bootcamp, though we make it clear that some people will not pass.

Identifying admissions priorities: This coming year (2017/2018), we are changing the boot camp so that students do not get to choose their classes. Students must complete Python projects before they arrive here to adequately prepare.
-Should we build a pre-bootcamp course? This year there was an avoidance of Jeff's class and lost of people filled into Terence's class.
-Do we give a qualifier exam?
-We are aiming for a 40/60 split international/domestic, gender balanced group, also increase in underrepresented groups. Those are our main goals with admissions.
-This group is less cohesive than last year. Lots of people have not said anything in class since day 1.

1.B. Side Topic 2: Grading Standards:
-Largely speaking, overall GPA is a B+ average. This has stayed mainly the same year to year. This can help set grading standards for this program.
-Questions: Can people still fail? What is the appropriate use of F?
-C is used more as a strong warning. F is used for clear cases where the person has stopped attending, stopped turning in homework, committed fraud, cheated, plagiarized etc. See what Terence sent out regarding grading:

BOOTCAMPS

This class is pass/fail and we expect most people to pass, but those getting below 80% raw average score are in the danger zone.

GRADED CLASSES

Grading standards. I consider an A grade to be above and beyond what most students have achieved. A B grade is an average grade for a student or what you could call "competence" in a business setting. A C grade means that you either did not or could not put forth the effort to achieve competence. Below C implies you did very little work or had great difficulty with the class compared to other students.

OPTIONS

We could add more specific language:

A, B, and C correspond to the usual percentages of 90-100, 80-89, 70-79, respectively. +/- modifiers are used near the thresholds.

2. Curriculum Discussion

We need to submit changes much earlier than we have in the past for them to be implemented next year. We will vote on these in the next few months.

Multivariate Statistics
-What are we doing with this course?
-Statisticians have agreed to assess the usefulness of this course, as well as other statistics courses.
-We want to survey students on what topics were most useful.
-Sometimes topics are repeated in courses redundantly.
-Rethinking the statistics curriculum as a whole.
-Is this the right time in the semester to teach a theoretical class like msa?
-Does this class move to an elective? And design of experiments move to a required course?
-A new course called applied computation and statistics?
-There is no time for extra topics in regression.
-A meeting is set to discuss these topics in October.

Electives
-Do we move to a full elective sequence in the summer?
-Web analytics has seen less interest over the years.
-Do we include more business content here?
-We would be doubling the faculty cost by changing a one unit class to a two unit class. We are wiping out the profitability by doing this.
-We will offer 4 elective choices this year.

-The original design of this course was to have an expert teach google analytics.
-There will be an NLP elective this year.
-Maybe they need an applied data challenge class? Applied machine learning to interviews with case studies?

This year: Faculty, ask your favorite students for their interview questions.

Note: 2018/2019 we are up for a program review. Lengthy report, external team's review, comprehensive evaluation, one-on-one in.

3. Program Goals

Faculty voted on removing the following program goals from the assessment report:

1) **Job placement success**: At least 90% of the program's graduates will be employed within three months of graduation.
2) **A strong enrollment pipeline:** By fiscal year 2021, the program will have at least 1000 applications.
3) **Strong enrollment performance:** By fiscal year 2021, the program will achieve a long-term steady-state enrollment of approximately 100 students.
4) **Strong financial performance**: The program will produce contribution margins in excess of 40%.

Faculty were uncomfortable with committing to hard numerical targets in connection with job placement, enrollment, and profitability. Faculty acknowledged that these sorts of metrics might be imposed by administration.

~~4. Practicum Update~~

5. DI Update

-We currently have four partners and growing. Development is bringing us more people.
-The deep learning certificate is the highest trafficked site on USF through social media.
-We now have 15 paid attendees for Deep Learning certificate, plus 5 alums. We will easily hit our desired enrollment.
-We are waiting on approval for 5 certificate courses for spring.

MSAN enrollment: Our application is open. We are aiming to let in 80-85.

Nick: We are leaving valuable students out there by not accepting more. We have already sacrificed the “small group” feel.

Others: We don’t know how growing will affect the culture at this point this year. We don’t have enough practicum mentors at this point. If we accepted 20 more people, we would throw off our demographics.

...

MSAN Faculty Meeting
January 23, 2017

Attendees: Kirsten Keihl, David Guy Brizan, David Uminsky, Jeff Hamrick, Michael Brzustowicz, Nick Ross, Nathaniel Stevens, Terence Parr, Paul Intrevado, Yannet Interian, Leslie Connolly Blakeman

1. Review of how cohort 5 is going, positives and negatives (please gather own thoughts). - David + All faculty
    - Practicum good
    - Slack - good/bad cuts down on email, allows for student to communicate, but students are using slack during lectures which has caused problems for Diane
    - Good batch of students, they are good at taking tests but inquisitiveness is down but could be attributed to larger cohort
    - Lack of participation, smart but lack of participation
    - Pick students randomly and ask question, students answer
    - First half vs. second half - basic first, second half think harder and then ask
    - Not randomize modules, create a “cohort” so students can create a closer bond.
        - Tracking vs. randomization trial
        - Set tracks until after practicums - done this way to make practicum meetings easier
        - People like the idea of one cohort
        - Nick had to be more aggressive and engaging with students due to the large cohort size but it worked to his benefit in the end
        - Paul - little nit-picky stuff students are going the distance, students asking basic questions. Could be worse than other years, but comes up every year
    - Job Placement - working on moving things up, Trying not to panic them by looking right now.
    - How do we measure inquisitiveness? No controls on the admission side to determine inquisitiveness.
    - Mostly comes down to the size of the classes, two cohorts could solve the problem. Monthly social activities - Movies, social events so they are hanging out with each other.
    - Demanding interaction
    - Diana - “Poll Everywhere” student seemed to be more interactive online. Asked Diana if other faculty would adopt this technology. Be careful because it could shut down vocal interaction.
    - Laptops down - Terrence

2. Review of grading in MSAN and discussion of grading policy. David, Terr, All faculty
    - Paul - Use percentages
    - GPA in mod 2 hit a new high.

- Calibrate - everyone has a different one to grade.
- Match where the scholarship is - 3.4 half of the class should be above the scholarship - hold them to the academic fire. More scholarships should have been pulled this year.
- We should have them alternate scheduling - can we pull that off moving forward? Even distribution of teaching
- Terence - you have to pass the final - Give him the leverage he needed.
- David - craft a policy around an avg GPA. Mapping percentages, not against. Policy can be a target but it can vary a little from it.
- Nathaniel - added a quiz and allowed students to drop their lowest grade but wouldn't do that moving forward.
- Terence - a standard policy that goes on all of the MSAN faculty syllabi.
- Above or below the mean and standard deviation if the only thing that Nick or David share. At the end they give them the weighted average.
- Just put in first paragraph into MSAN faculty syllabi
- Set and then review every semester - self regulatory behaviors to check that impulse

Review of MSAN Applicants

- Interviews completed: 120
- Domestic: recruitment pain point - good domestics this year, all of them have been processed
- Female recruitment is up

3. Announce Data Institute Annual Conference (James and Paul)

- October 15, 16 & 17 2017 @ 101 Howard
- Annual Conference - academic oriented
- James and Paul - Co-chairs, conducting research
- Whole building to use
- Has a conference schedule which includes workshops, plenary, 5 concurrent sessions.
- Michael Jordan - first Plenary. Will be research talks and practitioners, social science track
- MSF grant to get funding for this?
- Monday night - Poster session & Cocktails - up to 56/60 posters, Banquet Dinner
- Pricing $495 Early Bird + Association Member, $595 Early Bird, Regular $695, Practitioner NA, $1195, $1495 - Conference registration does not include lunch, dinner, banquet dinner, or workshops.
- No students at this conference
- List of Potential Workshops - just ideas on the PPT -- need help identifying really good track chairs.
    - A lot of work trying to get 24 speakers for a full track. Might be a lot to ask -- Nick Maybe split it up to different focuses.
- Looking to have 350 ⅔ being academic, ⅓ industry
- Website should be up on Thursday with password protected.

4. Discuss the increase in computing requirements and potential solutions to meet that need in our students, including a computing competency exam during bootcamp (Yannet/DGB/Terr/Diane)

- Terence - computational bootcamp - during interviews determine which students can skip certain bootcamp courses. Project/mini-labs for them to submit during bootcamp
- Yannet - Interview testing is very basic. If you don't know how to do computing. Have a very basic programming proficiency, let them take it three times. If you pass the basic things, you don't have to take Terence's bootcamp class. Terence make one at the beginning and one at the end.
- Early test to confirm that they proved themselves
- Add that to the orientation.
- Can they take all 4? Can't handle taking all 4.
- Worst coders in Python - did they take the class in bootcamp, did they pass the bootcamp? Any of the worse 6 skip the course during bootcamp?
- Want students to engage before. Those who study the homework before they arrive, do well in the program. Frontload an assignment at the beginning and then determining bootcamp courses when they arrive.
- Why limited to the CS bootcamp? Add placement exams to orientation.
- We have drifted towards a more technical program.
- Nick - does this get us the type of students we want to recruit?
- Yannet - students might be taking the wrong classes in bootcamp.
- Nick - let's not over react, let's teach the right courses and then see what happens moving forward.
- Terence - teach computing skills that they will need for everything else.
- Yannet - we are not identifying them in our current application process
- Nick - this problem is coming up because Terence didnt give the bootcamp as he should have. And weed out the people who can't do that.
- How do we make sure we didn't screw up in the interview? Do we have a small 30-minutes quick check?
- Nick - isn't that the purpose of the interview?
- Paul - practical piece - how do we grade all of these within one-day?
- Jeff - Keep is low scale at the start. Line up reviews.
- Steve - Drop linear algebra bootcamp?
- Terence - Yannet, Diane David will talk and determine
- David - Here is a deadline because he will send out info to students. Entrance exam.
- Diane - students need to practive more problem solving not there. What if we have programming classes that can use to practice. Everyday they had quizzes but they had a hard time.
- Terence - Data Aquistion might have been because they copy and pasted
- CS faculty get together and do a full redesign - david we have resources and opportunity. Rework the problem assignments so they are at a place that is better.
- Propose the solutions. Routing students in bootcamp.
-

5. review Idea of practicum incubator (Nick)
    - One page and it would have to be reviewed and approved
    - They have to be rockstars
    - If a student has an idea, it can be a practicum
    - Very little data science content
    - Micro version of this in Mike's Class - 5 week build product
    - Need VC's there for the final review and watch these presentations
    - All student based ideas
    - Pitch day - one page, voluntary data driven business idea. High GPA, solid student.
    - Monetary value - students would keep everything.
    - Nick will be a Mentor, throw it out there and see what happens.
    - Paul - how much entrpenure are they getting into? Nick - want them to code at the end of the day, turn into a CS thing. Business stuff would happen afterwards.
    - Goal: you have to have something on the internet that I can click on that does something.
    - Paul - How does this compare to the current practicum, make sure students are putting into the same effort as other students in traditional practicum.
    - Need a new IP/NDA for this.

6. Discuss partnering with international schools and companies (Nick)
    - German university looking to trade students. What if we brought them here and took them to see the practicum partners. Can charge for this type of thing. Offer them real floor space for these students and maybe a class/certificate piece.
    - Partnering with international univeristy's - take advantage of these opportunities. If we want to keep this program running, we need to take advantage of these sorts of thing. If we partner with a univeristy for a 2 month incubator, could help fund the program and data institute. We need to jump on them a lot quicker at this point.
    - If they can no longer get jobs here but their goal is to return back to their home countries.
    - Yannet - maybe these companies can help pay the students school loans. Maybe create a contract for students with companies regarding
    - Staying at apple, apple pays for her education. Creating pathways in that style but with international students. Domestic framework would pivot easily for internatioal students.
    - Trying to get companies to hire students right off the bat. Our name hurts us a lot because we aren't Berkeley. China/India might have more prospects.

7. Leverage current Meetup more and launch more potential events at night (Nick)
    - Mindi has a great idea
        - SUDS? Same place we had the holiday party. Current meetup groups. More evening meetups
        - Nick - data science meetup at night, once every three months. - Build something really quick - Chat Bot. On campus or at a different location. Need other faculty to pitch-in.

- Bay Area AI folks looking for a place right now DGB
- Women in Statistics - James idea
- Forward ideas/opportunities to Nick
- Recruit and get present, bay area numbers can be better.

8. Review statistical assessment exam (Jeff, James, Nathaniel, Paul)
- End of module 2
- To get an A+, you need both direct and indirect assessment. Would need to be mapped on the rubrix thing.
- Can try to put it into an automated system to make it easier - Canvas
- 2 hours to take this assessment
- Yannet - Randomize 15 questions
- Between module two and electives
- Nick - incorporate into the curriculum?
- CAS has a assessment process and would like to incorporate this into MSAN so they can just easily
- Make it manditory, random 15 questions so it's not two hours.

9. Propose and discuss curricular changes (all faculty)
- Diane - distributed computing - will need to take some thought and reworking. Change the course titles
    - Remind Diane that No SQL
    - Deadline next week for any changes
- Web Analytics - see her go through it once and then bring a porposal for next fall and then formalize.

Class assignments for next year? David will be emailing all faculty individulally and will determine class assignments.

10. Spend remaining time considering MSAN long term in lieu of potential changes to OPT and H1B thanks to new presidente

Look at your own class and rewrite the description as you see fit. We will do an approval in house and change them online. No deadline.
- Statistiions will get together to review

One unit to two unit - practicum is an insane amount of hours for students/faculty. Claim that they made a mistake to lower to one unit per module, currently four units.

Where are our peers, we are in the bottom half of our peers. Increase the cost of the program, $5k increase for next year. Current 45k.
Stay on this side of 50k, change one practicum to 1 unit to keep it under 50k
If we go up in price, it would not happen for next year.
Why would we do it? - take pressure off several areas, we would not need to grow as fast, more faculty resources per student. Faculty comp structure is exceptional for the university.
Generate a contingency place if OPT and H1-B visas are canceled - Would cause a complete redesign.
Vote now, wouldn't come online until 2018. When can we pull it back?
David will send out more information about this idea.
Leslie to research more on GRE/GMAT
Hard to find time to meet with everyone.

11. Items from the floor.
    - No Seminar next friday, we need a speaker for February
        - David Guy will do it
    - MS Data Engineering - Paul
        - Need to get this moving as things are changing. If started right now, wouldn't come online for two years. And just kill if it's not the right thing to do. Need faculty to volunteer and get this started. Nick will volunteer - need two more people to volunteer. Nick, David and Yannet. David will send out an email and include the board of advisors to determine what is needed from a program like this. Barret would be a good person to contact. Someone on the hiring side and someone from the engineering side to determine what is needed.
            - Needs: David will send a link to a propose a degree
            - David will show them the first few steps.
            - Conduct a marketing report

12. Go directly to Barrel head.

University of San Francisco
MS in Analytics
Faculty Meeting

Meeting Summary

Date:	05-16-2017	Start Time:		End Time:		Location:	

Attendees:

Note taker: Leslie Blakeman

Apologies:

Topics:

1. Course Description:

Are all Course descriptions in? Will update typos for us.
Submitting course descriptions as a batch job. Terence and Jeff have assisted in collecting and will submit on behalf of the MSAN dept.
Teaser descriptions for course descriptions, look for typos. If they are all in, electronic vote.
Once they were reviewed. All course descriptions are in.

2. Year in Review

DU - we made a lot of substantial changes this year and it's time to review.
Terence - year 1 to year 5. Started out with anyone with a checkbook, it is remarkable that we have created a massive non-liner climb. Easy eclipsed all programs at the university, and faculty quality.
Jeff - Pres said that it is the best program at the university. Word of mouth, in industry has heard comments.
James - Last module - student cohegian was pretty good, helped that some students are good at bringing the cohort together. There were some cliques. Some needy students, wanting a good grade and thinking they deserve a good grade. Entitled students.
Yanette - Weekly quiz, helped students gauge where they stand. Maybe a good way to prevent students feeling entitled to a certain grade.
James - Student asked "could I make my class easier?" hope that everyone stays rigurous and keeps the teaching in this program top-notch. Not sure where it's coming from.
Jeff - calibrating grades could help with this issue. Would let students know if they are an outlier and give them an idea where they stand.

David - high variance in grading, will get to later. Content wise, everything is harder than previous years. Consistently increasing in expectations and learning outcomes.
David G - pain points? First module was really tough, understanding what the students needed and how to teach it. Feels he didn't get it right from the get-go.
Diane - didn't understand the students backgrounds that much. They didn't have the NOSQL course, distributed computing harder for the students to understand. They think CS is writing code. Felt that she was being treated as an IT person and just looking to her to fix their code. Maybe they might need some very basic CS.
Nick - some students who don't know either and they still get good jobs.
Tere- Started with a blank screen for Comp Bootcamp. Overall problem solving strategy. Everything will be done in the abstract and then will move to code. Making a course on how to think like a computer scientist.
David - this has been run by the CS faculty as well.
Paul - how important to code or program in R?
Yanette - big functions preferred.
Paul - then why are we teaching it twice?
Nick - more feedback regarding student comments
Paul - Changing the focus from R to Python.
Ter- python and R are going to dovetail
Nick - should he focus more on packages or do more typical CS stuff? Personally would veir towards Packages.
Jeff - Go to packages because it is easier to implement in a 7 week course.
James - Packages and functions are going to be needed. But if the view is python mass scale, R for certain functions and packages.
Yanette - R was limiting because it was hard to implement. Some sort of testing from scratch. They end up not using the theory as much. Write some test because then they can explain how the tests are done.
David G - Agrees.
Jeff - develop mini regression package. Not sure how to manage the to a 7 week course.
David - compromise. One development oriented and one that is not. Need to understand more in R. Limited amount of development with packages.
Jeff - 601 and 604 make a considerable test/implementation of R.
David - Students need time series for their practicums.
Paul - practicum Last year vs. this year - practicum left tail shed. Worst projects now - colorox style of projects that are difficult because they aren't lead by data scientists. With help those sorts of things get on track. Students pitched their own projects. Students are everywhere now - coke, silicon valley bank, united health care for next year. Practicum are going very well - eventbrite had some legal issues. We now have the legal language can be hired as practicum without pay. Pauls connections with students is tenuous, not workable with this amount of students and number of projects.
Terence - powwow not bring back.
David G - Best/Worst

Jeff - miss-scaled bad choice - one man teams is an unhealthy structure. They don't have someone to bounce ideas off of except their practicum mentor.
Paul - that's a company thing. Do we say no to those companies?
Jeff - yes, maybe moving forward. 3 member teams are ideal.
David - how many successful teams of 1 or 2?
Yanette - Vungle is the only place it is working. Students are working on projects that are redundant William Sonoma is the worst - they don't have enough work. They are not managed well.
Paul - at NCU are they dealing with bigger companies?
Jeff - they are doing less work at their practicums. They have an MBA mentality.
Nick - UCLA 5 person teams, when they don't work the faculty will generate work for them.
Jeff - not a great idea.
Paul - maybe able to force people, but now while we are scaling up.
Yanette - quality of projects is getting better. Can we meet to go over what practicums are working or not working? Meet to go over the best/worst and things we would like to change. Maybe some companies can be let go is they aren't working.

3. App Development: Mike B
Students ran with it and half of the class did an awesome job, the other half didn't really try hard and did the bare minimum. Maybe class can be purely backend. Focus on entrepreneurial in Nick's class. Some students really came out of their shell. Brad and Kyle did a great job in their presentations.
Yanette - backup homework? Half of the class does this thing and create a lot of structure and the other half of the
Nick - in the end of my class, propose a business plan, need to be data focused so they would have something walking into Mike's class. So he would have a data point at the beginning. Team building was slow at the beginning of Mike's class.
Jeff - planning/stability issues - are we going to be ok if something happens to Mike? Diane and Nick can handle it.
Mike - let them pick their own groups. Asked them not to pick their friends. Is there a way we can break up some cliques so that students with different skills are working together. Mike would be happy to teach next year, not sure where he will be next year.
David - thank you, final day was pretty amazing. David brought in 5 VCs and they left the presentations feeling pretty good about the app ideas.
---Paul is meeting with vincent the engineering guy right now. From VC day.

Accomplishments:
Jeff - navigating the university through 7m of budget cuts. David - reinspired to work on a paper.
David G - Paper with Claire. Peer review committee loves student/faculty a lot.
James - excited for the conference, finishing up a paper. Working with kelsea on a paper.
Mindi - Yanett deep learning and course. Great tie in with the data institute.
Terence - new web monkey. And thanks to jeremy's help kick ass forest.

Diane - paper she wrote with brichen, egan accepted to publish. Undergrad student to write a paper data science student. Jesuit scholarship tentatively accepted - aleia in the data science undergrad.
Yanette - Machine Learning class went, much happier about how things are going. The two students at UCSF, both of their work got accepted. Very excited about deep learning.
Nathaniel - Submitted a few paper, a few other accepted. Particularly happy with how many students want to take D&A experiments.
Nick - surviving this year. David - nick did a lot, pretty much works 2.25 jobs.
Mike - finished his book, QC1 phase now. Will be printed soon, Maybe headed back to research.

Admissions:
Last year 122 soft offers to bring in 60 students.
This year, targeted 70 toward end move to 80. 183 soft offers.
Total deposits 2016 - 70
This year - 105 deposits
Declined informal offers - 11
Total dropped deposits so far, 16. Close to where we should be.
David's hope, we are going to lose another 12. We are a space of they better drop.
Bootcamp hoping to drop a few more.
Last year, we did not overshoot. As late drops came in, we made offers to top of the waitlist. We are fine if the models hold. Scarier to do it this way.
Right now we are at 62% international, internationals are the late drops.
Median GRE 165 last year, 167 this year.
Gmat held at 50
GPA is about holding
One PHD coming in next year, 4 PHDs this year.
Median work experience stayed down to two year. This year will hold at two. Work experience is so incredibly helpful when it comes to getting a job.
Jeff - maybe this has to do with the economy right now.
Discussed the overall makeup of the incoming cohort.
Jeff - utilizing the undergrad USF pipeline to help diversify future cohorts.

Thanks to mindi, kirsten, leslie for helping with the admission process.

Bootcamp Structure - skip for now

Future of web analytics course:
Yannett, sorry for Diane - could be outdated. Diane not 100% sure of what to teach. Can we have two classes that students can choose from? Students aren't that interested in that sort of stuff.
David - still open to making the whole end open to electives.

Jeff - not budgeted for. Change in curriculum structure. Elective implies choice, this violates the budget because we will have to pay two different teachers to teach the course.
David we have capped the number of electives we offered this year. 4 this year and 4 last year.
David - in favor of keeping business content, maybe a different course that has business contect built in. if they can not answer how this is going to improve the bottom line and need a course that needs to answer that question. They need to answer how this adds value.
Yanette - this also means work experience. Work experience is important.
Terene - focus on that in the bootcamp. First question to every client is how do you make money?
James - agrees that it should go in the practicum.
Nick - something that they do stress and harps on them in biz comm and don't go to the interview not knowing how the company makes money. How do you reinforce that learning?
David - the seminar does this very well, the students aren't really absorbing the learning from seminar? Web Analytics on the table for next week, in fall meeting we propose a new course if web analytics isn't working.

Integrating student life:
Leslie, Mindi, Kirsten - working on student life
David - events are great for students, helps them bond.
Diane - group project, mixing students. Cultural thing because international students don't want to speak english or makes it hard to connect.
Nick - reinforce this in class. Can be dinged for not speaking english in class. Non-english is a zero.
David - stronger statement than the one he makes in orientation. At the end of the day, chunk of students who didn't make strides in communication this year. In previous years, students got better while they were here in the program.
Nat - admitted students day, english was terrible.
David - everyone starts off at a very different level of communication
Nick - Different events, and just invite 20 randomly. 10 events, everyone gets to go to 2 and this will randomly pull together students.
Tere - CS grad students, went to a grad/fac mixer and this created a massive level of improvement in communication.
David - organized launched, faculty and staff volunteer. Faculty lunch, random selection of students. A way to get to know students.
Jeff - maybe before seminar?
Kirsten - big gap with bootcamp
As a way to create a bond with students.
David - things that don't work anymore -
Homework - dropping off homework at Kirsten's desk isn't working. Need to have a better way of distributing homework to students.
Concerns - timeliness, I don't have an office at the downtown campus, not handing back in class because it takes too long,
Tere - last 4 numbers of their CWID solves the privacy issue.

Nathan - but students can still take other students homework.
David - bootcamp is an issue, students are fragile,
Tere - aren't they getting their grade on Canvas?
Daine - can you leave comments on canvas?
Jeff - not entertaining something that is dependent on internet downtown
Paul - what if we generate solutions? They have access to solutions.
Jeff - students want to know what their written feedback is and solutions aren't cutting it. Does not have time to hand back homeworks in class with a compressed course schedule.
Terence - just use CWID numbers instead of names and students pass them out in class.
Yanette - how do you enter it into canvas? How do you learn their names?
David - propose solutions:
    Nathaniel Solutions - faculties job to hand out in class
    David is ok with hiring someone to sort
    Can student worker grade enter grades on canvas?
Jeff - student worker not a workable solution
James - handing it back during office hours? And if they don't pick-up you throw it away?

Solution - workstation behind Ashli's desk. Use hanging folders there, stack papers for students to go through until the student worker who can come in and file stuff.
If we can get that space, then we will do it. If you need to store stuff, give to david or terence's office to store.

Professional communication:
Nick - name change ok? Just took the work business out so it would be under 30
When we went from 2 units to one, the thing that got cut out was presentation. Doing it once without repetition meant students did it and then forgot. Toss notion of presentations and focus on soft interview skills in the last few weeks of class.
Yanette - in mentoring we can ask students to present to me and that will take some stress off of Nick.
Nick - 20 students, multiple students. This last time they did one presentation and there wasn't any follow-through.
How many students are doing presentations? Not many.
Jeff - what if we hire an outside consultant and we bring people into review their presentations. Nick only has 7 hours with them, would need to be parallel and not take up time in class. One in class and a secondary follow-up. Need to find 16 hours of people who can deal with it. Can he get rid of it and focus on interview skills.
David - nicks class is the only course where students are giving presentations but students are presenting at their practicums. Don't want to lose presentation skills and at least give the students feedback. Students will do a poor presentation if they don't get feedback.
Jeff - nick does the feedback once, then the mentor can do the follow-through and can review the slides and presentations. Standardized review policy
David - practicum used to meet two hours a week, and used to review procedure and how-to.

Nick - maybe reserve a day for practicum presentations? This could create issues with NDAs and students can't show what they are working one.
Could schedule time to do this.

The Blog - Thank you terence for running the blog.
We do a lot of amazing things, students, faculty, staff do amazing things - use the blog to proclaim the great things we are working on. PR funnel, published and then blasted. Use it as a place to post news.
Need to create the structure - use the data institute site for blog.
Terence - two way to go, blog page once a month something to add or …. Do you want it under the same URL? Eventually that page will get big and it won't let you search because it just a basic html site. Search engines might find direct sign first and doesn't actually drive traffic. Start with /blog
Yanette - get a student involved
Create a schedule for upcoming items.

Confernee:
James - looking good, we have like 10 registration. Need help from everyone to spread the word. Want to try and get it out to everyone as fast as possible. CS students have not been his yet.
ACM is officially but they have not gotten back about if we can use their logo. Terence said we can use it in accordance with their website.

Curriculum:
What do we want to do about the curriculum?
Swapping
Experimental design last year, and this year. Move multivariat as an advanced stats course.
Nick - should have multi before, students check out the last module.
Nathaniel - happy with it going either way

Any other proposals?
Reorganizing no sql
Move design before multi.
James - a lot of students want that to be required. Students are in interview mood, moving around interview skills. Students are gearing up in the final module.
Yanett - Move interview skills class later, one module different.
David G - students are nervous regarding interviews. Students don't get that interviewing is a big deal until they get into his class. Likes it where it is, but will entertain the idea of it moving.
David U - students are doing pretty well in interviews right now.
David G - students started interview very early.
Terence - keep it where it is, because we have a good success rate.

Nick - interview skills, makes a change in the curriculum. Start talking about the next phase. If we move it forward, students aren't going to get that click.
David G - can we change it to be a graded course from pass - fail?
DU - how can we get that changed - would need to involve June.
Jeff - we would need to argue that the change to pass/fail was a mistake.
Vote: everyone voted yes to this change.

Vote on anything else?
David - on web analytics - let Diane teach and then we will vote in the fall meeting

David - DOE with multivariate. Is there another place to do DOE instead? One-unit DOE?
Terence - why not just leave it that way.
Nathaniel - learning it too late.
Nick - students seems to forget it because they don't use it right away.
Terence - for at least 50% of the students, where it is is fine.
Student got to choose their electives.
Stays where it is, we will vote on it in the fall.

Terence - network update - has been bothering ITS about network. Late july they will get a network connection. New consultants have been hired again, third set of consultants. Will moderate youtube, facebook live, netflix.

Diane - swap nosql and dis computing. Ask diane what she wants to do here!!!
Adding some date structure in inner session
Nosql, dis comp. Change the order.dis comp, and leftovers will go into the 2 unit course will cover nosql and data structures which leads into the other classes.
Nick what are you cutting?
Casandra would be gone but would cover Nosql at the concept level.
Nick - don't base these changes on these students who are the outliers.
Terence - use mongo, amazon services, spawning clusters in the cloud all that sort of stuff can be crammed in.
Mike - sql how do we make it easier with plain old sql
These changes have been reflected. These changes have been emailed out.
Vote: Support diane's proposal - everyone.
Tenure packets - diance can file these so she is the faculty on record. It goes into your tenure packet. Course replacement and change graduation requirements.

Slack - Vote
Keep students - vote: everyone, stays as is.
Helps us keep on eye on what is going one.
Terence - too much stuff is missed on faculty slack. Want to talk to faculty is should be an email transaction. Threading on slack isn't great.
Jeff - determine better etiquette? Official business should be on the faculty listserv.

David - low level communication on slack, will keep it for chit-chat purposes. Subgrouping is good but official business needs to be on email.
Terence will be the email cop. Anything that needs to be discussed in a meeting should not be in the slack channel.

Jeff - Meeting proctoring?
Yanette - brought this up regarding test taking. Who will proctor these quizzes?
Faculty show up early for bootcamp to proctor each others exams.

Stats Survey: Nick and Jeff

Data institute Update: still exists. Update on what it has done this year -
Mindi conference in october james/david, certificates SQL, Data Acquisition. Some rescheduled, some canceled. Will run deep learning again. 5 formal partners. Taking on a nonprofit partner now, presenting their work this summer. Looking to host more events, will be hosting the a/b testing group this summer. Book launch party for mike.
David - we had the call to run revenue positive, running better than expected. Meeting and maintaining contacts. Yaneet, james, shumy, david grant. The data institute is resource for research, to teach. Members thing is exciting, 5 companies, 5 figures each and benefit directly from being a member. They want students, pipelines and trainings.

Book orders -
Send them to kirsten/leslie by end of day.
If anyone wants a standing desk:
Diane - bookshelf, look into sticker whiteboard for Diane, send me her ergonomic order,