



HCMG 357/857 – HEALTH CARE DATA & ANALYTICS

Fall 2020, 0.5 CU, Mondays 3-4:20pm

Instructor: Professor Matthew Grennan, PhD grennan@wharton.upenn.edu

Teaching Assistants: Stuart Craig stucraig@wharton.upenn.edu

Gi Heung Kim robinkim@wharton.upenn.edu

Joseph Martinez josma@wharton.upenn.edu

Office Hours: by appointment

Course Description: This course is appropriate for managers or investors who want to better understand data science as being applied in healthcare, or for beginning or experienced data scientists who seek to better understand the health care vertical. Students will become familiar with basic tools of modern data science through hands on experience. We will also explore current business applications of these tools to the available data in the health care industry.

As we discuss examples, you will notice how human creativity needs to be at the heart of the questions being probed. Making effective use of data and analytics in business requires combining that creativity with technical and business skills to solve problems and unlock value.

At the end of this course, students will:

1. Know the health care data landscape
2. Understand the “loop” that drives modern evidence-based businesses
3. Dive into real health care data analytics problems, developing a first-hand familiarity with basic tools and concepts
4. Anticipate the business opportunities evolving in health care data and analytics

Recommended Texts: The notes and handouts and class discussions will be sufficient. For additional reading, I recommend the following (and there are many other good sources):

1. For “Data Science”:
Provost, Foster & Tom Fawcett (2013); *Data Science for Business*, O’Reilly.
Annalyn Ng & Keneth Soo (2017); *Numsense! Data Science for the Layman: No Math Added*, Self-published
2. For “Causal Inference”: Joshua D. Angrist & Jörn-Steffen Pischke (2014); *Mastering 'Metrics: The Path from Cause to Effect*, Princeton.
3. For background and futurism: Pedro Domingos (2015), *The Master Algorithm: How the Quest for the Ultimate Learning Machine will remake our world*, Basic Books.

Course Requirements: A curious mind, diligent preparation (10% of your grade), enthusiastic contribution to the class discussion (20% of your grade), individual problem sets (40% of your grade), and a team-based project with milestone reports (2x5=10% of your grade) and a final presentation (20% of your grade).

A successful course will depend on your engagement with and contribution to the classroom discussion. Questions, sharing of experiences, and an intellectual curiosity around anything discussed is always welcome. Each lecture is designed to be part didactic and part conversation. Your professor, TAs, and guest lecturers don’t have all the answers but know how to ask good questions.

We want the same from you: every student is expected to participate in all aspects of each lecture, case study discussion, as well as in the guest lectures.

Final projects are your chance to tailor your course experience, put the tools we learn and the health care dataverse to use, and educate your classmates on a new part of the HCD&A space. Projects can either analyse an existing company or initiative, or propose a new startup or initiative. In either case, a project will consist of: (1) a clearly defined problem to be solved, (2) motivation for what we gain by solving this problem (think in terms of questions answered and value unlocked), (3) the types of data and toolkit needed to answer the question, (4) challenges and caveats (and/or how one might overcome them), and (5) what insights this venture provides and (importantly) what additional lines of questioning this might lead to in the future. Project reports should not exceed 10 pages, excluding supporting appendices.

Course Prerequisites: For HCMG 357: STAT 101 or similar. For HCMG 857: STAT 613/621 or similar. Other experience in data science can serve as a substitute for the prerequisites. Knowledge of basic statistics is a must. Coding experience is a plus. Experience coding to solve data/statistics problems is ideal. Further training in data science is a plus, and we welcome those with more advanced preparation.

The course will involve problem sets and a final project in the R programming language. Those unfamiliar with R should take the Coursera tutorial prior to the beginning of class (<https://wcai.wharton.upenn.edu/resources/workshops-courses/>). The problem sets will introduce you to doing data science work in R, working with major sources of health care data, analyzing the data to gain insights, and presenting results clearly. The problem sets will provide a great deal of guidance and support as you go through the process, but you still need to be ready and willing to “get your hands dirty” in data and coding.

Course Outline:

Class	Date	Topic	Readings/Preparation
			<i>see Canvas</i>
1	Sep 14	Class Motivation and Expectations + Intro to HC Dataverse	
			<i>see Canvas</i>
2	Sep 21	Challenges and Potential of (Big and Small) Data & Analytics in HC	Guest: Jessie Juusola, Evidation
		<i>* Group project check-in: Potential Questions/Topics</i>	<i>Toolkit prep</i>
3	Sep 28	Toolkit: Text Analysis and Merging Data Sets	
		<i>* Problem Set 1 due</i>	<i>Case: Kyruus; see Canvas</i>
4	Oct 5	Business Models for Big Data in HC + Data Curation and Prediction	Guest: Graham Gardner, Kyruus
			<i>Toolkit prep</i>
5	Oct 12	Toolkit: Data Exploration, Cleaning, and Dashboarding	
		<i>* Problem Set 2 due</i>	<i>see Canvas</i>
6	Oct 19	ML and AI in Healthcare	
		<i>* Group project check-in: Question, Motivation, and Feasibility</i>	<i>Toolkit prep</i>
7	Oct 26	Toolkit: ML for flexible prediction with large data sets	
		<i>* Problem Set 3 due</i>	<i>Case: Intermountain Health; see Canvas</i>
8	Nov 2	Precision Medicine and Population Health Management	Guest: Benson Hsu, Sanford Health
			<i>Toolkit prep</i>
9	Nov 9	Toolkit: Causal Inference	
		<i>* Problem Set 4 due</i>	<i>Case: Oscar; see Canvas</i>
10	Nov 16	More granular data: Claims and EHR	Guest: Oscar
		<i>* Group project check-in: Preliminary Analysis</i>	<i>see Canvas</i>
11	Nov 23	Sensor Data and Business Model Integration	
			<i>see Canvas</i>
12	Nov 30	Legal and Ethical Issues in HCD&A	Guest: Lisa Clark, Duane Morris
13	Dec 7	Group Project Presentations	
14	Dec 10	Group Project Presentations	