Syllabus for RDADA

***Research Design and Applications for Data and Analysis (“RDADA”)***

version 2b -- Spring 2015

Section Instructors: Benjamin Stokes & Peter Norlander

Coursework & Instructor of Record: Steve Weber & Andy Brooks

School of Information

University of California, Berkeley

***Contents:***

[Course Summary](#h.a5d8x7mzu8tj)

[Prerequisites](#h.nwlymyy4hzpl)

[Books for students to purchase on their own](#h.sh9l202acwtk)

[Office hours](#h.xmf5ssolgkc4)

[Course Structure and Readings](#h.a1q3bx1iwuk1)

[Introduction and Overview (weeks 1 and 2)](#h.cz2f2ogp92un)

[Decision Making (weeks 3 and 4)](#h.vijzxxs9cuo9)

[Research Design (weeks 5 through 8)](#h.b03gt4ababc6)

[Conveying Findings (weeks 9 through 11)](#h.nj16hd4bsd2h)

[Future of Data Science (weeks 12 and 13)](#h.hdxzjuw6nf0e)

[Wrap-up (weeks 14 and 15)](#h.r0tuosc7xdyj)

[Assignments and Projects](#h.tmpbx9l4cn2l)

[Grading](#h.cfb8he1r1vp7)

# Course Summary

This course introduces students to the burgeoning data sciences landscape, with a particular focus on learning how to apply data science techniques to uncover, enrich, and answer the questions they will encounter and originate in industry. After an introduction to data sciences and an overview of the course, students will explore how organizations make decisions and the emerging role of big data in guiding both tactical and strategic decisions. Lectures, readings, discussions, and assignments will teach how to apply disciplined, creative methods to ask better questions, gather data, interpret results, and convey findings to various audiences in ways that change minds and change behaviors. The emphasis throughout is on making practical contributions to real decisions that organizations will and should make. Industries and domains that we will explore include sports management, finance, energy, journalism, intelligence, healthcare, and media entertainment.

## Prerequisites

There are no prerequisites for this course.

## Books for students to purchase on their own

* Course pack of readings, available via Study.Net *(student services will help you secure the files)*
* Daniel Kahneman, Thinking, Fast and Slow. Publisher: Farrar, Straus and Giroux; Reprint edition (April 2, 2013), ISBN 978-0374533557
* Darrell Huff and Irving Geis, How to Lie with Statistics, Publisher: W. W. Norton & Company; Reissue edition (October 17, 1993), ISBN 978-0393310726
* Brian McDonald, Invisible Ink: A Practical Guide to Building Stories that Resonate, Publisher: Libertary Company (January 11, 2010), ISBN 978-0984178629
* Edward Tufte, Visual Display of Quantitative Information, Publisher: Graphics Press; 2nd edition (May 2001), ISBN 978-0961392147
* Allison, Graham, and Philip Zelikow. *Essence of Decision, 2nd edition*. Longman, 1999.

## Office hours

All students are encouraged to drop by and chat; even if you have no specific questions, it’s a good time to reconnect and catch up. Office hours are on Fridays from 11:30am-12:30pm PT. Drop-in office hours may also be available at other times by emailing your instructors.

# Course Structure and Readings

*The course is conceptualized in six sections, detailed below.*

## Introduction and Overview (weeks 1 and 2)

Introduce the logic/flow for the program. Insights here set you up for the rest of the program. Framing data science at the School of Information: What is data science? What is data science becoming? Who does it? Where? Why? Data science myths, fallacies, and misconceptions. Highlight some aspects of sectors doing data science: Internet technology, healthcare, advertising/marketing, and so on. Historical context: data science relative to business intelligence (and other similar-sounding techniques). Some elements of data in the wild: web logs, attitudinal surveys, and so on.

Week 1

* Data science at the School of Information: More than a technical discipline
* How organizations will benefit from broadly trained data scientists
* Data science today—reality, aspiration, controversy, possibility

Required readings:

* EMC Data Science Community, *Data Science Revealed: A Data-Driven Glimpse into a Burgeoning New Field*. 2012.
* LaValle, Steve, Eric Lesser, Rebecca Shockley, Michael S. Hopkins, and Nina Kruschwitz. “Big Data Analytics and the Path from Insights to Value.” *MIT Sloan Management Review* 52, no. 2 (Winter 2011). <http://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/>
* Le Grand, Julian, and Zack Cooper. “The Geeks Must Quash the Believers in Gut Instinct.” *Financial Times* (February 21, 2012). <http://www.ft.com/intl/cms/s/0/5a996db2-5c93-11e1-8f1f-00144feabdc0.html>.
* Davenport, Thomas H. “Competing on Analytics.” *Harvard Business Review* (January 2006).
* Martin, Roger. “Beyond the Numbers: Building Your Qualitative Intelligence.” Rotman School of Management, 2010.
* Salsburg, David. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. Holt Paperbacks, 2002, chapter 2.

Optional:

* Dutcher, Jenna. “What is Big Data?” *Berkeley Data Science Program* (September 2014). <http://datascience.berkeley.edu/what-is-big-data/>
* Anderson, Chris. “The End of Theory, Data Deluge Makes Scientific Method Obsolete.” *Wired* (July 2008). <http://www.wired.com/science/discoveries/magazine/16-07/pb_theory>.

Week 2

* + When big data meets big decisions
  + History of business intelligence

Required readings:

* Voytek, Bradley. “Automated Science, Deep Data, and the Paradox of Information.” *O’Reilly Radar* (March 30, 2012). <http://radar.oreilly.com/2012/03/data-science-deep-data-information-paradox.html>
* Shah, Shvetank, Andrew Horne, and Jaime Capellá. “Good Data Won't Guarantee Good Decisions.” *Harvard Business Review* (April 2012).

*(Note that the readings are lighter this week: You’ll need that extra time for the group assignment.)*

Readings to dive deeper (optional) for Weeks 1 and 2:

* Alamar, Benjamin, and Vijay Mehrotra. “Beyond ‘Moneyball’: The Rapidly Evolving World of Sports Analytics, Part I.” *Analytics Magazine* (September–October 2011).
* Heuer, Richard J., Jr. “Psychology of Intelligence Analysis.” Center for the Study of Intelligence, 1999, chapter 6.
* Lewis, Michael. “Beane Counter.” *Sports Illustrated* (May 12, 2003).
* Loveman, Gary. “Diamonds in the Data Mine.” *Harvard Business Review* (May 2003).
* Hunter, Kathryn Montgomery. *Doctors Stories*. Princeton University Press, 1993, 21–26 and 51–57.

## Decision Making (weeks 3 and 4)

The second section focuses on learning some key principles underlying individual and organizational decision making. We’ll learn traditional and emerging decision-making models through examining iconic case studies in high-stakes decision making, with and without data science, and however rudimentary or advanced. We’ll identify the types of data needed in decision making, what decisions warrant *big* data, and the type and availability of such data. We’ll also consider the powerful role of cognitive biases in everyday decision making, such as anchoring, groupthink, inertia, recency, and attribution asymmetry, and the role of data in extending and countering those biases. Learnings will be applied to contemporary decisions confronting decision makers in various domains.

Week 3

* Cuban Missile Crisis case study
* Decision-making models, including analogical reasoning and analysis of competing hypothesis
* McNamara and the Vietnam War’s body count
* Identifying data needs; when do we need *big* data?; data availability challenges and opportunities

Required readings:

* Allison, Graham, and Philip Zelikow. *Essence of Decision, 2nd edition*. Longman, 1999, chapters 3 and 4.
* Neustad, Richard, and Ernest May. *Thinking in Time: The Uses of History for Decision Makers, 2nd edition*. Free Press, 1988, chapter 1.

Week 4

* Biases in individual and organizational decision making, including: anchoring, groupthink, inertia, recency, attribution asymmetry. Countering bias in everyday life

Required readings:

* Stauffer, David. “How Good Data Leads to Bad Decisions.” *Harvard Business Publishing Newsletters* (2002).
* Davenport, Thomas H. “Make Better Decisions.” *Harvard Business Review* (November 2009).
* Hammond, John S., Ralph L. Keeney, and Howard Raiffa. “The Hidden Traps in Decision Making.” *Harvard Business Review* (September–October 1998).
* Kahneman, Daniel. *Thinking, Fast and Slow*. Farrar, Strauss and Giroux, 2011, chapters 10–18.

Readings to dive deeper (optional):

* Allison, Graham, and Philip Zelikow. *Essence of Decision, 2nd edition*. Longman, 1999, chapters 5 and 6.
* Beal, Dave. “For Numbers Crunchers, Minnesota Twins' Old-School Methods Don't Add Up.” *Twin Cities Pioneer Press* (June 27, 2012). <http://www.twincities.com/twins/ci_20952060/numbers-crunchers-twins-old-school-methods-dont-add>.
* Davenport, Thomas H., and Brook Manville. *Judgment Calls: Twelve Stories of Big Decisions and the Teams That Got Them Right*. Harvard Business Review Press, 2012, chapter 2, “WGB Homes: How Can We Sell This House?” and chapter 8, “Mabel Yu and the Vanguard Group: Should We Recommend This Bond to Investors?”
* Davenport, Thomas H., and Brook Manville. *Judgment Calls: Twelve Stories of Big Decisions and the Teams That Got Them Right*. Harvard Business Review Press, 2012, chapter 10, “Should We Restructure for a New Strategy?”
* Heuer, Richard J., Jr. “Psychology of Intelligence Analysis.” Center for the Study of Intelligence, 1999, chapters 2, 9–13.
* Kahneman, Daniel, and Gary Klein. “Conditions for Intuitive Expertise: A Failure to Disagree.” *American Psychologist* 64 (2009).
* Kahneman, Daniel, and Amos Tversky. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47 (1979).
* Neyer, Rob. “Phillies Keep Winning without Your Fancy Numbers.” *Baseball Nation* (March 2, 2012). <http://mlb.sbnation.com/2012/3/2/2839053/phillies-keep-winning-without-your-fancy-numbers>.
* Tverksy, Amos, and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (1974).
* Verducci, Tom. “The Art of Winning an (Even More) Unfair Game.” *Sports Illustrated* (September 26, 2011). <http://sportsillustrated.cnn.com/vault/article/magazine/MAG1190632/index.htm>.

## Research Design (weeks 5 through 8)

This section of the course focuses on the general principles of efficient *research design*, how to best construct a question, gather data, and interpret results in order to meet a specific need. Starting with a review of the scientific method and how it has evolved over time, we then delve into identifying which questions are worth asking, how we as data scientists should best ask those questions, and how to evaluate the answers we uncover. We’ll consider the differences between *prediction* and *explanation* techniques, and the complex and complexly related concepts and uses of correlation and causation. We’ll touch on essential statistical concepts in plain English, as well as how to instrument existing and new environments to gather actionable data for analysis.

Week 5 (note that the async videos take up more time than usual this week)

* The “scientific method”—and its discontents
* Perspectives on “knowing,” from philosophy to action
* Paradigms and constructive conversations

Required readings:

* (revisit Anderson from wk1)
* Engineering and Public Policy Committee on Science. *On Being a Scientist: A Guide to Responsible Conduct in Research*. National Academies Press, 2009.
* Burton, Robert. *On Being Certain*. St. Martin’s Griffin, 2009, chapters 1 and 2.
* Kuhn, Thomas. *The Structure of Scientific Revolutions*. University of Chicago Press, 2012, chapter 12.

Week 6

* The linear model of research design
* The iterative reality of research design
* Asking better questions

Required readings:

* Creswell, John W. *Research Design: Qualitative, Quantitative, and Mixed Methods*. Sage Publications, 2008, chapters 1 (approach), 6 (research questions), 7 (theory/RQs), 8 (quantitative methods). *→ Hint: Consider reading chapter 6 first, then chapter one to the point that you can make good sense of Figure 1.2, then the first part of chapter 7, and skim all of chapter 8*

Week 7

* Tempting fallacies of argumentation
* Uncovering buried assumptions and dead conventions
* Simple rules of inference

Readings: suggestion

* Rao, Venkatesh. “The Dangerous Art of the Right Question.” *Trailblazers* (July 20, 2010). <http://blog.trailmeme.com/2010/07/the-dangerous-art-of-the-right-question/><http://bobulate.com/post/838164346/the-dangerous-art-of-the-right-question> for key summary, and the [full PDF](https://www.dropbox.com/s/wzhqphghro8ahbg/Venktatesh%282010%29TheDangerousArtoftheRightQuestion.pdf?dl=0) (a temporary link).
* Huff, Darrell Huff, and Irving Geis. *How to Lie with Statistics*. W. W. Norton, 1993.
* Best, Joel. *Stat-Spotting: A Field Guide to Identifying Dubious Data*. University of California Press, 2008, part 1 (p.3-13).

Week 8

* Practicum on research design: A case study from start to “finish”
* *Assignment due before class:* ***Case Studies*** *(Individual)*

Readings to dive deeper (optional):

* Panger, Galen. “Why the Facebook Experiment is Lousy Social Science Research.” <https://medium.com/@gpanger/why-the-facebook-experiment-is-lousy-social-science-8083cbef3aee>
* de Vaus, David. *Research Design in Social Research*. Sage Publications, 2001, chapters 1–3.
* Heuer, Richard J., Jr. “Psychology of Intelligence Analysis.” Center for the Study of Intelligence, 1999, chapters 4–5, 8.
* Juliano, William. “Was Branch Rickey the Father of Sabermetrics?” *The Yankee Analysts* (March 28, 2011). <http://www.yankeeanalysts.com/2011/03/was-branch-rickey-the-father-of-sabermetrics-27771>.

## Conveying Findings (weeks 9 through 11)

The ability to skillfully convey one’s findings to others is critical to the organizational decision-making process. We encounter multiple types of audiences for our findings, from those well-versed in data science methods to those who may be unfamiliar or skeptical of such methods. Ultimately we need to be skilled storytellers. We’ll explore how data science can be a persuasive tool to change people’s attitudes and behaviors, with special attention to the application of information visualization techniques. We’ll examine case studies of persuasion gone bad, and how aspects of our research design and findings can be used, both appropriately or inappropriately, to persuade others.

Week 9

* Conveying findings through stories and visualizations
* Overview of storytelling and what makes a good story

Readings:

* McDonald, B. **Invisible Ink**: A Practical Guide to Building Stories That Resonate (Libertary, 2013).

Week 10

* Conveying findings to various audiences
* Revisiting the stories told in our earlier case studies
* Introduction to information visualization
* Analyzing information visualizations through history

Required readings:

* Offenhuber, Dietmar. “Visual Anecdote.” *Leonardo* 43, no. 4 (August 2010): 367–74.
* Tufte, Edward. *Visual Display of Quantitative Information*. Graphics Press, 2001. Skim the book and concentrate on a couple of specific areas that you find most interesting.

Week 11

* Persuasion and everyday life: changing attitudes and behaviors
* Examples of persuasion gone bad, including propaganda and misinformation
* The impact of research design on persuadability
* Wrap-up: conveying findings

Readings:

* Williams, Harold S. “Informing vs. Persuading.” *Innovating* 1, no. 2. The Rensselaer Institute.
* Kleiner, Art, and George Roth. "How to Make Experience Your Company's Best Teacher.” *Harvard Business Review* (September 1997).

Readings to dive deeper (optional):

* Gray, Jonathan, Liliana Bounegru, and Lucy Chambers. “Data Journalism in Perspective.” *The Data Journalism Handbook*. 2012. <http://datajournalismhandbook.org/1.0/en/introduction_4.html>.
* Laurila, David. “Jon “Boog” Sciambi: Broadcasting the Stats.” *FanGraphs Baseball* (March 12, 2012). <http://www.fangraphs.com/blogs/index.php/jon-boog-sciambi-broadcasting-the-stats/>.
* Steele, Julie, and Noah Iliinsky. *Beautiful Visualization: Looking at Data through the Eyes of Experts (Theory in Practice)*. O’Reilly Media, 2010.

## Future of Data Science (weeks 12 and 13)

We consider some important elements of the future of data science, as well as our individual roles in that future. As capabilities increase, what are some of the ethical and legal issues that we may encounter? Where will data scientists most effectively apply their skills; where will you apply yours?

Week 12

* The future of data science: More than *Minority Report*?
* Ubiquitous data science’s potential ethical and legal issues

Readings:

* Duhigg, Charles. “How Companies Learn Your Secrets.” *New York Times* (February 16, 2012). <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

Week 13

* Identifying your role in the future of data science: potential employers, roles, and projects.
* Working as a data scientist, including interviews with current data scientists

Required readings:

* Nonaka, Ikujiro. “The Knowledge-Creating Company.” *Harvard Business Review* (November 1991).

## Wrap-up (weeks 14 and 15)

In our final weeks we’ll summarize and review key takeaways and highlight how those insights can be used going forward. Week 14 introduces the course project.

Readings to dive deeper (optional):

* Enriquez, Juan, Gary P. Pisano, and Gaye L. Bok. “In Vivo to in Vitro to in Silico: Coping with Tidal Waves of Data at Biogen.” *Harvard Business School*, 2002.
* Matz, Eddie. “Saviormetrics.” ESPN, The Magazine (August 13, 2012). <http://espn.go.com/mlb/story/_/id/7602264/oakland-brandon-mccarthy-writing-moneyball-next-chapter-reinventing-analytics-espn-magazine>.
* Boudway, Ira. “Baseball: Running the New Numbers.” *Bloomberg Businessweek* (March 31, 2011). <http://www.businessweek.com/magazine/content/11_15/b4223072802462.htm>.
* Lewis, Peter H. “For the Love of the Technology, the Bay Area Is Reinventing Baseball (Again).” *The New York Times* (April 26, 2012). <http://www.nytimes.com/2012/04/27/us/for-the-love-of-the-technology-san-francisco-is-reinventing-baseball-again.html>.
* Neyer, Rob. “FIELDf/x Is Going to Change Everything.” ESPN (August 30, 2010). <http://espn.go.com/blog/sweetspot/post/_/id/5041/fieldfx-is-going-to-change-everything>.

Week 15

* Final exam, building on what your small teams learned in week 14

# Assignments and Projects

Coursework assignments will include a mix of short papers, brief presentations, and in-class debates. These individual and small-group projects will offer hands-on exercises of real-world decisions and events. For details, see the assignments document.

## 

## 

## Grading

Grades will be assessed on a combination of the following:

* 30 % is group projects (except final)
* 30% is individual projects (except case study)
* 10 % is the case study (week 8, as individuals)
* 10 % is final project (final weeks, as a group)
* 20% is participation/attendance

The grading scale is University-wide: <http://registrar.berkeley.edu/gradeskey.html>