Syllabus for RDADA

Research Design and Applications for Data and Analysis (RDADA)

Spring 2020

Course Designers:
Steve Weber & Andy Brooks

Instructors of Record:
Corey Jackson, Carlos Rivera, Mike Rivera, & Anna Zaitsev

Course Coordinator:
Mike Rivera

School of Information
University of California, Berkeley
Masters of Information and Data Science (MIDS)

This document outlines the weekly readings.
Your instructor will provide an assignments document.

Major revisions are pending until the first class. Minor revisions may be made up until a week before a scheduled class. Check the live syllabus every week to stay current.

Contents

Week 14 | Finals Week

Previous Term Archives

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 you will encounter and originate in industry. After an introduction to data science and an overview of the course, students will explore decision-making in organizations 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.

Course Coordinator

Your section instructor is your first point of contact. If you would like to chat with the course coordinator, Mike Rivera, please feel free to reach out via Slack @mikerivera. If you prefer email, you can reach him at michaelrivera@ischool.berkeley.edu

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 Amazon
- Darrell Huff and Irving Geis, How to Lie with Statistics, Publisher: W. W. Norton & Company; Reissue edition (October 17, 1993), ISBN 978-0393310726 Amazon
- Brian McDonald, Invisible Ink: A Practical Guide to Building Stories that Resonate, Publisher: Libertary Company (January 11, 2010), ISBN 978-0984178629 Amazon
- Knaflic, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals.* John Wiley & Sons, 2015. ISBN: 978-1119002253 Amazon

- Allison, Graham, and Philip Zelikow. *Essence of Decision, 2nd edition*. Longman, 1999. ISBN: 978-0321013491Amazon

Other Readings & VPN

Stared (*) readings are available in the Study.net course pack you will purchase. Other readings are available via link through the library. Please make sure to route traffic through the library VPN to ensure access. Info on how to set up the **VPN can be found here.** Select "Library Access and Full Tunnel" to access the content. Some readings may be found via a linked google drive file. To access these, please make sure you're logged into you Berkeley Google Suite.

Office hours

All students are encouraged to drop by and chat; even if they have no specific questions, it's a good time to reconnect and catch up. Regular office hours will be communicated by your instructor; feel free to request an individual appointment if you cannot attend.

Overview of Assignments

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, including class onboarding procedures, see the Assignments document distributed by your instructor.**

W201 has four types of assignments:

- Individual Assignments
- Group Presentations
- Final Group Project
- Discussion Questions (not graded)

Individual and group assignment will be graded. Discussion questions will help you digest the material and will prepare you for live session; however, they are not graded.

The breakdown of your final grade is:

- 30% individual assignments (3)
- 30% group projects (3)
- 20% final project

- 20% class engagement (This includes, but is not limited to, attendance, preparation for live session, participation in discussion, successful collaboration with your project team, and peer review).
    - Completion of the on-boarding is part of your participation grade.

Students can find their assignment grades on the I School Virtual Campus (ISVC). You will receive a letter grade at the end of the term. You instructor is solely responsible for assigning final grades.

Course Structure and Readings

Read & Watch Material Before Live Session

To ensure that we have a productive and informed conversation in live session, we expect you to read the required materials outlined in the syllabus, and watch the videos in the ISVC **before** we meet in live session.

Async Strategy

We recommend you watch the videos first, then read the articles, as videos provide important context for the readings. Also, your strategy to read academic literature may be different than how you read a novel. One recommended technique is to quickly read the abstract/intro, then skim the headings, tables and charts, and then read the conclusion. After you skim the reading in the manner described above, give the article a full read. This will ensure you know the punchline before you start. Good academic writing should make the main point clear.

The course is conceptualized in six sections, detailed below.

I. Introduction and Overview

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: More Than a Technical Discipline

- 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. **Link here (via Google Drive)**
- 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). **Link here (via Google Drive)**
- Le Grand, Julian, and Zack Cooper. "The Geeks Must Quash the Believers in Gut Instinct." *Financial Times* (February 21, 2012). https://www.ft.com/content/5a996db2-5c93-11e1-8f1f-00144feabdc0
- Davenport, Thomas H. "Competing on Analytics." *Harvard Business Review* (January 2006). **Link Here (via google drive)**
- *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 readings:

- DataHack Radio #11: Decision Intelligence with Google Cloud's Chief Decision Scientist, Cassie Kozyrkov. October 2018 https://www.analyticsvidhya.com/blog/2018/10/datahack-radio-decision-intelligence-google-cloud-cassie-kozyrkov/
- *Big Data Uncovered: What Does a Data Scientist Actually Do?* Forbes. 2016. http://www.forbes.com/sites/bernardmarr/2016/01/07/big-data-uncovered-what-does-a-data-scientist-really-do/
- datascience@berkeley. "Big Data Isn't a Concept — It's a Problem to Solve" *Berkeley Data Science Program* (2019). http://datascience.berkeley.edu/what-is-big-data/

II. Decision Making

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 2 | When Big Data Meets Big Decisions

- 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). **Missing image in article HERE** 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). http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=73552652&site=eds-live

Optional readings:

- Heuer, Richard J., Jr. "Psychology of Intelligence Analysis." Center for the Study of Intelligence, 1999, chapter 6. See left hand side "view in OskiCat" then click link below "link to online version" http://search.ebscohost.com/login.aspx?direct=true&db=cat04202a&AN=ucb.b18541235&site=eds-live
- Kroll, Joshua A., Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu. "Accountable Algorithms." *University of Pennsylvania Law Review*. Vol 165:633. http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9570&context=penn_law_review
- Loveman, Gary. "Diamonds in the Data Mine." *Harvard Business Review* (May 2003). http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9721850&site=eds-live (see "PDF Full Text" at left)
- Lewis, Michael. "Beane Counter." *Sports Illustrated* (May 12, 2003). https://www.si.com/vault/issue/703333/97
- Alamar, Benjamin, and Vijay Mehrotra. "Beyond 'Moneyball': The Rapidly Evolving World of Sports Analytics, Part I." *Analytics Magazine* (September–October 2011). http://analytics-magazine.org/beyond-moneyball-the-rapidly-evolving-world-of-sports-analytics-part-i/

Week 3 | High-Pressure Decision Making: People and Organization

- Cuban Missile Crisis case study
- Decision-making models, including analogical reasoning and analysis of competing hypotheses
- 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. download here

Week 4 | Biases in Decision Making

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

Required readings:

- Kahneman, Daniel. *Thinking, Fast and Slow*. Farrar, Strauss and Giroux, 2011, chapters 10–18.
- Stauffer, David. "How Good Data Leads to Bad Decisions." *Harvard Business Publishing Newsletters* (2002). http://search.ebscohost.com/login.aspx?direct=true&db=f5h&AN=8587204&site=eds-live
- Davenport, Thomas H. "Make Better Decisions." *Harvard Business Review* (November 2009). http://search.ebscohost.com/login.aspx?direct=true&db=edswss&AN=000271040500023&site=eds-live
- Hammond, John S., Ralph L. Keeney, and Howard Raiffa. "The Hidden Traps in Decision Making." *Harvard Business Review* (September–October 1998). http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=1040942&site=eds-live (see "PDF Full Text" at left)

Optional readings:

- Allison, Graham, and Philip Zelikow. *Essence of Decision, 2nd edition*. Longman, 1999, chapters 5 and 6.
- Heuer, Richard J., Jr. "Psychology of Intelligence Analysis." Center for the Study of Intelligence, 1999, chapters 2, 9–13. See left hand side "view in OskiCat" then click link below "link to online version" http://search.ebscohost.com/login.aspx?direct=true&db=cat04202a&AN=ucb.b18541235&site=eds-live
- Kahneman, Daniel, and Gary Klein. "Conditions for Intuitive Expertise: A Failure to Disagree." *American Psychologist* 64 (2009). https://search.proquest.com/docview/614512461/812520E976C042ECPQ/
- Kahneman, Daniel, and Amos Tversky. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (1979). Click "Full Text from JSTOR" http://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.10.2307.1914185&site=eds-live

- Tverksy, Amos, and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (1974). Click "Full Text from JSTOR" http://search.ebscohost.com/login.aspx?direct=true&db=rgr&AN=523865061&site=eds-live
- 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?"
- Verducci, Tom. "The Art of Winning an (Even More) Unfair Game." *Sports Illustrated* (September 26, 2011). https://www.si.com/vault/2011/09/26/106111997/the-art-of-winning-an-even-more-unfair-game
- Neyer, Rob. "Phillies Keep Winning without Your Fancy Numbers." *Baseball Nation* (March 2, 2012). http://www.sbnation.com/2012/3/2/2839053/phillies-keep-winning-without-your-fancy-numbers
- 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/2012/06/26/for-numbers-crunchers-minnesota-twins-old-school-methods-dont-add-up/

III. Research Design

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 | What Is Knowing?

(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:

- Anderson, Chris. "The End of Theory, Data Deluge Makes Scientific Method Obsolete." *Wired* (July 2008). https://www.wired.com/2008/06/pb-theory/
- *Burton, Robert. *On Being Certain*. St. Martin's Griffin, 2009, chapters 1 and 2.

- *Kuhn, Thomas. University of Chicago Press, 2012, chapter 12.

Optional reading:

- Engineering and Public Policy Committee on Science. *On Being a Scientist: A Guide to Responsible Conduct in Research*. National Academies Press, 2009. (Download as a guest) https://www.nap.edu/catalog/12192/on-being-a-scientist-a-guide-to-responsible-conduct-in

Week 6 | Practical Research Design for Real People

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

Required readings:

- *Creswell, John W. and J. David Creswell. *Research Design: Qualitative, Quantitative, and Mixed Methods*. Sage Publications, 2018, 5th Edition. Please read the chapters in the following order:
    - Ch 7, Research questions and hypotheses. Ch 1, The selection of a research design.

Optional readings:

- * Creswell, John W. and J. David Creswell: Ch 3, The use of theory. Ch 8, Quantitative methods.
- de Vaus, David. *Research Design in Social Research*. Sage Publications, 2001, chapters 1–3. https://books.google.com/books?id=9yurQt7T65oC&printsec=frontcover#v=onepage&q&f=false

Week 7 | Good Logic, Bad Logic, and Everything in Between

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

Required Readings:

- Huff, Darrell Huff, and Irving Geis. *How to Lie with Statistics*. W. W. Norton, 1993.
    - We will discuss chapters 1-4 and 7-10 in class. Chapters 5-6 are about visuals.

Optional readings:

- Rao, Venkatesh. "The Dangerous Art of the Right Question." *Trailblazers* (July 20, 2010). http://bobulate.com/post/838164346/the-dangerous-art-of-the-right-question for key summary, and the full PDF (a temporary link).
- Best, Joel. *Stat-Spotting: A Field Guide to Identifying Dubious Data*. University of California Press, 2008, part 1 (p.12-19 in 2013 updated edition. Link below).

http://search.ebscohost.com/login.aspx?direct=true&db=cat04202a&AN=ucb.b20847052&site=eds-live
Left hand side "view in OskiCat" then click "ebrary"

Week 8 | Research Design Case Study

- Practicum on research design: A case study from start to "finish"

Note: No readings are required this week to encourage you to make a strong finish for your Week 8 Case Study Assignment

Optional readings:

- Panger, Galen. "Why the Facebook Experiment is Lousy Social Science Research." https://medium.com/@gpanger/why-the-facebook-experiment-is-lousy-social-science-8083cbef3aee
- Heuer, Richard J., Jr. "Psychology of Intelligence Analysis." Center for the Study of Intelligence, 1999, chapters 4–5, 8. http://search.ebscohost.com/login.aspx?direct=true&db=cat04202a&AN=ucb.b18541235&site=eds-live
- Juliano, William. "Was Branch Rickey the Father of Sabermetrics?" *The Yankee Analysts* (March 28, 2011). http://www.captainsblog.info/2011/03/28/was-branch-rickey-the-father-of-sabermetrics/6097/

IV. Conveying Findings

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 | Storytelling Through Words and Pictures

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

Required reading:

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

Week 10 | How Visualizations Work

- 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:

- Knaflic, Cole Nussbaumer. *Storytelling with Data: A Data Visualization Guide for Business Professionals.* John Wiley & Sons, 2015.
    - Quickly click through this example of how to remove to improve.
- Offenhuber, Dietmar. "Visual Anecdote." *Leonardo* 43, no. 4 (August 2010): 367–74. http://search.ebscohost.com/login.aspx?direct=true&db=edspmu&AN=edspmu.S1530928210400043&site=eds-live
- Huff, Darrell Huff, and Irving Geis. *How to Lie with Statistics*. W. W. Norton, 1993.
    - Ch. 5-6.

Optional reading:

- Steele, Julie, and Noah Iliinsky. *Beautiful Visualization: Looking at Data through the Eyes of Experts (Theory in Practice)*. O'Reilly Media, 2010. http://berkeley.worldcat.org/title/beautiful-visualization-looking-at-data-through-the-eyes-of-experts/oclc/611902370&referer=brief_results
- If you would like an overview of some design fundamentals, we recommend Edward Tufte, Visual Display of Quantitative Information, Publisher: Graphics Press; 2nd edition (May 2001), ISBN 978-0961392147 *Amazon*. You may find the following sections most useful. Ch. 1: pp. 13, 51; Ch. 2: pp. 53, 76-7. Ch; 3, 5, 6, 9.

Week 11 | Persuasion in Business and Real Life

- 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

Required readings:

- Gallo, Carmine. "The Art of Persuasion Hasn't Changed in 2,000 Years." *Harvard Business Review*. (July 2019) https://hbr.org/2019/07/the-art-of-persuasion-hasnt-changed-in-2000-years
- *Williams, Harold S. "Informing vs. Persuading." *Innovating* 1, no. 2. The Rensselaer Institute.

Optional readings:

- Kleiner, Art, and George Roth. "How to Make Experience Your Company's Best Teacher." *Harvard Business Review* (September 1997). http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9709112726&site=eds-live
- 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/jon-boog-sciambi-broadcasting-the-stats/

V. Future of Data Science

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 | Data Science Futures & Life as a Data Scientist: Domains, Employers, Projects

- The future of data science: More than *Minority Report*?
- Ubiquitous data science's potential ethical and legal issues
- Identify your role in the future of data science: potential employers, roles, and projects.
- Work as a data scientist, including interviews with current data scientists

Required 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?pagewanted=print
- Nonaka, Ikujiro. "The Knowledge-Creating Company." *Harvard Business Review* (November 1991). http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=9201061306&site=eds-live

Optional reading:

- Marwick, Alice E., & boyd, danah. (2014). Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, 16(7), 1051–1067. http://doi.org/10.1177/1461444814543995

VI. Wrap-up

In our final weeks we'll summarize and review key takeaways and highlight how those insights can be used going forward.

Week 13 | Wrap-Up, Key Lessons, and the Path Forward

- Review
- Work-in-progress updates

Optional readings:

- *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). https://www.bloomberg.com/news/articles/2011-03-31/baseball-running-the-new-numbers
- 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?pagewanted=print
- Neyer, Rob. "FIELDf/x Is Going to Change Everything." ESPN (August 30, 2010). http://www.espn.com/blog/sweetspot/print?id=5041

Week 14 | Finals Week

- Build on what your small teams learned in Week 14. See assignments document for more details.

Previous Term Archives

- 2019 Fall
- 2019 Summer
- 2019 Spring
- 2018 Fall
- 2018 Summer
- 2018 Spring
- 2016 Fall
- 2016 Summer
- 2016 Spring
- 2015 Fall
- 2015 Summer
- 2015 Spring

- 2014 Fall