Understanding Behavior with Big Data

COURSE SYLLABUS

Instructor: Seth Stephens-Davidowitz

Email: seth.stephens@gmail.com

Prerequisites: No advanced mathematical training is required for this course, although having taken an introductory probability/statistics course may be useful.

1.2 Objectives

This course has two objectives. The first is to enable students to better understand studies that use data. By the end of this course, students will be able to make sense of data-rich studies, whether in academic journals, newspapers, or business presentations. Students will learn how to rigorously critique numerical arguments.

The second objective is to enable students to understand how data might help them answer questions they have during their careers. How do you determine the right data to collect or analyze? How do you find it? How do you know if the data can be trusted?

1.3 Course Structure

The course is divided into three parts.

The first part introduces the value of Big Data. This will explain the importance of rigorous data analysis -- and how the internet is making it easier and easier to collect data to help make important decisions.

The second part focuses exclusively on prediction. We will discuss how to best use data to predict what people will do in the future. Ultimately, data analysis is most valuable not when it can help you make sense of what already happened but if it can offer you new insights into what might happen. This is how you really improve decision-making.

The third part explores the weaknesses and pitfalls of data analysis. We will discuss ethical concerns Big Data can raise and the importance of using human intuition to guide data analysis. We will also focus on the value of incorporating many sources of data to correct blind spots in any one particular source.

1.4 Instruction Format

The course is primarily lecture based, although it includes some cases and activities. Class discussion is strongly encouraged.

1.5 Final Group Project

Students will think up a difficult decision that a corporation might face. They will discuss how data might be used to help with the decision. They will propose what data they would collect and how they would analyze it. They will present their proposal. The goal of the proposal would be to convince a potentially skeptical manager that their data collection and analysis plan would help make a better decision.

Course Outline

Number	Topic	Assignment/Readings Due
1	Introduction	
2	The Limitations of Human Intuition -- and the Need for Data	**Reading**: Ayers, Chapters 1-3, Stephens-Davidowitz, Chapters 1-2; Kahneman, Chapters 1-3;
3	The Value of New Sources of Data, Part I	**Reading**: Stephens-Davidowitz, Chapter 3 **Due**: Homework # 1
4	The Value of New Sources of Data, Part II	**Reading**: Dubner and Levitt, Chapters 1-3 **Reading**: Varian and Stephens-Davidowitz (2015)
5	Honest Data, Part I	**Reading**: Stephens-Davidowitz, Chapter 4
6	Honest Data, Part II	**Reading**: Rudder, Chapters 1-2
7	Personalization, Part I	**Reading**: Stephens-Davidowitz, Chapter 5 **Due**: Homework # 2
8	Personalization, Part II	**Reading**: Smith and Telang, Chapter 1
9	Correlation Versus Causation, Part I	**Reading**: Stephens-Davidowitz, Chapter 6 Stephens-Davidowitz, Varian, and Smith (2016) Levitt (1994)

10	Correlation Versus Causation, Part II (A/B Testing)	**Reading**: Christian (2012)
11	The Science of Prediction, Intro	**Reading**: Silver, Chapters 1-2 Due: Homework # 3
12	Sports Analytics	**Reading**: Silver, Chapter 3; Silver, Chapter 8
13	Data and Finance	**Reading**: Silver, Chapter 11
14	Data and Marketing	**Reading**: Duhigg (2012)
15	Data and Economic Output	**Reading**: Choi and Varian (2009)
16	Data and Politics	**Reading**: Gabriel and Stephens-Davidowitz (2016) **Prepare**: Visit and familiarize yourself with fivethirtyeight.com **Due**: Homework # 4
17	People Analytics	**Reading**: Bersin, Chapters 1-4
18	Midterm Review	
19	Midterm Exam	
20	The Dangers of Data, Intro	**Reading**: Stephens-Davidowitz, Chapter 6
21	Ethical Issues in Data Analysis	**Reading**: Stephens-Davidowitz, Chapter 7
22	Data Privacy	**Reading**: Stephens-Davidowitz, Chapter 8; Schneier, Chapters 1-3 **Due**: Homework # 5
23	Value of Small Data	**Reading**: Peysakhovich and Stephens-Davidowitz (2015)
24	An Unpredictable World	**Reading**: Taleb, Chapters 1-3
25	In-Class Group Project Preparation	
26	Group Presentations, 1	
27	Group Presentations, 2	
28	Conclusion	

DETAILED READING SCHEDULE

CLASS 1

Introductory Lecture; No Assigned Reading

CLASS 2

- Ian Ayers, *Supercrunchers*, 2007, Chapters 1-3
- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapters 1-2
- Daniel Kahneman, *Thinking Fast and Slow*, 2011, Chapters 1-3

CLASS 3

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 3

CLASS 4

- Stephen Dubner and Steven Levitt, *Freakonomics*, 2005, Chapters 1-3
- Hal Varian and Seth Stephens-Davidowitz, “A Hands-on Guide to Google Data,” 2015.

CLASS 5

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 4

CLASS 6

- Christian Rudder, *Dataclysm*, 2014, Chapters 1-2

CLASS 7

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 5

CLASS 8

- Michael D. Smith and Rahul Telang, *Streaming Sharing Stealing*, 2016, Chapter 1

CLASS 9

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 6
- Steven Levitt, “Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House,” *The Journal of Political Economy* 102(4), 1994.
- Seth Stephens-Davidowitz, Hal Varian, and Michael D. Smith, “Super Returns to Super Bowl Ads?” Working Paper, 2016.

CLASS 10

- Brian Christian, “The A/B Test: Inside the Technology that’s Changing the Rules of Business,” *Wired*, April 25, 2012.

CLASS 11

- Nate Silver, *The Signal and the Noise*, 2015, Chapters 1-2

CLASS 12

- Nate Silver, *The Signal and the Noise*, 2015, Chapters 3 and 8

CLASS 13

- Nate Silver, *The Signal and the Noise*, 2015, Chapter 11

CLASS 14

- Charles Duhigg, “How Companies Learn Your Secrets,” *New York Times*, February 16, 2012.

CLASS 15

- Hyunyoung Choi and Hal Varian, “Predicting the Present with Google Trends,” Working Paper, Google Inc., 2009.

CLASS 16

- Stuart A. Gabriel and Seth Stephens-Davidowitz, “If They Google You, Do You Win?” *New York Times*, October 23, 2016.

CLASS 17

- Josh Bersin, *The Training Measurement Book*, Chapters 1-4

CLASS 18

Midterm Review; No Reading

CLASS 19

Midterm; No Reading

CLASS 20

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 6

CLASS 21

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 7

CLASS 22

- Seth Stephens-Davidowitz, *Everybody Lies*, 2016, Chapter 8
- Bruce Schneier, *Data and Goliath*, 2015, Chapters 1-3

CLASS 23

- Alex Peysakhovich and Seth Stephens-Davidowitz, “How Not to Drown in Numbers,” *New York Times*, May 3, 2015.

CLASS 24

- Nassim Taleb, *The Black Swan*, 2010, Chapters 1-3

CLASS 25

Group Presentation Preparation; No Reading

CLASS 26

Group Presentations; No Reading

CLASS 27

Group Presentations; No Reading

CLASS 28

Concluding Lecture; No Reading