The History of Big Data
Professor Benjamin Schmidt
Email: bschmidt@neu.edu
Office Hours: Tuesdays 3-5pm, 219 Meserve Hall; or by appointment

Overview

This course helps students to understand contemporary excitement and fears about "Big Data" in a long historical context. Much is new about the way corporations, governments, and individuals use massive computational resources to search for patterns. But those who use big data draw on legacies from well before the computer age for data management, for structuring a complicated world into measurable quantities.

Course Themes

We will trace the long history of big data through four parallel strands:

1. The rise of massive systems of data collection by the government in the 19th century through institutions like the census and the military.

2. The attempts of American businesses to collect and use data to control their markets and their workers.

3. The turn to data by the sciences.

4. The development of computers from the 1940s on, and the ways that social forces shaped the development of computing.

Course Goals

In this course, you will:

1. Debate and describe the ways that contemporary practices of "Big Data" are shaped by and differ from a long historical context;

2. Interpret historical sources of data and recast them into a contemporary terms you and your peers can understand; and

3. Understand some of the major turning points in the history of computing, data collection, and social control.

Requirements

Classroom You must complete all the readings for the course and attend class prepared to discuss them.

Response Posts to Blackboard 6 times in the semester, you will post a short *response* to one or more of the readings for that day on Blackboard. These must be posted by 5pm the day *before* class meets so that your peers have time to read them. You must also write 6 *responses* to your peer's posts. Both posts and responses will be included as part of your participation grade.

Papers You will write one 5 to 7 page paper for this class, based on the readings; no outside research is expected.

Archival Project Mid-semester, we will take a trip to the university archives to look at some archival documents. You will write up a description of another document from elsewhere in the archives. This can take the form of a straightforward narration, or you can adapt the information

Final Project Final project assignments will be distributed in October, but you should start thinking early about which one you will want. It will consist of either 1) an 8-10 page paper in which you extend one of the weeks of the course with additional readings; or 2) a digital project in which you analyze a dataset created before the year 1994 using modern tools. In either case, you must discuss the project in advance with me.

Academic Integrity

You are expected to have read, and follow at all times, the University's Academic Integrity Policy.

Grading

Your grade will come from the following sources:

- Discussion: 35%

- Short Papers: 20%

- Blackboard Posts: 15%

- Final Project: 30%

The end of this syllabus includes a longer description of what sort of work will receive an "A," a "B," and so forth.

Schedule and Readings

Required for purchase: these may be placed in the University bookstore, but you will probably be better off ordering used or new copies online.

- Frederick Taylor, *The Principles of Scientific Management*

- Sarah Igo, *The Averaged American*

- Nicholas Lemann, *The Big Test*

- Tracy Kidder, *The Soul of a New Machine*

- Tim Berners-Lee, *Weaving the Web*

All other readings will be made available through Blackboard.

In addition to the readings listed below, some short primary sources may be added to Blackboard as the course develops to accompany a week's readings.

Week 1: Introductions

Week 2: Early Modern Information Overload

September 9

- Blair, "Reading Strategies for Coping With Information Overload ca.1550–1700"

- (In class): Plato, *Phaedrus*, on the invention of writing.

September 12

- Müller-Wille and Charmantier, "Natural history and information overload"

Week 3: Ordering the World

September 16

- Borges, *The Analytical Language of John Wilkins*

- Foucault, *The order of things*, Introduction and Chapter 3

September 19

- Try the Foucault again

- In class: Playfair, *The commercial and political atlas and Statistical breviary*

Week 4: 19th Century Records

September 23: Accounting

- Edwards, "Early Bookkeeping and Its Development into Accounting"

September 26: Slave economies

- Garvey and Gitelman, ""facts and FACTS" : Abolitionists' Database Innovations"

- Fogel and Engerman, *Time on the cross*, pp. 192-209,233-258

- Wilentz, *Major problems in the early republic, 1787-1848*, "GW Hammond, Instructions to his Overseer"

Week 5: Managed Information in the late nineteenth century

September 30: Industrial Revolutions

- Beniger, *The control revolution*, Chapter 6, "Industrial Revolution and the Crisis of Control"

- In Class: Dun and Bradstreet. J., "Mercantile Agencies.," J., "Mercantile Agencies."

October 3: The Census

- Kinnahan, "Charting Progress"

- Anderson, *The American census*, Chapters 3 to 5

- In Class: Blank Census forms, 1790-1940; and completed forms for Greenleaf St., Boston, 1940 Census.

First Paper Assignments handed out October 3.

Week 6: Sciences of Management

October 7: Taylorism

- Taylor, *The principles of scientific management*

- In class: punchcards and sorting.

October 10: Fordism

- Meyer, *The five dollar day*, Chapters 5 and 6

- In class: Chaplin, *Modern Times*, First fifteen minutes

Weeks 7 and 8: A Culture of Data

(No class October 14/Columbus Day)

First Papers Due Wed. October 16 at noon over Blackboard

October 17: Meritocracy

- Lemann, *The big test*, pp. 1-173

- In class: Plato, *The Republic of Plato*, Book III, 412b-415c

Week 8: A culture of data: Part II

October 21: Quantified Selves

- Igo, *The averaged American*, Introduction, Chapters 5 and 6, and Epilogue

- In class: Foucault, *The history of sexuality*, 157-159

October 24: Archival Trip

- Meet in University Archives for backstage tour and to see sample holdings.

Week 9: Early Computers

October 28: Imagining Computers

- Bush, "As We May Think"

- Bush, "Memex Revisited"

October 31: Gender and Computing

- *Cosmpolitan,* "The Computer Girls," 1967

- Light, "When computers were women"

- Film selections in class: Lang, *Desk Set*

Week 10: Computers in the Mainstream

November 4: The Mainframe Age

- Miller, *The assault on privacy*, 34-69,254-274

November 7: Computer Workers

- Kidder, *The soul of a new machine*

Week 11: Personal Computing

(No class November 11/Veterans' Day)

Data Exploration papers due November 13.

November 14: Personal computing

- Ted M. Lau, "Total Kitchen Information System", *Byte Magazine*, 1977

- Berners-Lee and Fischetti, *Weaving the Web*, Introduction; Chapters 1, 2, and 3

Week 12: The Age of Google

November 18: Information Overload revisited

- Gleick, *The information*, "After the Deluge"

- Vaidhyanathan, *The Googlization of everything*, Chapter 2

November 21: The emergence of "Big Data"

- The End of Theory, Wired Magazine, 2010 (Read the Introduction and pick two fields you are interested in to discuss in class)

- The Norvig-Chomsky Debate

Week 13: Surveillance and the state

November 25

- MacAskill and MacAskill, "NSA files decoded"

- Gellman and Soltani, "NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say"

- Doctorow, "Lockdown"

(No class November 28/Thanksgiving)

Week 14: Wrapup

December 4: Reflections

- Watch Edward Snowden's interview with Glenn Greenwald

- (In class) Foucault, *Discipline and punish*, "Panopticism," pp. 200-204

Optional Reading for catchup discussion posts

- (Optional catchup blackboard posts after class) Listen to "Photo Op," *This American Life* Episode 493

Final Projects Due December 10

Full Citations for Readings

Anderson, Margo J. *The American census: a social history* [in English]. New Haven: Yale University Press, 1988. ISBN: 0300040148 9780300040142 0300047096 9780300047097.

Beniger, James R. *The control revolution: technological and economic origins of the information society* [in English]. Cambridge, Mass.: Harvard University Press, 1986. ISBN: 0674169859 9780674169852 0674169867 9780674169869.

Berners-Lee, Tim, and Mark Fischetti. *Weaving the Web: the original design and ultimate destiny of the World Wide Web by its inventor* [in English]. San Francisco: HarperSanFrancisco, 1999. ISBN: 0062515861 9780062515865 006251587X 9780062515872.

Blair, Ann. "Reading Strategies for Coping With Information Overload ca.1550–1700." *Journal of the History of Ideas* 64 (1 2003): 11–28. ISSN: 1086-3222, doi:10.1353/jhi.2003.0014, http://dash.harvard.edu/bitstream/handle/1/3228379/blair%202003.pdf?sequence=2".

Borges, Jorge Luis. *The Analytical Language of John Wilkins*. Translated by Lilia Graciela Vázquez, 1999.

Bush, Vannevar. "As We May Think." *The Atlantic*, July 1945. ISSN: 1072-7825, http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

———. "Memex Revisited." In *From Memex to hypertext*, edited by James M. Nyce and Paul Kahn, 197–216. San Diego, CA, USA: Academic Press Professional, Inc., 1991. ISBN: 0-12-523270-5, http://dl.acm.org/citation.cfm?id=132180.132193.

Chaplin, Charles. *Modern Times*. In collaboration with Charles Chaplin, Charles Chaplin, Paulette Goddard, Henry Bergman, and Tiny Sandford. IMDB ID: undefined; IMDB Rating: undefined (undefined votes). February 25, 1936.

Computers, IRS, 1960s. In collaboration with Footage File. August 6, 2010. http://www.youtube.com/watch?v=WePY7OhZNOY&feature=youtube_gdata_player.

Doctorow, Cory. "Lockdown: The coming war on general-purpose computing." Boingboing. January 10, 2012. http://boingboing.net/2012/01/10/lockdown.html.

Edwards, James Don. "Early Bookkeeping and Its Development into Accounting." ARTICLETYPE: research-article / Full publication date: Winter, 1960 / Copyright © 1960 The President and Fellows of Harvard College, *The Business History Review* 34 (4 1960): 446–458. ISSN: 0007-6805, doi:10.2307/3111429, http://www.jstor.org/stable/3111429.

Ensmenger, Nathan. *The computer boys take over computers, programmers, and the politics of technical expertise* [in English]. Cambridge, Mass.: MIT Press, 2010. ISBN: 9780262289351 0262289350, http://books.google.com/books?id=nixSAQAAIAAJ.

Fogel, Robert William, and Stanley L Engerman. *Time on the cross: the economics of American negro slavery* [in English]. New York: Norton, 1989. ISBN: 0393306208 9780393306200 0393312186 9780393312188.

Foucault, Michel. *Discipline and punish: the birth of the prison*. New York: Pantheon Books, 1977. ISBN: 0394499425.

———. *The history of sexuality*. Translated by Robert Hurley. New York: Pantheon Books, 1978. ISBN: 0394417755.

———. *The order of things: an archaeology of the human sciences* [in English]. New York: Vintage Books, 1994. ISBN: 0679753354 9780679753353.

Garvey, Ellen Gruber, and Lisa Gitelman. ""facts and FACTS" : Abolitionists' Database Innovations." In *"Raw Data" is an oxymoron*, 89–102. Cambridge: MIT Press, 2013.

Gellman, Barton, and Ashkan Soltani. "NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say" [in en-US]. *The Washington Post*, October 30, 2013. ISSN: 0190-8286, http://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story_2.html.

———. "NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say" [in en-US]. *The Washington Post*, November 1, 2013. ISSN: 0190-8286, http://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html.

Gleick, James. *The information: a history, a theory, a flood* [in English]. New York: Pantheon Books, 2011. ISBN: 9780375423727 0375423729.

Igo, Sarah Elizabeth. *The averaged American: surveys, citizens, and the making of a mass public*. Cambridge, Mass: Harvard University Press, 2007. ISBN: 0674023218.

J., D. O. "Mercantile Agencies." *New York Daily Times* (New York, N.Y., United States), November 7, 1851, 1. http://search.proquest.com/hnpnewyorktimesindex/docview/95765241/abstract/142445A46F336CD6D70/11?accountid=12826.

———. "Mercantile Agencies." *New York Daily Times* (New York, N.Y., United States), October 29, 1851, 4. http://search.proquest.com/hnpnewyorktimesindex/docview/95772455/abstract/142445A46F336CD6D70/12?accountid=12826.

Kidder, Tracy. "The Computer Girls." *Cosmopolitan*, 1967, 52–56.

———. *The soul of a new machine* [in English]. Boston: Little, Brown, 1981. ISBN: 0316491705 9780316491709.

Kinnahan, Thomas P. "Charting Progress: Francis Amasa Walker's Statistical Atlas of the United States and Narratives of Western Expansion." <p>Volume 60, Number 2, June 2008</p>, *American Quarterly* 60 (2 2008): 399–423. ISSN: 1080-6490, doi:10.1353/aq.0.0012, http://muse.jhu.edu.ezproxy.neu.edu/journals/american_quarterly/v060/60.2.kinnahan.html.

Lang, Walter. *Desk Set*. In collaboration with Phoebe Ephron, Henry Ephron, Spencer Tracy, Katharine Hepburn, Gig Young, and Joan Blondell. IMDB ID: undefined; IMDB Rating: undefined (undefined votes). 20th Century Fox, May 1, 1957.

Lemann, Nicholas. *The big test: the secret history of the American meritocracy* [in English]. New York: Farrar, Straus / Giroux, 1999. ISBN: 0374299846 9780374299842.

Light, Jennifer. "When computers were women." *Technology and Culture* 40 (3 1999): 455.

Lubar, Steven. ""Do Not Fold, Spindle or Mutilate": A Cultural History of the Punch Card" [in en]. *Journal of American Culture* 15 (4 1992): 43–55. ISSN: 1542-734X, doi:10.1111/j.1542-734X.1992.1504_43.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1542-734X.1992.1504_43.x/abstract.

MacAskill, Ewen, and Ewen MacAskill. "NSA files decoded: Edward Snowden's surveillance revelations explained." The Guardian. November 1, 2013. http://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded.

Meyer, Stephen. *The five dollar day: labor management and social control in the Ford Motor Company, 1908–1921*. Suny series in American social history. Albany, N.Y: State University of New York Press, 1981. ISBN: 0873955080.

Miller, Arthur Raphael. *The assault on privacy: computers, data banks, and dossiers* [in eng]. In collaboration with Internet Archive. Ann Arbor, University of Michigan Press, 1971. http://archive.org/details/assaultonprivacyoomill.

Müller-Wille, Staffan, and Isabelle Charmantier. "Natural history and information overload: The case of Linnaeus." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1 2012): 4–15. ISSN: 1369-8486, doi:10.1016/j.shpsc.2011.10.021, http://www.sciencedirect.com/science/article/pii/S1369848611001130.

Plato. *The Republic of Plato*. Translated by 1930-1992. Bloom Allan David. New York: Basic Books, 1968. ISBN: 0465069355.

Playfair, 1759-1823., William. *The commercial and political atlas and Statistical breviary*. Edited by Howard Wainer and 1944- Spence Ian. New York: Cambridge University Press, 2005. ISBN: 0521855543.

Taylor, Frederick Winslow. *The principles of scientific management* [in English]. New York: Cosimo, 2006. ISBN: 1596058897 9781596058897.

"The Analytical Language of John Wilkins." http://www.alamut.com/subj/artiface/language/johnWilkins.html.

Vaidhyanathan, Siva. *The Googlization of everything: (and why we should worry)* [in English]. Berkeley: University of California Press, 2011. ISBN: 9780520258822 0520258827 9780520948693 0520948696 9780520272897 0520272897.

Wilentz, Sean. *Major problems in the early republic, 1787–1848: documents and essays* [in English]. Lexington, Mass.: D.C. Heath, 1992. ISBN: 0669243329 9780669243321.

Last revised 10/08/2013

Grading guidelines

Written Work

An **A** or **A**- thesis, paper, or exam is one that is good enough to be read aloud in a class. It is clearly written and well-organized. It demonstrates that the writer has conducted a close and critical reading of texts, grappled with the issues raised in the course, synthesized the readings, discussions, and lectures, and formulated a perceptive, compelling, independent argument. The argument shows intellectual originality and creativity, is sensitive to historical context, is supported by a well-chosen variety of specific examples, and, in the case of a research paper, is built on a critical reading of primary material.

A **B+** or **B** thesis, paper, or exam demonstrates many aspects of A-level work but falls short of it in either the organization and clarity of its writing, the formulation and presentation of its argument, or the quality of research. Some papers or exams in this category are solid works containing flashes of insight into many of the issues raised in the course. Others give evidence of independent thought, but the argument is not presented clearly or convincingly.

A **B**- thesis, paper, or exam demonstrates a command of course or research material and understanding of historical context but provides a less than thorough defense of the writer's independent argument because of weaknesses in writing, argument, organization, or use of evidence.

A **C+**, **C**, or **C**- thesis, paper, or exam offers little more than a mere a summary of ideas and information covered in the course, is insensitive to historical context, does not respond to the assignment adequately, suffers from frequent factual errors, unclear writing, poor organization, or inadequate primary research, or presents some combination of these problems.

Whereas the grading standards for written work between A and C- are concerned with the presentation of argument and evidence, a paper or exam that belongs to the D or F categories demonstrates inadequate command of course material.

A **D** thesis, paper, or exam demonstrates serious deficiencies or severe flaws in the student's command of course or research material.

An **F** thesis, paper, or exam demonstrates no competence in the course or research materials. It indicates a student's neglect or lack of effort in the course.

Discussions and Seminars

A student who receives an **A** for participation in discussion in discussions or seminars typically comes to every class with questions about the readings in mind. An 'A' discussant engages others about ideas, respects the opinions of others, and consistently elevates the level of discussion.

A student who receives a **B** for participation in discussion in discussions or seminars typically does not always come to class with questions about the readings in mind. A 'B' discussant waits passively for others to raise interesting issues. Some discussants in this category, while courteous and articulate, do not adequately listen to other participants or relate their comments to the direction of the conversation.

A student who receives a **C** for discussion in discussions or seminars attends regularly but typically is an infrequent or unwilling participant in discussion.

A student who fails to attend discussions or seminars regularly and adequately prepared for discussion risks the grade of **D** or **F**.

–Taken from the department of history at Princeton University.