

## LITERARY DATA: SOME APPROACHES

Thursdays, 1:10–4:10 p.m. in Murray 305 (Plangere Seminar Room)

Professor Andrew Goldstone ([andrew.goldstone@rutgers.edu](mailto:andrew.goldstone@rutgers.edu))

<http://www.rci.rutgers.edu/~ag978/litdata>

Office hours: TBA

### COURSE DESCRIPTION

In the last ten years, the strange quasi-disciplinary formation known as *DH* or *Digital Humanities* has renewed the struggle over methods in literary studies. Analyses of digitized texts using computer-assisted techniques promise to transform the kinds of evidence, the methods of interpretation, and the modes of argument which matter to literary scholarship. *Data* is now a subject of energetic debate in literary studies: what constitutes literary data, and how should it be analyzed and interpreted? How might aggregation and quantification produce new knowledge in literary scholarship? What methods are most appropriate for grappling with the enormous, and enormously messy, world of digitized literary texts and data about literature?

This course pursues two aims in parallel: to engage with the history and current practice of literary data analysis, and to introduce the foundational skills of literary data analysis in the R programming language. Class time will be divided between seminar and practical instruction. The seminar discussions trace theoretical debates about literary data from structuralism and scientific bibliography, to experiments in computational stylistics, to contemporary scholarly controversies in and around DH. The practicum surveys the fundamentals of programming and data manipulation, with an introduction to selected numerical techniques and data visualizations. Short homework exercises supplement the in-class instruction, with an emphasis on handling actual literary data of various kinds.

There are two major assignments. A short position paper on a theoretical question about literary data and DH is due at midterm. The final assignment is to plan, carry out, and report on a small-scale project in literary data analysis. This project is to be undertaken in small groups; the report will detail methods and interpretations together with code and data.

No special technical expertise of any kind is expected; instruction begins from first principles. However, the work of programming does require willingness to experiment, patience in the face of frustration, and the nerve to ask for help as often as needed.

Bring your own laptop to class, if you have one; loaner laptops will also be available for in-class workshops. MacOS X and Linux are the preferred operating systems for work in the course, but Windows will be accommodated as well.

## LEARNING GOALS

1. Engage critically, through discussion and formal writing, with contemporary debates about computational methods in literary study.
2. Understand contemporary discussions of the digital humanities in the context of arguments about data in the human sciences in the last fifty years.
3. Develop basic competence in analyzing data using the R environment, including obtaining, reformatting, tabulating, and visualizing multiple forms of data.
4. Understand the fundamentals of computation by mastering the most important constructs in the R programming language.
5. Gain practical experience in analyzing literary data for the purpose of answering research questions in literary scholarship through a collaborative data-analysis research project.

This course fulfills the department's B distribution requirement.

## READING LIST (SUBJECT TO CHANGE)

- Blei, David M. "Probabilistic Topic Models." *Communications of the ACM* 55, no. 4 (April 2012): 77–84. <http://www.cs.princeton.edu/~blei/papers/Blei2012.pdf>.
- Bourdieu, Pierre. *The Rules of Art: Genesis and Structure of the Literary Field*. Translated by Susan Emanuel. Cambridge: Polity Press, 1996. Excerpt on Sakai.
- Burrows, J. F. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon, 1987. Excerpt on Sakai.
- Cecire, Natalia. "Ways of Not Reading Gertrude Stein." *ELH* (2014).
- Chambers, John M. *Software for Data Analysis: Programming with R*. New York: Springer, 2008. <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1007/978-0-387-75936-4>. Not required. An advanced reference on the programming language.
- Clement, Tanya E. "'A thing not beginning and not ending': Using Digital Tools to Distant-Read Gertrude Stein's *The Making of Americans*." *Literary and Linguistic Computing* 23, no. 3 (September 2008): 361–81.
- Darnton, Robert. "Reading, Writing, and Publishing in Eighteenth-Century France: A Case Study in the Sociology of Literature." *Daedalus* 100, no. 1 (1971): 214–56. <http://www.jstor.org/stable/20023999>.

- DiMaggio, Paul, Manish Nag, and David Blei. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41, no. 6 (December 2013). <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1016/j.poetic.2013.08.004>.
- Elson, David, Nicholas Dames, and Kathleen McKeown. "Extracting Social Networks from Literary Fiction." In *ACL 2010*, 138–47. 2010. <http://www.aclweb.org/anthology/P10-1015>.
- Folsom, Ed. "Database as Genre: The Epic Transformation of Archives." *PMLA* 122, no. 5 (October 2007): 1571–1579. <http://www.mlajournals.org.proxy.libraries.rutgers.edu/doi/abs/10.1632/pmla.2007.122.5.1571>.
- Gitelman, Lisa, ed. *"Raw Data" Is an Oxymoron*. Cambridge: MIT Press, 2013. Not required, except for Rosenberg's chapter, but an important collection.
- Goldstone, Andrew, and Ted Underwood. "Quiet Transformations: A Topic Model of Literary Studies Journals." 2014. <http://www.rci.rutgers.edu/~ag978/quiet>. This site accompanies the authors' essay.
- . "The Quiet Transformations of Literary Studies: What Thirteen Thousand Scholars Could Tell Us." *NLH* 45, no. 3 (Summer 2014): 359–84. [http://muse.jhu.edu.proxy.libraries.rutgers.edu/journals/new\\_literary\\_history/v045/45.3.goldstone.pdf](http://muse.jhu.edu.proxy.libraries.rutgers.edu/journals/new_literary_history/v045/45.3.goldstone.pdf).
- Gries, Stefan Thomas. *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York: Routledge, 2009. Excerpt on Sakai.
- Grimmer, Justin, and Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21, no. 3 (Summer 2013): 267–97. <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1093/pan/mpso28>.
- Griswold, Wendy. "Number Magic in Nigeria." *Book History* 5, no. 1 (2002): 275–82. <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1353/bh.2002.0008>.
- Healy, Kieran, and James Moody. "Data Visualization in Sociology." *Annual Review of Sociology* 40, no. 1 (2014). <http://www.annualreviews.org.proxy.libraries.rutgers.edu/doi/abs/10.1146/annurev-soc-071312-145551>.
- Hockey, Susan. "The History of Humanities Computing." Chap. 1 in Schreibman, Siemens, and Unsworth, *A Companion to Digital Humanities*. Available online.
- Hoover, David. "Quantitative Analysis and Literary Studies." Chap. 28 in Schreibman, Siemens, and Unsworth, *A Companion to Digital Literary Studies*. Available online.
- Jockers, Matthew L. "Metadata." Chap. 5 in *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press, 2013. Excerpt on Sakai. The rest of the book, though not required, gives a striking range of different kinds of quantitative text analyses.

- Jockers, Matthew L. *Text Analysis with R for Students of Literature*. New York: Springer, 2014. ISBN: 9783319031637. <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1007/978-3-319-03164-4>. This textbook is available to you in PDF through the library. You are likely to find it easier to work from in paper. Springer sells both an ordinary hardcover and a print-on-demand softcover.
- Kirschenbaum, Matthew. "What Is 'Digital Humanities,' and Why Are They Saying Such Terrible Things about It?" *differences* 25, no. 1 (2014): 46–63. <http://differences.dukejournals.org.proxy.libraries.rutgers.edu/content/25/1/46>.
- Klein, Lauren F. "The Image of Absence: Archival Silence, Data Visualization, and James Hemings." *American Literature* 85, no. 4 (December 2013): 661–88. <http://americanliterature.dukejournals.org.proxy.libraries.rutgers.edu/content/85/4/661>.
- Krippendorff, Klaus. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: SAGE, 2013. Excerpt on Sakai.
- Lévi-Strauss, Claude. "The Structural Study of Myth." *The Journal of American Folklore* 68, no. 270 (October 1955): 428–444. <http://www.jstor.org.proxy.libraries.rutgers.edu/stable/536768>.
- Liu, Alan. "The Meaning of the Digital Humanities." *PMLA* 128, no. 2 (2013): 409–23. <http://www.mlajournals.org/doi/abs/10.1632/pmla.2013.128.2.409>.
- Lunenfeld, Peter, Anne Burdick, Johanna Drucker, Todd Presner, and Jeffrey Schnapp, eds. *Digital Humanities*. Open access ed. Cambridge: MIT Press, 2012. [http://mitpress.mit.edu/sites/default/files/titles/content/9780262018470\\_Open\\_Access\\_Edition.pdf](http://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf).
- Manovich, Lev. "What Is Visualisation?" *Visual Studies* 26, no. 1 (2011): 36–49. <http://dx.doi.org.proxy.rutgers.edu/10.1080/1472586X.2011.548488>.
- McCarty, Willard. "Knowing...: Modeling in Literary Studies." Chap. 21 in Schreibman, Siemens, and Unsworth, *A Companion to Digital Literary Studies*. Available online.
- McGann, Jerome J. *Radiant Textuality: Literature after the World Wide Web*. New York: Palgrave, 2001. Excerpt on Sakai.
- McKenzie, D. F. "Printers of the Mind: Some Notes on Bibliographical Theories and Printing-House Practices." *Studies in Bibliography* 22 (1969): 1–75. <http://www.jstor.org.proxy.libraries.rutgers.edu/stable/40371475>. Only an excerpt of this long essay is assigned.
- McPherson, Tara. "Why Are the Digital Humanities So White? or Thinking the Histories of Race and Computation." In *Debates in the Digital Humanities*, open access ed., edited by Matthew K. Gold. Minneapolis: University of Minnesota Press, 2013. <http://dhdebates.gc.cuny.edu/debates/text/29>.

- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Joseph P. Pickett, Dale Hoiberg, et al. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331, no. 6014 (2011): 176–182. <http://www.sciencemag.org.proxy.libraries.rutgers.edu/content/331/6014/176>.
- Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso, 2005. ISBN: 9781844671854. Available at the bookstore.
- . "Style, Inc. Reflections on Seven Thousand Titles (British Novels, 1740–1850)." *Critical Inquiry* 36, no. 1 (Autumn 2009). <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1086/606125>.
- Navarro, Dan. *Learning Statistics with R*. <http://health.adelaide.edu.au/psychology/ccs/teaching/lsr/>. Not required. This statistics textbook, free online, includes a good introductory discussion of R.
- Petzold, Charles. *Code: The Hidden Language of Computer Hardware and Software*. Redmond, WA: Microsoft Press, 2000. Excerpt on Sakai.
- Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. Urbana: University of Illinois Press, 2011. Excerpt on Sakai.
- Rosenberg, Daniel. "Data before the Fact." Chap. 1 in Gitelman, "*Raw Data*" Is an Oxymoron. Available on Sakai.
- Schmidt, Benjamin M. "Words Alone: Dismantling Topic Models in the Humanities." *Journal of Digital Humanities*. 2, no. 1 (Winter 2012). <http://journalofdigitalhumanities.org/2-1/words-alone-by-benjamin-m-schmidt/>.
- Schreibman, Susan, Ray Siemens, and John Unsworth, eds. *A Companion to Digital Humanities*. Oxford: Blackwell, 2004. <http://www.digitalhumanities.org/companion/>. Not required, except for Hockey's essay on the history of digital humanities. Overlaps substantially with the same editors' 2008 anthology.
- , eds. *A Companion to Digital Literary Studies*. Hardcover. Oxford: Blackwell, 2008. <http://www.digitalhumanities.org/companionDLS/>. Not required, except for the chapters by Hoover and McCarty. This anthology, freely available online, gives a somewhat dated overview of literary humanities computing.
- So, Richard Jean, and Hoyt Long. "Network Analysis and the Sociology of Modernism." *boundary 2* 40, no. 2 (June 2013): 147–182. <http://boundary2.dukejournals.org.proxy.libraries.rutgers.edu/content/40/2/147>.
- Spector, Phil. *Data Manipulation with R*. New York: Springer, 2008. <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1007/978-0-387-74731-6>. Not required. A further reference on R.

- Tector, Paul. *R Cookbook*. Sebastopol, CA: O'Reilly, 2011. ISBN: 9780596809157. Strongly recommended: a clear reference with lots of examples.
- Trumpener, Katie. "Paratext and Genre System: A Response to Franco Moretti." *Critical Inquiry* 36, no. 1 (2009): 159–71. <http://www.jstor.org.proxy.libraries.rutgers.edu/stable/10.1086/606126>.
- Underwood, Ted. "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago." *Representations* 127, no. 1 (Summer 2014): 64–72. <http://www.jstor.org.proxy.libraries.rutgers.edu/stable/10.1525/rep.2014.127.1.64>.
- "What Is Data in Literary Studies?" *Arcade* (January 16–20, 2014). <http://arcade.stanford.edu/content/what-data-literary-studies-1>. Not required. A collection of statements from an MLA 2014 Roundtable.
- Wickham, Hadley. *Advanced R*. Chapman & Hall, 2014. <http://adv-r.had.co.nz/>. Not required.
- . *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer, 2009. <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1007/978-0-387-98141-3>.
- Wilkins, Matthew. "The Geographic Imagination of Civil War-Era American Fiction." *American Literary History* 25, no. 4 (2013): 803–840. [http://muse.jhu.edu.proxy.libraries.rutgers.edu/journals/american\\_literary\\_history/v025/25.4.wilkins.html](http://muse.jhu.edu.proxy.libraries.rutgers.edu/journals/american_literary_history/v025/25.4.wilkins.html).
- Wilkinson, Leland. *The Grammar of Graphics*. 2nd ed. New York: Springer, 2005. <http://dx.doi.org.proxy.libraries.rutgers.edu/10.1007/0-387-28695-0>.