LITERARY DATA: SOME APPROACHES

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


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](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.


