Data Mining in the Humanities

Wednesdays, 2:50-4:10 p.m., Honors College, room S126

Francesca Giannetti, Digital Humanities Librarian, Alexander Library
Office Hours: Tuesday, 10:00 a.m. - 12:00 p.m., in Digital Humanities Lab
francesca.giannetti@rutgers.edu
(848) 932-6097

Course Description: Popular media often portray “big data” as the exclusive province of information scientists, but data collection in the humanities can swiftly exceed the capacity of the human brain to analyze. Increasingly, humanists turn to digital tools to conduct quantitative research on literary texts, websites, tweets, images and sound recordings. How does one create or reuse a humanities data set? What tools are used to store, manipulate and process that data? How does one begin to analyze humanities research data and share findings in the form of visualizations? This course will explore some methodologies of quantitative analysis in the humanities using free and open source digital tools to yield insights into data that would otherwise be difficult to obtain. Through lectures, discussion, labs, and a digital final project, students will familiarize themselves with the tools of digital humanities scholarship and learn to form arguments on the basis of a few simple computational techniques.

Prerequisite(s): None

Credit Hours: 1, pass/no credit

Course Objectives:

At the completion of this course, students will be able to

1. evaluate the use of humanities data (and metadata) through readings and by looking critically at several scholarly digital projects;
2. apply several methods for collecting, processing and presenting humanities data;
3. understand some of the technical, ethical, and social factors that influence which data gets produced and analyzed, and by whom.
Grade Distribution:

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussion</td>
<td>30%</td>
</tr>
<tr>
<td>Labs</td>
<td>40%</td>
</tr>
<tr>
<td>Final Project</td>
<td>30%</td>
</tr>
</tbody>
</table>

Course Policies:

- **Attendance**
  - This course will involve lots of active, engaged participation in class discussions and labs. We won't have many lectures and there is no final exam. You should come to class having read the required reading and prepared to discuss it with your classmates. Maintaining an animated class discussion also requires that you be *present*, both physically and mentally. This means
  * Arriving on time;
  * Being attentive to the conversation;
  * Preparing assigned readings before the start of class;
  * Bringing assigned readings to class. If we are reading online articles, load them on your workstation or device, or print them and bring that hard copy.
  - Byrne seminars meet for only ten weeks. You may miss one class without penalty, but if you miss two, your discussion grade will be affected.
  - It is your responsibility to find out what you missed and to make up any pertinent assignments. If you miss a lab due to an excused absence you should attempt to make up that work.

- **Digital Etiquette**
  - Phones, tablets and other mobile devices should be muted and stowed away for the duration of class. If your phone rings once during class, I will ask you to turn it off. If it happens a second time, I may ask that you leave the classroom, which will count as an absence.
  - You may use a laptop during this course, and indeed you will need it for the digital labs (we may move the class session to a computer lab, if necessary). Please don’t use your device to text, follow a game, post on Reddit, or check your friend’s Tumblr. This distracting behavior affects not only you, but also those around you. There is one big exception to the social media rule—you are allowed to tweet—provided that the tweet relate to our discussions and that you use the class hashtag #humdata17.
  - This course relies heavily on access to computers, free and/or open source software applications, and the Internet. At some point during the semester you will have a problem with technology. Just consider it a given that, with no preamble and little provocation, your hard drive will fail, a server will go down, a piece of software will not behave the way you expect it to, or something else will go wrong. Now is the time to learn some good habits to safeguard your work. Start assignments early and SAVE SAVE SAVE often. Always keep a backup copy of your work, stored some place secure, and ideally remotely accessible like in a Dropbox or Google Drive folder. It is your responsibility to ensure that your work is not irrecoverably lost. If your laptop crashes, check one out at the libraries or use a computer lab. I will not grant extensions based on last-minute problems with technology.

- **Grades**
  - If you participate in class and turn in all assignments on time, you won’t have difficulty passing this course. There will be one mid-semester grade check-in, allowing you to track your progress, in which I’ll report the assignments I have collected for you.
• Learning Support

  – Rutgers has Learning Centers on each campus where any student can get tutoring and other help. Rutgers also has a Writing Program where students can get assistance with writing skills and assignments.

• Students with Disabilities

  – Students with disabilities may request academic accommodations from the Office of Disability Services. Phone: (848) 445-6800. Website: https://ods.rutgers.edu.

Academic Integrity
In addition to knowledge and skills, Rutgers University aims to teach students appropriate ethical and professional standards of conduct. The Academic Integrity Policy exists to inform students and faculty of their obligations in upholding the highest standards of professional and ethical integrity. All student work is subject to this policy. Professional and academic practice provides guidance about how to properly cite, reference, and attribute the intellectual property of others.

The principles of academic integrity require that a student:

  • properly acknowledge and cite all use of the ideas, results, or words of others.
  • properly acknowledge all contributors to a given piece of work.
  • make sure that all work submitted as his or her own in a course or other academic activity is produced without the aid of impermissible materials or impermissible collaboration.
  • obtain all data or results by ethical means and report them accurately without suppressing any results inconsistent with his or her interpretation or conclusions.
  • treat all other students in an ethical manner, respecting their integrity and right to pursue their educational goals without interference. This requires that a student neither facilitate academic dishonesty by others nor obstruct their academic progress.
  • uphold the canons of the ethical or professional code of the profession for which he or she is preparing.

The full text of the Rutgers Academic Integrity Policy may be found at http://academicintegrity.rutgers.edu/academic-integrity-policy/.

Course Schedule:
The weekly coverage might change depending on the progress of the class.
| Week       | Content                                                                                                                                                                                                                                                                                                                                 |