**Digital Methods (syllabus by Donald Sturgeon)**

**Instructor:** Donald Sturgeon

**E-mail:** [sturgeon@fas.harvard.edu](mailto:sturgeon@fas.harvard.edu)

|  |
| --- |
| * **CHNSHIS 202** * **Harvard College/Graduate School of Arts and Sciences:** 161272 * **Term:** Spring 2015-2016 * **Course Instructors:** Donald Sturgeon. * **Location:** CGIS South S001 (FAS) * **Meeting Time:** Monday 1:00pm - 3:59pm * **Exam Group:** FAS15\_A * **Course Description:** This course introduces graduate students to programming skills and digital humanities techniques of direct practical relevance to research in their discipline. It will consist of weekly lectures, each introducing a specific type of technique, followed by an interactive lab session during which students practice applying the technique to data appropriate to their own research. No background in digital methods or programming is assumed, but students are expected to have basic computing skills and are required to bring a suitable laptop to use during the lab sessions. The techniques covered in this course all have broad applicability and students will be expected to apply them to their own research topics and relevant texts as arranged during the first few sessions. The course will end with student presentations in which students apply an appropriate selection of the techniques studied to their own research questions. While examples and coursework will draw upon Chinese language source materials, students working with other languages are also encouraged to take this course. * **Notes:** Enrollment limited to 16. This course will be taught by Donald Sturgeon, Postdoctoral Fellow in Chinese Digital Humanities and Social Sciences at the Fairbank Center for Chinese Studies, Harvard University. |

**Schedule**

**Week 1 (Jan 25) - Introduction**

* Background and basic concepts
* Representing text on a computer
* Setting up the Python environment

**Week 2 (Feb 1) - Introduction to programming**

* Variables, functions, loops, and files
* Reading: [An Informal Introduction to Python (Links to an external site.)](https://docs.python.org/3/tutorial/introduction.html" \t "_blank), Python Software Foundation.

**Week 3 (Feb 8) - Regular expressions**

* String manipulation and data extraction.
* Reading: [Regexone (Links to an external site.)](http://regexone.com/" \t "_blank), interactive tutorial, lessons 1-14

***Feb 15 - University holiday, no class***

**Week 4 (Feb 22) - Working with structured data**

* Associative arrays, tables, CSV files
* Reading: Matthew L. Jockers. [Macroanalysis: Digital Methods and Literary History](http://id.lib.harvard.edu/aleph/013629251/catalog), University of Illinois Press 2013. Pages 3-32.

**Week 5 (Feb 29) - Practical data manipulation**

* Automated extraction of data from the web

**Week 6 (Mar 7) - Textual similarity**

* Introduction to information retrieval

***Mar 14 - Recess week, no class***

**Week 7 (Mar 21) - Topic modeling**

* Generating and interpreting data using Mallet
* Readings: [Where to start with text mining (Links to an external site.)](http://tedunderwood.com/2012/08/14/where-to-start-with-text-mining/" \t "_blank), Ted Underwood. [Topic Modeling for Humanists: A Guided Tour (Links to an external site.)](http://www.scottbot.net/HIAL/?p=19113" \t "_blank), Scott Weingart.

**Week 8 (Mar 28) - Network visualization with Gephi**

* Representing data as a network graph
* David Easley and Jon Kleinberg, [Networks, Crowds, and Markets: Reasoning About a Highly Connected World (Links to an external site.)](http://www.cs.cornell.edu/home/kleinber/networks-book/" \t "_blank), Cambridge University Press 2010, Chapter 2, p. 23-46.

**Week 9 (Apr 4) - Principal component analysis**

* Exploratory data analysis in Python
* Reading: Ian H. Witten, Eibe Frank and Mark A. Hall, [Data Mining: Practical Machine Learning Tools and Techniques](http://id.lib.harvard.edu/aleph/013335825/catalog), Morgan Kaufmann, 2011. Sections 1.1-1.3 and 2.1-2.3.

**Week 10 (Apr 11) - Machine learning**

* Features, classification, regression

**Week 11 (Apr 18) - Review and discussion**

* What worked, what didn’t, and why
* Debugging of issues arising during project work

**Week 12 (Apr 25) - Student presentations and discussion**