INFO/CS 3300: Data-driven Web Applications
INFO 5100: Visual Analytics for the Web
Spring 2016
Prof. David Mimno

Location: Olin 155
Time: MWF 11:15--12:05
Credits: 3
Websites: http://mimno.infosci.cornell.edu/info3300/, CMS, Piazza
Prereq: CS 2110 (or familiarity with object-oriented programming), INFO 2300 (or experience developing interactive web pages with HTML, CSS, and Javascript)

Course Description

The web has become an outstanding environment for telling stories with data. This course will cover technologies for representing, modeling and displaying data in the context of interactive web pages. Practical skills for building web pages will be mixed with data mining algorithms and theoretical approaches to graphics. We will use the D3 Javascript library to build both static and dynamic visualizations. Students will also learn functional programming style in Javascript and scalable vector graphics (SVG) as necessary to use D3. We will introduce a number of popular data mining models and algorithms such as Naïve Bayes classifiers, k-means clustering, and network layout methods, which we will incorporate into web visualizations.

Contact Information

Email Prof. Mimno for course administrative questions. Post to Piazza about course content questions.

Professor:
David Mimno, Gates 205. 607-255-8919. Best contact method is email mimno@cornell.edu. I will reply within 24 hours.

Head TA:
Moontae Lee. moontae@cs.cornell.edu.

TAs:

Grading

Grades will be based on attendance (10%), weekly homework (40%), two open-ended group projects (15% each), and a take-home final exam (for undergraduates in 3300) OR a third
group project (for masters students in 5100) (20%). Homework will be assigned on
Mondays, due at midnight the following Monday, and returned the Monday after. Work will
be turned in through CMS. Regrade requests should go in writing to mimno@cornell.edu.
The first project will be to design a static (non-interactive) web visualization. The second
project will be to design an interactive visualization. Each group will be assigned to a TA.
You will send weekly progress reports to your TA listing what each team member is
responsible for and what has been accomplished in the past week, and flagging any
problems or questions.

Absences and late/missing work

Class time will mix lecture, discussion, and hands-on programming exercises, so attendance
is important. If you will be absent, write to mimno@cornell.edu with an explanation. Late
homework will not be accepted, but your two lowest homework grades will be discarded. In
the unlikely event that you are having difficulty with CMS, work received by email before the
deadline will be accepted.

In-class work

Web programming is a complicated skill, and there's no way to learn without practice. Each
class will involve a daily programming problem that we will work on together. Template
HTML files will be made available before class.

Laptops

In order to facilitate interactive in-class work, you are allowed to bring a laptop. If you don't
have one or don't choose to bring one, work with someone sitting near you. If you have a
laptop, you will be expected to use it for relevant work. "Multitasking" is a myth.
Distractions limit your ability to learn, and the ability of those near you. If your laptop is
open, expect to show the results of your in-class work, or to have thoughtful questions.

Academic Integrity

We will follow university policies as outlined in the Academic Integrity Handbook. You are
encouraged to discuss homework, but each student will complete assignments alone.

Using other people's code is an important part of programming, but for group projects the
code should be substantially the work of the group members except for standard libraries
such as D3, lodash, and jQuery. Any code used in projects that was not written by the group
members should be placed in separate files and clearly labeled with their source URLs. If
you have benefitted from online resources such as examples or StackOverflow answers, list
the URLs in comments in your own code, even if you did not directly copy anything.

Project work that relates to your other classes or research is encouraged, but you may not
recycle assignments. There must be no doubt that the work you turn in for this class was done for this class.

**Students with Disabilities**

We will make every effort possible to ensure that the class works for all students. Contact Prof. Mimno if there is anything we should know about. If there is a specific event such as an exam that you are concerned about, please inform us at least two weeks in advance so that we have time to make arrangements.

**SONA Credits**

Many researchers on campus need participants for user studies and other types of experiments. The SONA system allows you to register for such studies. You will get 0.5 percent extra for this course, up to a maximum of 2.0 percent, for each 30 minute study (or equivalent). Participating in studies is a great way to find out what real research looks like. To register, go to this URL:

https://cornell-comm.sona-systems.com

**Course Outline**

The following is a tentative course outline, subject to change.

Week 1, Jan 27 (WF) Course objectives. Web environments. Javascript, JSON
Week 2, Feb 1 SVG and D3, Grammar of Graphics
Week 3, Feb 8 D3: circles, text, scales; naive Bayes classifiers
   [Winter break]
Week 4, Feb 17 D3: lines; Linear regression; Color theory
Week 5, Feb 22 D3: transitions; k-means
Week 6, Feb 29 Networks: force-directed layouts
Week 7, Mar 7 [Project 1 due Monday 3/7] Network sampling
Week 8, Mar 14 D3: paths, mouseovers; maps and geographic data
Week 9, Mar 21 More paths: Time series smoothing
   [Spring break]
Week 10, Apr 4 Perceptrons, SVMs
Week 11, Apr 11 Locality-sensitive Hashing
Week 12, Apr 18
Week 13, Apr 25 [Project 2 due Monday 4/20]
Week 14, May 2 Randomization, permutation testing, bootstrapping
Week 15, May 9 (MW) Summary and review

Take home final. Dates to be announced.