Day 1: Introduction to Working with Open Data

Raymond Yee

January 21, 2014 (`http://is.gd/wwod1401`)

Course Overview

INFO 290T- Working with Open Data
http://www.ischool.berkeley.edu/courses/290t-wod Spring 2014 /
CCN: 41620
T,Th 2:00-3:30pm 202 South Hall
Office Hours: T, Th 3:30-4:30pm, 303A South Hall
Instructor: Raymond Yee, Ph.D.
Contact info: `yee@berkeley.edu` (@WorkingOpenData / @rdhyee)
Tutor: AJ Renold (`arenold@ISchool.Berkeley.EDU`)

Goals Today

- Introduce prospective students as to the purpose, structure, content of the course
- Begin to think together about open data and using Python to analyze open data
- Start building a learning community that will work together
- Build basic communication structures

Course Description

Open data – data that is free for use, reuse, and redistribution – is an intellectual treasure-trove that has given rise to many unexpected and often fruitful applications. In this course, students will

1. learn how to access, visualize, clean, interpret, and share data, especially open data, using Python, Python-based libraries, and supplementary computational frameworks
2. understand the theoretical underpinnings of open data and their connections to implementations in the physical and life sciences, government, social sciences, and journalism.

Working with Open Data (WwOD) is a *technical* course with a strong focus on the social-political context and domains of application of open data.

Prerequisite

Info 90 (Programming for Computing Applications) or equivalent background with Python.

Expectations

- Participants will work on tangible projects related to the overall theme. They can select from a list of projects I design or they can propose other projects of comparable scope and intent. However, we will find ways for the projects to combine together into a larger super-project.
- Participants will be heavily involved in learning from and teaching each other, depending on each other for the course's collective success. The class will provide support for students to work together, not only in their own project group, but also course-wide and even with people working with us outside of the class.
- To ensure that our projects remain grounded in the "real world," we'll be working to engage outside users for our projects from the outset.
- The course will be designed to enable the larger community to participate.

Main Textbook

Wes McKinney. `Python for Data Analysis`. (O'Reilly Media, 2012). I strongly recommend getting a paper copy as well as accessing any electronic versions

- oreilly.com
- Proquest.safaribooksonline at UCB

Supplementary Materials

I plan to supplement the book with materials covering the following topics:

- open data, open content in various fields
- using JavaScript, HTML5, CSS together with Python for data presentation, analysis, and visualization, (e.g., d3.js)

In addition to survey materials on the public domain, creative commons, and open data movements, I'll focus us on

- Wikipedia, dbpedia, Freebase data
- census data

and other data sets still to be determined, probably large open scientific data sets