

# Day 2: Review of Python & iPython (Chap 3 and Chap 13)

Raymond Yee

January 24, 2013 (<http://bit.ly/wwod1302>)

# Goals Today

- ▶ make sure we have an iPython environment set up for everyone
- ▶ Work through exercise in census iPython notebook:  
`https://github.com/rdhyee/working-open-data/blob/master/notebooks/Day\_02\_class\_starter.ipynb`

# Assigned Readings for today

- ▶ read PfDA, Chap 1 Preliminaries, especially the installation instructions for EPD Free for your computer platform. I want you to try installing EPD Free (or EPD Academic) before class on Thursday.
- ▶ read PfDA, Chap 3
- ▶ skim PfDA, Appendix: Python Language Essentials – to help remind yourself of key elements of standard Python
- ▶ skim PfDA, Chap 2 Introductory Examples

## Examples of Open Data

- ▶ <http://www.data.gov/>
- ▶ <https://data.sfgov.org/>
- ▶ <https://data.acgov.org/>
- ▶ <http://data.openoakland.org/>
- ▶ <http://www.socrata.com/discover/video-case-study-somerville-ma/>
- ▶ <http://sunlightfoundation.com/projects/>
- ▶ [http://oad.simmons.edu/oadwiki/Data\\_repositories#Astronomy](http://oad.simmons.edu/oadwiki/Data_repositories#Astronomy)
- ▶ <http://opencontext.org/>
- ▶ <http://courtlister.com/>
- ▶ <http://aws.amazon.com/publicdatasets/>
- ▶ <http://openmetadata.lib.harvard.edu/bibdata>
- ▶ <http://www.bart.gov/schedules/developers/api.aspx>  
/ <http://www.bart.gov/schedules/developers/index.aspx>
- ▶ <http://www.ncdc.noaa.gov/>
- ▶ <http://www.propublica.org/tools/>

# Motivation: why I care about open data and why you might care

Traditional motivations given for open government data:

- ▶ transparency
- ▶ accountability
- ▶ efficiency
- ▶ innovation

My personal interests in the area:

- ▶ Open data useful testbed for working on data of all sorts, because of zero financial costs and minimal restrictions on use, reuse, redistribution
- ▶ Growing community around open data because of these low barriers. . . democratization of data. . . many more of us can participate in working with open data and attract a wide range of people I love to learn and to think and to understand, a big believer of computational and information systems as mind augmenters/extenders and open data (as

## Quick Examples

### money and politics

I read a story in Sunday's New York Times in connections made between political fund-raising and specific decisions being made in Congress. So how can we participate in this process we can certainly read stories which have made these connections. We first verify that yes, The New York Times is correct in making these connections. We can use this process to discover new connections – wasn't obvious from reading article on paper but online article has links to [opensecrets.org](http://www.opensecrets.org) – e.g.,

<http://www.opensecrets.org/politicians/contrib.php?cycle=2012&cid=N00004643&type=C>

### desire for transparency in China

Fascinating to read NYT saying that the Chinese middle class demanding government transparency:

<http://www.nytimes.com/2013/01/20/world/asia/in-china-discontent-among-the-normally-faithful.html>

# Random and not-so random questions for me that open data can help answer

Reading the news, world news, local news, tech news, understanding new contexts, deepening old interests, controlled serendipity.

- ▶ What music to listen to
- ▶ What book to read?
- ▶ When can I see the next episode of White Collar?
- ▶ When was BWV 156 (Ich steh mit einem Fuß im Grabe) first performed?
- ▶ How to invest our money?
- ▶ What programming language to learn next?
- ▶ What charity to give money to?
- ▶ What restaurant to try?
- ▶ What to cook?
- ▶ How should we take care of our physical health?

# Some Big Questions for the Course

- ▶ What are the essential characteristics of open data?
- ▶ What are the costs and benefits of open data?
- ▶ How to map the universe of open data? What's out there?  
What data is not available in open form?
- ▶ What can we learn from open data?
- ▶ What business models?
- ▶ What are people doing with open data?
- ▶ What are the different common formats used to represent open data? (e.g., CSV, XML, KML, SHP in data.gov) – and how can we use Python to process those formats?
- ▶ What are the issues that we face in combining open data with closed data



# Homework

- ▶ Submit to `yee@berkeley.edu` by Wednesday, Jan 30 at noon. the iPython notebook.