Course Title: Principles of Urban Informatics  
Course number: CUSP-GX-5003  
Time: Thursday, 9.00 AM - 11.30 AM  
Dates: 09/03/2013 - 12/13/2013  
Room: Magnet, 2MTC, 8FL, Room 820  
Credits: 3  
Session: 001

Teaching Team

Instructor: Claudio Silva  
TAs: Nivan Ferreira and Fabio Miranda

(Discipline Core) (3 credits) – Introduction to the core disciplines of data acquisition and management, integration, and analytics; presents software tools, frameworks for problem-solving and model selection parameters using data science in the urban context, including basic modeling and analytical methods; visualization techniques, including geographic information systems; working with large datasets and understanding data sources, including instrumentation, physical sensors, imagery, and mobile sensing platforms; explores issues of data ethics, privacy, etc.; and provides an introduction to citizen science, crowdsourcing, and participatory sensing.

Learning objectives

In this course, the student will learn the major concepts, tools, and techniques for what informatics can do for cities. It includes background in data management, visualization, and data science, and also includes material not usually covered in computer science courses, including how to best handle spatial-temporal data, and issues related to citizen science and participatory sensing. The course is essentially divided into four major parts, each of which will have a written and programming assignment. The course will not have a final, instead, we expect each student to apply the techniques learned in the class in a mini-project, which is likely to involve working on making an open city dataset more useful (e.g., through curation or by creating a visualization tool).

Major references:

The Practice of Programming  
Brian W. Kernighan, Rob Pike  
Addison-Wesley, 1999
SQL in a Nutshell
Kevin Kline, Brand Hunt, Daniel Kline
O'Reilly, 2008

Visualization Analysis and Design: Abstractions, Principles, and Methods
Tamara Munzner

Data Science for Business: What you need to know about data mining and data-analytic thinking
Foster Provost, Tom Fawcett
O'Reilly, 2013

Tentative Schedule

Week 1 (9/5/2013): Introduction to the course; urban data types; public data sources; examples of existing systems.

Programming & Tools Module

Week 2 (9/12/2013): Programming, scripting, and software engineering concepts and tools. We will give an introduction to Python and R. We will cover some good software practices, the user of version control systems. We will also demonstrate the use of important tools for building software, including emacs.

Week 3 (9/19/2013): This is a continuation of the previous class, mostly focused on extending our coverage of Python, and potentially also showing some initial Javascript code.

Database Module


Week 6 (10/10/2013): Database concepts, tools, and techniques, Part III. Introduction to Big Data management. Introduction to Big data management. MapReduce: Introduction, programming model / paradigm, HADOOP, examples. MongoDB + MapReduce, handling big data with a combination of MongoDB and MapReduce, examples.

Week 7 (10/17/2013): Database concepts, tools, and techniques, Part IV. Spatial-temporal datasets. Spatial datasets: introduction, types, queries of interest. Spatial Indexes: R-tree, Quad tree, KD-tree. GIS – Postgis, spatial data support for PostgreSQL, examples. Spatial temporal datasets examples, types of queries: queries wrt current time, queries over time intervals.

Visualization Module


Week 9 (10/31/2013): Visualization concepts, tools, and techniques, Part II, 2D plotting techniques. Go through different 2D plot types, highlighting the channels and marks (bar chart, scatterplot, treemap, etc.). Interaction. Data filtering and aggregation. Focus+Context. Crossfilter introduction. Create dimensions using it, apply filter and aggregation. Example of Crossfilter (filter and aggregation).


Data Science Module


Extras

Week 14 (12/5/2013): Invited industrial lecture or make up lecture. (The data might float earlier as needed.)