

# CUSP-GX-6001: Urban Computing Skills Lab (UCSL) Summer 2014

## Instructors

**Module 1:** J.C. Bonilla - [jb3379@nyu.edu](mailto:jb3379@nyu.edu)

**Module 2:** Hong Yang - [hong.yang@nyu.edu](mailto:hong.yang@nyu.edu)

**Module 3:** Huy Vo - [huy.vo@nyu.edu](mailto:huy.vo@nyu.edu)

## Overview

The UCSL at CUSP is a series of online sessions designed to build a common skillset and familiarity with techniques, concepts, and models for urban informatics computing. The online sessions focus on data exploration, programming skills and statistical methods needed for scientific computing in the field of urban informatics.

## Accessing Lectures

Lectures can be accessed via NYUHome>>Academics>>lynda.com.

<https://home.nyu.edu/academics>

## Schedule & Assignments

Lectures are designed as introductions to the theory and techniques in each module and as foundations for completing all assignments. Students who are experienced in the computing platforms in each module (Excel for module 1, R for module 2 and Python for module 3) are can initiate assignments without completing each video lecture.

Modules and assignments will follow the schedule below:

Weeks	Module	Assignment Due Date
June 30 - July 6	Data Exploration in Excel	Assignment #1: July 6
July 7 - 27	Statistical Methods in R	Assignment #2: July 13 Assignment #3: July 27
July 28 - 17	Introduction to Python	Assignment #4: August 3 Assignment #5: August 17

## Module 1: Data Exploration in Excel (July 1 – 6)

---

**Instructor:** JeanCarlo (J.C) Bonilla - [jb3379@nyu.edu](mailto:jb3379@nyu.edu)

### Lecture 1.1: Managing and Analyzing Data

<http://www.lynda.com/Excel-tutorials/Excel-2013-Managing-Analyzing-Data/158657-2.html>

<http://www.lynda.com/Excel-tutorials/Using-Top-10-value-percent-option/158657/168481-4.html>

When you work with large amounts of data, you need tools to manage the sheer volume of information and get the most value from it. This video lecture shows a variety of ways to use sorting to order your spreadsheets, insert subtotals into sorted lists, see just data you need with filtering, identify and delete duplicate data, and use the SUMIF and database functions to further analyze your data.

Topics include:

- Multiple key sorting
- Filtering single and multiple columns
- Creating a top-ten list with values or percentages
- Setting up subtotals
- Creating multiple-field criteria filters
- Creating unique lists from repeating field data
- Using the Remove Duplicates command
- Finding duplicate data with specialized arrays
- Counting the number of unique items in a list
- Using SUMIF and COUNTIF functions
- Working with the database functions such as DSUM and DMAX
- Converting lists to tables

### Lecture 1.2: Up and Running with Excel What-If Analysis

<http://www.lynda.com/Excel-tutorials/Up-Running-Excel-What-Analysis/150165-2.html>

This video lecture shows how to analyze hypothetical business cases using formulas and variable data. Specifically, you will learn how to build data tables, define alternative data sets with scenarios, find target values with Goal Seek, and analyze complex problems with Solver. The last two chapters show you how to create a configurable model with sliders and check boxes and how to run a Monte Carlo simulation.

Topics include:

- Creating a scenario
- Defining data tables
- Finding a target value with Goal Seek
- Defining constraints and running Solver
- Creating a configurable model
- Manipulating your model
- Running a Monte Carlo simulation

**Assignment #1:** MonteCarlo Simulation on Service Delivery

**Data:** NYC Open Data

**Due Date:** July 6th, 2014 at 11:59pm. Submission via *NYUClasses*

## Module 2: Statistics Methods in R ( July 7 – 27)

---

**Instructor:** Hong Yang - [hong.yang@nyu.edu](mailto:hong.yang@nyu.edu)

### Lecture 2.1: Introduction to R

<http://www.lynda.com/R-tutorials/R-Statistics-Essential-Training/142447-2.html>

<http://www.lynda.com/R-tutorials/R-Statistics-Essential-Training/142447-2.html>

R is the language of big data—a statistical programming language that helps describe, mine, and test relationships between large amounts of data. In this lecture you will learn how to use R to model statistical relationships using graphs, calculations, tests, and other analysis tools. Students will learn how to enter and modify data; create charts, scatter plots, and histograms; examine outliers; calculate correlations; and compute regressions, bivariate associations, and statistics for three or more variables.

Topics include:

- Installing R on your computer
- Using the built-in datasets
- Importing data
- Creating bar and pie charts for categorical variables
- Creating histograms and box plots for quantitative variables
- Calculating frequencies and descriptives
- Transforming variables
- Coding missing data
- Analyzing by subgroups
- Creating charts for associations
- Calculating correlations
- Creating charts and statistics for three or more variables
- Creating crosstabs for categorical variables

### Assignment #2: R Basics

**Data:** R\_Assignment 1\_Data\_CherryTrees.csv

**Due Date:** July 13th, 2014 at 11:59pm. Submission via *NYUClasses*

### Assignment #3: Examine the Parking Violation Data Set

**Data:** TBD

**Due Date:** July 27th, 2014 at 11:59pm. Submission via *NYUClasses*

## Module 3: Introduction to Python (July 28 – August 17)

---

**Instructor:** Huy Vo - [huy.vo@nyu.edu](mailto:huy.vo@nyu.edu)

### Lecture 3.1: Up and Running with Python

<http://www.lynda.com/Python-tutorials/Up-Running-Python/122467-2.html>

Get a quick intro to Python, the popular and highly readable object-oriented language. This video lecture provides an overview of the installation process, basic Python syntax, and an example of how to construct and run a simple Python program. Learn to work with dates and times, read and write files, and retrieve and parse HTML, JSON, and XML data from the web.

Topics include:

- Installing Python
- Choosing an editor/IDE
- Working with variables and expressions
- Writing loops
- Using the date, time, and datetime classes
- Reading and writing files
- Fetching Internet data
- Parsing and processing HTML

### Lecture 3.2: Handling of Spatio-Temporal data

The growth of social networks together with the availability of low cost sensors has resulted in a rapid increase of spatial-temporal data. These data, especially pertaining to urban environments, are being actively collected through various means ranging from conventional sensors, such as power consumption and noise, to more “unconventional” means such as GPS in vehicles, mobile devices, and social media.

Moreover, cities all over the world are not only collecting these data, but they are also making the data available (see e.g., NYC Open Data). In brief, almost all urban datasets have temporal (time) or spatial (geo location) components or both. Thus, it is crucial for us to handle spatio-temporal data appropriately if we want to work with urban datasets. In this lecture, we will be introduced to shapefiles as well as how to visualize them with matplotlib.

In this lecture, we will be introduced to shapefiles as well as how to visualize them with matplotlib.

<http://www.lynda.com/ArcGIS-tutorials/Understanding-shapefile-data/162135/172475-4.html> (the first 3 chapter)

<http://scipy-lectures.github.io/intro/matplotlib/matplotlib.html>

[http://introtopython.org/visualization\\_earthquakes.html](http://introtopython.org/visualization_earthquakes.html)

- Understanding geo-spatial data
- Plotting made easy with matplotlib
  - How to do simple plots (similar to R)
  - How to plot geo-spatial data (shapefiles)

**Assignment #4:** Tracking real-time bus information (using Python to fetch real-time bus information directly from MTA through their GPFS interface)

**Data:** MTA Bus-time real-time feeds

**Due Date:** August 3rd, 2014 at 11:59pm. Submission via *NYUClasses*

**Assignment #5:** Visualizing real-time bus information on a map (built on top of the assignment #4)

**Data:** MTA Bus-time data

**Due Date:** August 17th, 2014 at 11:59pm. Submission via *NYUClasses*