Urban Big Data Analytics - Syllabus

Andy Hong

July 18, 2018

Course information

- Instructor: Andy Hong (andyhong@gmail.com)
- Course Credit(s): 3.00
- Term: July 14 to August 14, 2018
- Day/Time: 9:00 – 12:00 M-F
- Location: Room 110, West Mall Annex

Course Description:

With the advent of open data movement, knowledge and skills for collecting and analyzing big data become increasingly important for urban planners. This course will teach students how to harness the power of big data by mastering the way they are collected, organized, and analyzed to support better decision making in urban planning context. Students will learn the basic tools needed to manipulate large datasets derived from various open-data platforms, from data collection to storage and approaches to analysis. Students will be able to capture and build data structures, perform basic queries in order to extract key metrics and insights. In addition, students will learn how to use various data analytic tools, such as Tableau and Exploratory, to analyze and visualize data. The course will also give students some exposure to statistical programming with R, and introduce them to basic machine learning techniques.

Prerequisites:

- Experience with coding is a plus but not required

Course requirement:

- This course requires students to have access to a computer. Please bring your laptop or talk to your VSP coordinator about having access to computers in the library.

Course references:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. New York, NY, USA: Springer. (Can be downloaded for free: https://www-bcf.usc.edu/~gareth/ISL/
- Introduction to R, Free online course offered by DataCamp https://www.datacamp.com/courses/free-introduction-to-r
- Intro to Python for Data Science, Free online course offered by DataCamp https://www.datacamp.com/courses/intro-to-python-for-data-science
- (Additional reading) EMC Education Services. (2015). Data Science & Big Data Analytics. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Indianapolis, IN, USA: John Wiley & Sons, Inc.

Grading Scheme:

Your final grade for the course will be based on the following three items

- Course participation (3% for each of the 10 classes): 30%
- Four assignments (10% for each assignment): 40%
- Final group project (on-line submission): 30%

Course schedule

PART I. DATA SCIENCE BASIC

Class 1 - Introduction to urban big data - `Jul 18 (Wed) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Course overview and syllabus
2. Introduction of data science and big data
3. Emergence of urban data science and open data initiatives
4. Linking data, city management, and policy making
5. Overview of the assignments and group project

ASSIGNMENT #1 OUT [web] [pdf]

Class 2 - Data acquisition through open-data platform - `Jul 19 (Thu) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Introduction to Exploratory
2. Vancouver open data catalogue
3. NYC open data
4. Chicago open data – OpenGrid
5. Group project topic assignment

ASSSIGNMENT #1 DUE / ASSIGNMENT #2 OUT [web] [pdf]

Class 3 - Data wrangling - `Jul 20 (Fri) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Data types, conversion, and categorization
2. Data filtering, sorting, and reordering
3. Summarizing and joining data
4. Hands-on group session

PART II. GEOSPATIAL DATA BASIC

Class 4 - Database and SQL - `Jul 23 (Mon) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Introduction to database system
2. Difference between tabular and text-based data
3. Choosing which tools to use for which purpose
4. Loading database using SQL
5. Importing data to SQLite
6. Querying and filtering data

ASSSIGNMENT #2 DUE / ASSIGNMENT #3 OUT [web] [pdf]

Class 5 - Processing spatial data and GeoJSON data - `Jul 24 (Tue) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Basic cartography and projection
2. Spatial data basic: point, line, polygon, and raster
3. Reading conventional spatial data: shapefiles and geo-database
4. Reading GeoJSON data
5. Merging tabular data with spatial data
6. Hands-on group session

Class 6 - Cloud computing and Google Big Query - `Jul 25 (Wed) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. What is big data and why do we need to know?
2. Basic intro to cloud database system
3. Difference between Google BigQuery, Amazon Web Services, and Microsoft Azure
4. Basic data querying steps
5. SQL basic and examples
6. Group project mid term check-in

Class 7 - Exploratory data analysis (EDA) - `Jul 26 (Thu) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Intro to exploratory data analysis
2. Intro to R and basic programming skills
3. Summary statistics and data types
4. Basic plotting and correlations
5. Missing data and handling outlier

ASSSIGNMENT #3 DUE / ASSIGNMENT #4 OUT [web] [pdf]

PART III. ADVANCED MODELING AND VISUALIZATION**

Class 8 - Data visualization and web mapping - `Jul 30 (Mon) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Intro to leaflet.js
2. Choosing the right form of data visualization
3. Coloring schemes, plotting, and stratification
4. Spatial data handling and mapping
5. Hands-on group session

Class 9 - Statistical learning with Exploratory - `Jul 31 (Tue) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Basic probability and statistics – signal-to-noise ratio
2. Statistical inference and modeling
3. Linear regression with continuous data
4. Data transformation
5. Hands-on group session

Class 10 - Advanced statistical modeling - `Aug 1 (Wed) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Regression with skewed data
2. Poisson and negative binomial regression
3. Logistic regression and interpretation
4. Model selection and goodness-of-fit
5. Hands-on group session

ASSSIGNMENT #4 DUE

Class 11 - Basic machine learning and future of urban data science - `Aug 2 (Thu) 9:00-12:00`
Lecture slides: [web] [pdf]
Group session: [web] [pdf]

1. Intro to machine learning
2. Data clustering with k-nearest neighbours
3. Random Forrest (short example)
4. Support Vector Machine (short example)
5. Application to urban data science
6. Group project presentation prep

Class 12 - Final presentation and speciall session

Session A: Final group project presentation - `Aug 2 (Thu) 1:30-4:00`
- Each group will have 15 minutute to present their project.

Session B: Special event will follow after the presentation - `Aug 2 (Thu) 4:00-5:00`
1. Data Science for Social Good (DSSG) Fellow will come and discuss their work applying data science skills to solving various social issues 2. Special mixer event will take place after the group discussion.

FINAL PROJECT DUE BY Aug 8, 12:00 MIDNIGHT
Submit your final group project to the instructor (andyhong@gmail.com) by Aug 8 (Wed)
(-3% for each day of late submission)

Please use the following email title format:
`VSP BigData - [group number] - [project name]`
ex), `VSP BigData - Group 1 -`