Urban Big Data Analysis

Instructor Information:
Instructor: Andy Hong
Email: andyhong@gmail.com
Phone: (604) 822 - 3687
Office room: #372 SPPH, University of British Columbia

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

Textbooks:

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. New York, NY, USA: Springer.
- 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
- 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 four items:

- Course participation 20%
- Five assignments (6% each) 30%
- Mid-term 30%
- Group project and presentation 20%

Course Schedule:

1. Introduction to urban big data analysis
    a. Course overview and syllabus
    b. Introduction of data science and big data
    c. Emergence of urban data science and open data initiatives
    d. Linking data, city management, and policy making
    e. Overview of weekly assignments and group project (Assignment #1)

2. Data acquisition through open-data platform
    a. Introduce Exploratory and Tableau
    b. Vancouver open data catalogue
    c. NYC open data
    d. Chicago open data – OpenGrid
    e. Hands-on lab session

3. Working with JSON and MongoDB data
    a. Introduction to database system
    b. Difference between tabular and text-based data
    c. Choosing which tools to use for which purpose
    d. Viewing and processing JSON
    e. Importing data to MondoDB
    f. Querying and filtering data (Assignment #2)

4. Data wrangling
    a. Data types, conversion, and categorization
    b. Data filtering, sorting, and merging
    c. Transforming data between wide and long format
    d. Hands-on lab session

5. Processing spatial data and GeoJSON data
    a. Basic cartography and projection
    b. Spatial data basic: point, line, polygon, and raster
    c. Reading conventional spatial data: shapefiles and geo-database

   d. Reading GeoJSON data
   e. Merging tabular data with spatial data
   f. Hands-on lab session – Census data (Assignment #3)

6. Cloud computing and Google Big Query
   a. What is big data and why do we need to know?
   b. Basic intro to cloud database system
   c. Difference between Google BigQuery, Amazon Web Services, and Microsoft Azure
   d. Basic data querying steps
   e. SQL basic and examples
   f. Hands-on lab session with Google BigQuery and NYC Taxi trip data

7. **Mid-Term**
   a. Multiple choice test on the fundamentals of data science
   b. Multiple choice test on data types, conversion, and merging
   c. Short answers to JSON and MongoDB processing
   d. Short answers to GeoJSON and spatial data processing
   e. Describe basic steps of data cleaning and wrangling
   f. Multiple choice test on big data and cloud computing

8. Exploratory data analysis (EDA) with Exploratory
   a. Intro to Exploratory – Intuitive and user-friendly R coding
   b. Intro to R and basic programming skills
   c. Summary statistics and data types
   d. Basic plotting and correlations
   e. Missing data and handling outlier
   f. Hands-on lab session (Assignment #4)

9. Data visualization and web mapping
   a. Intro to Tableau and CartoDB
   b. Choosing the right form of data visualization
   c. Coloring schemes, plotting, and stratification
   d. Spatial data handling and mapping
   e. Hands-on lab session

10. Statistical learning with Exploratory
   a. Basic probability and statistics – signal-to-noise ratio
   b. Statistical inference and modeling
   c. Linear regression with continuous data

d. Data transformation
e. Hands-on lab session (Assignment #5)

11. Advanced statistical modeling
    a. Regression with skewed data
    b. Poisson and negative binomial regression
    c. Logistic regression and interpretation
    d. Model selection and goodness-of-fit
    e. Hands-on lab session

12. Basic machine learning and future of urban data science
    a. Intro to machine learning
    b. Data clustering with k-nearest neighbours
    c. Random Forrest (short example with hands-on exercise)
    d. Support Vector Machine (short example with hands-on exercise)
    e. Application to urban data science and decision making

13. Final project presentation
    a. A group of 3-4 people
    b. Assignment of group project based on students' interests
    c. Project deliverables: document + codes
    d. Group presentation: power point
    e. Peer evaluation