

Urban Big Data Analytics - Syllabus

Andy Hong

July 17, 2019

Course Information

- Instructor: Andy Hong, The George Institute for Global Health, University of Oxford
- Contact: andyhong@gmail.com, <http://andyhong.org>
- Course Credit(s): 3.00
- Term: July 15 to August 8, 2019
- Day/Time: 9:00 – 12:00 M-Th
- 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 Gapminder 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.

Course requirement

- Experience with coding is a plus but not required
- This course requires students to have a computer. Please bring your laptop.
- [ESSENTIAL] Exploratory: <https://exploratory.io/plan?plan=Community>
 - Use your university email to get the free version. If you don't have a university email, ask your university to create one for you.
- R: <https://mirror.its.sfu.ca/mirror/CRAN>
- R-Studio: <https://www.rstudio.com/products/rstudio/download>
- QGIS: <https://qgis.org/en/site>

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%
- Pop quiz (one random in-class quiz, closed book): extra 3% point

Course Schedule

I. Data Science Basic

Lecture 1 - Introduction to urban big data - Jul 17 (Wed) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session: [\[link\]](#)

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 [\[link\]](#)

Lecture 2 - Data acquisition through open-data platform - Jul 18 (Thu) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session: [\[link\]](#)

Group project part 1: [\[link\]](#)

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

ASSIGNMENT #1 DUE / ASSIGNMENT #2 OUT [\[link\]](#)

Lecture 3 - Data wrangling + Special guest speaker - Jul 19 (Fri) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session: [\[link\]](#)

1. Data wrangle principles: NICE(R)
2. Data types, conversion, and categorization
3. Data filtering, sorting, and reordering
4. Summarizing and joining data
5. Hands-on group session
6. Guest speaker: Akansha Vashisth, Masters in Data Science, UBC (11:30 - 12:00)

II. Database & Computing

Lecture 4 - Database and SQL - Jul 22 (Mon) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session: [\[link\]](#)

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

ASSIGNMENT #2 DUE / ASSIGNMENT #3 OUT [\[link\]](#)

Lecture 5 - Spatial data and GeoJSON data - Jul 23 (Tue) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session: [\[link\]](#)

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

Lecture 6 - Cloud computing and Google Big Query - Jul 24 (Wed) 9:00-12:00

Lecture slides: [\[link\]](#)

Group project part 2: [\[link\]](#)

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

Lecture 7 - Exploratory data analysis (EDA) Jul 25 (Thu) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session and group project part 3: [\[link\]](#)

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

ASSIGNMENT #3 DUE

III. Advanced Analytics

Lecture 8 - Data visualization and web mapping - Jul 29 (Mon) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session: [\[link\]](#)

1. Intro to data visualization
2. Tufte's 10 rules
3. Data visualization with ggplot2
4. Interactive web mapping with leaflet
5. Hands-on group session

ASSIGNMENT #4 OUT [\[link\]](#)

Lecture 9 - Statistical analysis with Exploratory - Jul 30 (Tue) 9:00-12:00

Lecture slides: [\[link\]](#)

Group project part 4: [\[link\]](#)

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

Lecture 10 - Advanced statistical analysis - Jul 31 (Wed) 9:00-12:00

Lecture slides: [\[link\]](#)

Group session: [\[link\]](#)

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

ASSIGNMENT #4 DUE

Lecture 11 - Basic machine learning and future of urban data science - Aug 1 (Thu) 9:00-12:00

Lecture slides: [\[link\]](#)

1. Intro to machine learning
2. Supervised vs. Unsupervised learning
3. Data clustering with k-means technique
4. Future of urban data science
5. Group project presentation prep

Lecture 12 - Final group presentation + Special guest speaker - Aug 6 (Tue) 9:00-12:00

- Each group will have 15 minutes to present their project.
- Guest speaker: Sayanti Ghosh, Masters in Data Science, UBC (11:30 - 12:00)

Final Project

Final project is DUE BY Aug 8 (Thu), 11:59 MIDNIGHT

Submit your final group project to the instructor (andyhong@gmail.com) by Aug 8 (Thu)
(-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 - Boston Crime Analytics