

# PH244

## Big Data: A Public Health Perspective

### General Information

- **Instructor:** Lexin Li  
Email: lexinli@berkeley.edu  
Campus address: 344B Li Ka Shing Center, MC 3370  
Office Phone: 510-664-4584
- **Lectures:** MW 10am-12pm, Rm 440 / 714C, University Hall

### Course Description

- **Prerequisite:** PB HLTH 142 (basic concepts of probability and distributions, point and interval estimation, hypothesis testing), 145 (regression analysis of continuous outcome), 241 (categorical data analysis, some modern statistical learning techniques), or equivalent. If you have any concerns about prerequisite, please come talk to me right after the first class.
- **Recommended books:** There is no required textbook, but the following books are good reading for this course.
  - *The Elements of Statistical Learning – Data Mining, Inference and Prediction*, Trevor Hastie, Robert Tibshirani, and Jerome Friedman, 2009, 2nd edition, Springer.
  - *Mining of Massive Datasets*, Anand Rajaraman, Jure Leskovec, and Jeffrey D. Ullman, 2014, Stanford University
- **Learning objectives:** by the end of semester, students will be expected
  - to have an in-depth understanding of key statistical principles as well as a wide range of data mining and machine learning techniques;
  - to gain some experience with modern optimization and computing paradigms;
  - to have a good understanding of some key Big Data applications in public health, including electronic health record, omics, mobile health, among others.

### Projects and Grading

- There are totally 3 data analysis projects + 1 computing project.

- You can do the project **on your own**, or **pair up** with a classmate. Each team should be no more than 2 persons.
- Big Data applications in public health:
  - Project I: Electronic Health Record
    - Heritage Health Prize:  
<https://www.kaggle.com/c/hhp>
    - Predict HIV Progression:  
<https://www.kaggle.com/c/hivprogression>
  - Project II: Mobile Health
    - Predicting Parkinson’s Disease Progression with Smartphone Data:  
<https://www.kaggle.com/c/predicting-parkinson-s-disease-progression-with-smartphone-data/data>
    - Accelerometer Biometric Competition:  
<https://www.kaggle.com/c/accelerometer-biometric-competition/data>
  - Project III: Computing
  - Project IV: Omics/Imaging
    - Second Annual Data Science Bowl: Transforming How We Diagnose Heart Disease:  
<https://www.kaggle.com/c/second-annual-data-science-bowl>
    - Data Science Bowl 2017: Can You Improve Lung Cancer Detection?  
<https://www.kaggle.com/c/data-science-bowl-2017/>
- Data repository: [www.kaggle.com](http://www.kaggle.com)
- Project **due dates**:

	Project	Topic	Due date
(1)	Project I	EHR	February 12
(2)	Project II	Mobile Health	March 12
(3)	Project III	Computing	April 9
(4)	Project IV	Omics/Imaging	May 7

- Data analysis project:
  - Under each project category, **choose one data** to work on — please coordinate with me.
  - Ask questions: (a) the competition question; (b) a question of your own.

- Data processing: handle and clean the real world big and messy data.
- Data analysis: (a) summarize the winning methods; (b) propose your own solution and do a comparison; (c) carry out an analysis to address your own question.     **only if winning methods are available**
- Written report:
  - The report is **no more than 3 pages, including** tables, figures, references, and everything.
  - At the beginning of the report, prepare a summary of your findings (what kind of useful knowledge you have obtained via your analysis).
  - Make sure to include a description on how you process the data.
- Oral presentation:
  - There is an oral presentation **in class** in the week of April 23 to April 25.
  - You can choose to present any one of the 3 data analysis projects you have done — please coordinate with me.
- There is no other homework or exam.
- Grading: four projects 20% each + oral presentation 20%.

## University Policies

- **ACCOMMODATION OF RELIGIOUS CREED** In compliance with California Education Code, Section 92640(a), it is the official policy of the University of California at Berkeley to permit any student to undergo a test or examination, without penalty, at a time when that activity would not violate the student's religious creed, unless administering the examination at an alternative time would impose an undue hardship that could not reasonably have been avoided. Requests to accommodate a student's religious creed by scheduling tests or examinations at alternative times should be submitted directly to the faculty member responsible for administering the examination by the second week of the semester.

Reasonable common sense, judgment and the pursuit of mutual goodwill should result in the positive resolution of scheduling conflicts. The regular campus appeals process applies if a mutually satisfactory arrangement cannot be achieved. The link to this policy is available in the Religious Creed section of the Academic Calendar webpage.

- **CONFLICTS BETWEEN EXTRACURRICULAR ACTIVITIES AND ACADEMIC REQUIREMENTS** The Academic Senate has established Guidelines Concerning Scheduling Conflicts with Academic Requirements to address the issue of conflicts that arise between extracurricular activities and academic requirements. These policies specifically concern the schedules of student athletes, student musicians, those with out-of-town interviews, and other students with activities (e.g., classes missed as the result of religious holy days) that compete with academic obligations.

These policies were updated in Spring 2014 to include the following statement:

+The pedagogical needs of the class are the key criteria when deciding whether a proposed accommodation is appropriate. Faculty must clearly articulate the specific pedagogical reasons that prevent accepting a proposed accommodation. Absent such a reason, the presumption should be that accommodations are to be made.

The guidelines assign responsibilities as follows:

+It is the instructor's responsibility to give students a schedule, available on the syllabus in the first week of instruction, of all class sessions, exams, tests, project deadlines, field trips, and any other required class activities.

+It is the student's responsibility to notify the instructor(s) in writing by the second week of the semester of any potential conflict(s) and to recommend a solution, with the understanding that an earlier deadline or date of examination may be the most practicable solution.

+It is the student's responsibility to inform him/herself about material missed because of an absence, whether or not he/she has been formally excused.

The complete guidelines are available on the Academic Senate website. Additionally, a checklist to help instructors and students comply with the guidelines is available on the Center for Teaching and Learning website.

- **ABSENCES DUE TO ILLNESS** Instructors are asked to refrain from general requirements for written excuses from medical personnel for absence due to illness. Many healthy people experience a mild-to-moderate illness and recover without the need to seek medical attention. University Health Services does not have the capacity to evaluate such illnesses and provide documentation excusing student absences. However, UHS will continue to provide documentation when a student is being treated by Tang for an illness that necessitates a change in course load or an incomplete.

From time-to-time the Academic Senate has issued guidance concerning missed classes and exams due to illnesses such as influenza advising that students not attend class if they have a fever. Should a student experience repeated absences due to illness, it may be appropriate for the faculty member to ask the

student to seek medical advice. The Senate guidelines advise faculty to use flexibility and good judgment in determining whether to excuse missed work, extend deadlines, or substitute an alternative assignment. Only the Committee on Courses of Instruction (COCI) can waive the final exam. However, a department chair can authorize an instructor to offer an alternative format for a final exam (e.g., paper, take-home exam) on a one-time basis (<http://academic-senate.berkeley.edu/committees/coci/toolbox#16>).

- **ACCOMMODATION FOR DISABILITY** Instructors are reminded of their responsibilities for accommodating disabilities in the classroom in the following areas:

**Confidentiality:** Information about a student's disability is confidential, and may not be shared with other students. **Role of Instructor:** Course instructors play a critical role in enabling the University to meet its obligation to appropriately accommodate students with disabilities who are registered with the Disabled Students Program (DSP) and who have been issued a Letter of Accommodation. **Reading Assignments:**

+In advance: Because students with print disabilities usually need assistance from the DSP Alternative Media Center, reading materials should be provided well in advance (two or more weeks) before the reading assignment due date.

+Required or Recommended: Always indicate which course readings (including bCourse postings) are either "required" readings, or "recommended."

+Accessible Format: Reading materials (especially bCourse postings) should be provided in an "accessible format," e.g., clearly legible, "clean" (without stray marks, highlighting, or mark-ups), and whenever possible, in a Word Document or word-searchable PDF.

For more information about accommodations for students with disabilities, please contact the Disabled Students Program at 510-642-0518 or email DSP Director Paul Hippolitus [hippolitus@berkeley.edu](mailto:hippolitus@berkeley.edu). For more information about providing reading assignments in an accessible format, please contact Martha Velasquez directly at [dspamc@berkeley.edu](mailto:dspamc@berkeley.edu).

- **ACCOMMODATION FOR PREGNANCY AND PARENTING** In compliance with Title IX of the Education Amendments of 1972, and with the California Education Code, Section 66281.7, it is the official policy of the University of California at Berkeley to not discriminate against or exclude any person on the basis of pregnancy or related conditions, and to provide reasonable accommodations to students as appropriate. Instructors are reminded of their responsibilities for excusing medically necessary absences for pregnancy and related conditions and making reasonable accommodations in the areas of class sessions, exams, tests, project deadlines, field trips, and any other required activities. For graduate

students, faculty advisors are reminded of policies regarding parental leave and the extension of normative time for academic milestones, as set out in the Guide to Graduate Policy.

Reasonable common sense, judgment and the pursuit of mutual goodwill should result in the positive resolution of conflicts. The regular campus appeals process or Title IX complaint process apply if a mutually satisfactory arrangement cannot be achieved. For more information about accommodations for student who are pregnant or parenting, please contact the Office for the Prevention of Harassment and Discrimination or (specific to graduate students) email the Graduate Division at [graddean@berkeley.edu](mailto:graddean@berkeley.edu).

- **READING, REVIEW, RECITATION (RRR) WEEK** The Reading, Review, Recitation (RRR) period?which are instructional days?before final exams provides students time to prepare for exams, to work on papers and projects, and to participate in optional review sessions and meetings with instructors. For the coming semesters, please keep these dates in mind:

In Fall 2015, formal classes end on Friday, December 4, 2015. RRR Week will take place between the last day of classes (December 4) and the first day of the final exam period (Monday, December 14, 2015).

In Spring 2016, classes end on Friday, April 29, 2016. RRR week will take place between the last day of classes (April 29) and the first day of the final exam period (Monday, May 9, 2016).

Please note that the regular semester classroom will NOT be available during the RRR week unless the instructor requests it through the departmental scheduler.

Presentations of capstone projects, oral presentations, and performances are permitted, although flexibility in scheduling may be required to accommodate students' individual schedules. The introduction of new material is not permitted. Mandatory exams or quizzes and other mandatory activities are also not permitted, with some very limited exceptions (capstone presentations, for example).

Please keep in mind that final exams and papers or projects substituting for final exams may not be due before the final exam week.

Detailed, updated guidelines on RRR week activities are available on the Academic Senate web site. The Office of the Registrar has posted answers to frequently-asked questions about the academic calendar.

In addition, the Center for Teaching and Learning has prepared some suggestions on making RRR week productive for instructors and students. If you have tips or ideas you would like to have added to this page, please email [teaching@berkeley.edu](mailto:teaching@berkeley.edu).

- COMMENCEMENT CEREMONIES AND FINAL EXAMS Campus policy stipulates that graduation ceremonies must take place after the conclusion of final examinations, with the exception of professional school ceremonies with graduate students only.