BIOSTATS 597D (1 credit)
Intermediate Statistical Computing
Sprint 2016 :: Weds 6:00-6:50pm :: LGRC A210

INSTRUCTOR
Xiangdong Gu
Data Scientist
MassMutual Financial Group
Email: ustcgxd@gmail.com

Teaching assistant: TBD
Email: TBD

CLASS RESOURCES
Course Website
Github
Recommended Textbook
Holden Karau et al (2015) *Learning Spark: Lightning-Fast Big Data Analysis*
Software (free downloads)
R
Rstudio
Python
Spark
SQLite

PREREQUISITES
None.

COURSE GOALS
The main goal of this course is to prepare students with necessary computing skills for future career as a statistician or data analyst/scientist. By end of this course, you should be able to use various tools to extract data from different sources, either structured for unstructured, and transform them into forms that are ready for analysis and modeling. You will also be able to build web based tools to deliver your data products using R Shiny.

COURSE OUTLINE

- SQL (4 weeks)
- Web scraping and regular expression (2 weeks)
- Introduction to Python (1-2 weeks)
- Spark (3-4 weeks)
- Shiny (2 weeks)
- Typical data science project lifecycle (1 week)

Expectations

Types of Assignments and Activities, with Grade Contributions

Participation/citizenship (10%) : I consider course citizenship to be a vital part of your grade. A few of the characteristics of good class citizens are: attending all course meetings, using office hours, asking questions, offering to answer questions, actively listening when others are talking, and posting to online discussion forums, among others. Citizenship is more a function of quality than quantity. Note that the "default" citizenship score is 5 out of 10, which allows students who actively and productively contribute to class to substantially increase their grade. Please note that good citizenship is different from "talking a lot," and it is quite possible to earn a low citizenship score because you fail to let others contribute.

Homework (90%): There will be regular homework assignments. Some of them will be graded as complete/incomplete, others will receive more detailed feedback and grading. Late assignments will not be accepted under any circumstances. Your lowest homework grade will be dropped.

Course Policies

Collaboration on homework is expected and encouraged, although you must write up your own assignment. No copying or cutting and pasting. Quizzes must be completed without assistance from your classmates.

Late assignments: Completing homework assignments on time will be vital to not falling behind in this course. It is expected that you hand in assignments on time. Late assignments will not be accepted.

Attendance is required. Absences (excused or not) will impact your participation grade.

All mobile devices that can/will be distracting to you or others during class must be turned off at the start of class and may not be used during class time.

Formal CEPH Course Competencies

- Describe the role biostatistics serves in public health.
- Describe conceptual frameworks (statistical literacy) in biostatistics
- Apply biostatistical methods to the design of studies in public health.
- Use computers to appropriately store, manage, manipulate and process data for a research study using modern software.
- Apply descriptive techniques commonly used to summarize public health data.
- Describe the basic concepts of probability, random variation and selected, commonly used, probability distributions.
- Select and perform the appropriate descriptive and inferential statistical methods in selected basic study design settings.
- Apply basic informatics techniques with vital statistics and public health records in the description of public health characteristics.
- Develop written and oral presentations based on statistical analyses for both public health professionals and educated lay audiences.

Academic Honesty Policy Statement Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst.

Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. The procedures outlined below are intended to provide an efficient and orderly process by which action may be taken if it appears that academic dishonesty has occurred and by which students may appeal such actions.

Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent. For more information about what constitutes academic dishonesty, please see the Dean of Students website.

Disability Statement The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier-free campus. If you are in need of accommodation for a documented disability, register with Disability Services to have an accommodation letter sent to your faculty. It is your responsibility to initiate these services and to communicate with faculty ahead of time to manage accommodations in a timely manner. For more information, consult the Disability Services website.