GMS 6803: Data Science for Clinical Research (3 credit hours)

Fall 2017

LOCATION: HPNP Room G-201

CLASS HOURS: Thursday 1:55 pm – 4:55 pm

INSTRUCTOR:

Jiang Bian, PhD
Clinical and Translational Research Building (CTRB) 3228
Phone: (352) 273-8878
Email: bianjiang@ufl.edu

François Modave, PhD
Clinical and Translational Research Building (CTRB) 3224
Phone: (352) 294-5984
Email: modavefp@ufl.edu

COURSE OVERVIEW:

GMS 6803 provides students with an introduction to a wide range of concepts and techniques in data science as they apply to biomedical and clinical research. Data is an essential component of biomedical and clinical research. It is critical for students to understand and gain practical experience with the entire life cycle of data, from data collection to data analysis to the dissemination and archiving of valuable results. In this course, students are introduced to the broad landscape of data science for biomedical and clinical research: learn how to design and implement computerized databases for data collection, perform basic query and reporting operations, prepare databases for analytical tasks, perform quality assurance procedures, and understand basic data analytical methods and approaches. Further, today's researchers are mining big datasets for patterns and trends that lead to new hypotheses and new discoveries. This course also aims to give students insight into tools, methods, and approaches for big data analytics in the biomedical domain. This course will be a foundation for students who are interested in becoming the next generation biomedical data scientists.

COURSE OBJECTIVES:

Teaching methods include lecture, discussion, and hands-on data assessment, analysis, and presentation. The goals of the course are:

- To provide basic understanding of the steps in the life cycle of data in biomedical and clinical research.
- To familiarize students with basic principles of data management.
- To deepen students' understanding of data structure, data standards, and data quality issues in data-intensive biomedical researches.
- To help students understand ethic and legal issues when dealing with biomedical and clinical data.
- To introduce the concepts of big data, and the associated tools, methods and approaches.
- To give students access to exploratory data analysis techniques and tools.

TEXTBOOKS/READING MATERIALS:

The following are suggested reading materials, however, they are not required. The instructor will distribute lecture handouts when necessary.

Recommended text:

- Grus J. Data Science from Scratch: First Principles with Python. 1st Edition. O'Reilly Media. 2015.
- McKinney W. Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. Sebastopol, Calif: O'Reilly Media, Inc.; 2012.

Optional text:

- Prokscha S. Practical Guide to Clinical Data Management. 3rd ed. Boca Raton: CRC; 2012.
- McFadden E. Management of Data in Clinical Trials. 2nd ed. Hoboken, N.J.: Wiley-Interscience; 2007.

Additional references:

- Krishnankutty B, Bellary S, Kumar NB, Moodahadu LS (2012). Data management in clinical research: an overview. Indian J Pharmacol 2012; 44(2):168.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009; 42(2):377-381.
- Greenes RA, Pappalardo AN, Marble CW, Barnett GO. Design and implementation of a clinical data management system. Comput Biomed Res 1969; 2(5):469-485.
- Sarkar IN. Biomedical informatics and translational medicine. J Transl Med 2010; 8(1):22.
- Shortliffe EH, Cimino JJ. Biomedical informatics. Springer Science+ Business Media, LLC; 2006.
- Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill D, Kania R, Schaeffer M, St Pierre S, Twigger S, White O, Rhee SY. Big data: The future of biocuration. Nature 2008; 455(7209):47-50.
- Antezana E, Kuiper M, Mironov V. Biological knowledge management: the emerging role of the Semantic Web technologies. Brief Bioinform 2009; 10(4):392-407.

OFFICE HOURS:

Office hour is by request. Please email the instructor for an appointment in advance. Likely we can address the questions over email. If not, please make an appointment, and I will try to accommodate your schedule.

PREREQUISITES:

Although the main goal of the course is not to teach programming, a fair amount of coding is required to complete the course work. Students are required to have basic computer-related skills and knowledge (e.g., operating system(s)). Prior experience with one or more data management software systems (e.g., MS Excel, Access, SQL, etc.) is required. Prior experience with one or more programming languages (e.g., general programming languages such as Java and C/C++; scripting languages such as Python, Lua and Ruby; and statistical computing languages such as SAS, SPSS, R and Matlab) is required.

GRADE COMPOSITION:

Attendance: 5%
Homework assignments: 40%
Midterm (project proposal and presentation): 25%
Final (project report and presentation): 30%

Homework assignments:

Assignments consist of writing critiques for research papers (6) and a small programming project (1).

Students will be asked to read research articles in topics related to data science and to write a critique of the paper. Students are expected to discuss these research articles in class. Please do not copy and rearrange the sentences in the original paper, which will result in low a grade.

Students are expected to finish one to two small programming projects. These programming exercises are often simple and their implementations are often easy to find online. You can use these online implementations as references. However, you will be penalized if you merely copy others' work. I reserve the rights to ask you to explain your code line-by-line.

Assignment rules: You are required to compliant with these rules.

- Your assignment must be turned in no later than 11:59 pm on the day that it is due.
- Late homework assignments will NOT be accepted, unless you have a ***prior*** approval from the instructor.
- No handwritten assignment. All assignments need to be submitted electronically either by email or the online system (will be clarified at the beginning of the course).
- DO NOT COPY OTHERS' HOMEWORK. There is zero-tolerance. The one who copy the homework will receive 0 point; and the one who is copied will get only 50% of the points that he/she should have received.
- You can work with others (e.g., discuss, consult, etc.) on a homework assignment. And, if you work on a homework assignment with other students in the course, you are required to list their names when you turn in the assignment. Plagiarism will receive 0 point.
- Searching for a solution on the web—and then submitting it as your answer for a homework assignment—will be considered a violation.

Course project:

The final product of the course is a course project, which consists of 55% of overall the grade. Each student is required to complete a course project. You can collaborate with other students as a team. However, each team can have up to three (3) members. Exception can only be made with written explanation and subject to the instructor's approval. And, please clearly delineate roles and responsibilities of each team member. Your final grade of the course project will be adjusted based on your contribution (e.g., merely presenting the project in the final presentation is NOT a contribution).

I will distribute a list of project ideas in the first week of the class as a reference. A systematic review of related topics is acceptable. You are encouraged to come up novel ideas related to the course. You will conduct extensive background research (e.g., literature review, identify and collect potential datasets, conduct preliminary data analysis, etc.), and you are expected to write a project proposal and give a presentation during the midterm. Please follow the following requirements for the project proposal and presentation.

Project proposal requirements:

- Cover Page: Include title and list of team members.
- Abstract: Up to 1 page. Explain the motivation for the work to be accomplished.
- Project description: Up to five (5) pages, and please include the following:
    - Specific Aims/Objectives
    - Background and Significance
    - Approach/Research Design (preliminary data and analysis if applicable)
    - Timeline
- Literature cited (no page limit); please follow the Vancouver or JAMA style.

Proposals must use single column and single spacing; Arial or Times New Roman font; font size no smaller than 11 point; tables and figure labels can be in 10 point; minimum 0.5 inch margins.

Midterm (proposal) presentation:

- Up to fifteen (15) slides and no more than 15 minutes of presentation with 10 minutes Q&A.
- Please send the slides to the instructor at least three (3) days in advance.

Each project team is expected to turn in a final project report, associated code and datasets (or reference to used datasets), and a group presentation.

Project report requirements: the project report can be up to ten (10) pages (including references), and please structure the report to include:

- Title (14 point typeface) and names of each team member
- Abstract: no more than 250 words summarizing the project.
- Introduction: a short background and objective(s) of the study.
- Methods: design, setting, dataset, approaches, and main outcome measurements.
- Results: key findings
- Discussion: key conclusions with direct reference to the implications of the methods and/or results.
- References: please follow the Vancouver style.

Final project presentation:

- Up to twenty-five (25) slides and no more than 25 minutes of presentation with 10 minutes Q&A.
- Please send the slides to the instructor at least three (3) days in advance.

Attendance policy:

Class attendance is mandatory. Excused absences follow the criteria of the UF Graduate Catalogue (e.g., illness, serious family emergency, military obligations, religious holidays), and should be communicated to the instructor prior to the missed class day when possible. UF rules require attendance during the first two course sessions. Missing more than three scheduled sessions will result in a failure. Regardless of attendance, students are responsible for all material presented in class and meeting the scheduled due dates for class assignments. Finally, students should read the assigned readings prior to the class meetings, and be prepared to discuss the material for each session.

Grading scale:

Letter Grade	Grade Points	Grade Percentage
A	4.0	95-100
A-	3.67	90-94
B+	3.33	87-89
B	3.0	83-86
B-	2.67	80-82
C+	2.33	77-79
C	2.0	73-76
C-	1.67	70-72
D+	1.33	67-69
D	1.0	63-66
D-	.67	60-62
E	0	< 59

For more detail on letter grades and related University of Florida policies, please see the Grades and Grading Policies at http://gradcatalog.ufl.edu/content.php?catoid=6&navoid=1219#grades.

Make-up policy: Students are allowed to make up work only as the result of illness or other unanticipated circumstances. In the event of such emergency, documentation will be required in conformance with University policy. Work missed for any other reason will earn a grade of zero.

UF POLICIES:

University policy on accommodation students with disabilities: Students requesting accommodation for disabilities must first register with the Dean of Students Office (http://www.dso.ufl.edu/drc/). The Dean of Students Office will provide documentation to the student who must then provide this documentation to the instructor when requesting accommodation. You must submit this documentation prior to submitting assignments or taking the quizzes or exams. Accommodations are not retroactive, therefore, students should contact the office as soon as possible in the term for which they are seeking accommodations.

University policy on academic misconduct: Academic honesty and integrity are fundamental values of the University community. Students should be sure that they understand the UF Student Honor Code at http://www.dso.ufl.edu/students.php. You are expected and required to comply with the University's academic honesty policy (University of Florida Rules 6C1-4.017 Student Affairs: Academic Honesty Guidelines, available at http://regulations.ufl.edu/chapter4/4017.pdf). Cheating, plagiarism, and other forms of academic dishonesty will not be tolerated. Note that misrepresentation of the truth for academic gain (e.g., misrepresenting your personal circumstances to get special consideration) constitutes cheating under the University of Florida Academic Honesty Guidelines

Netiquette – communication courtesy: All members of the class are expected to follow rules of common courtesy in all email messages, threaded discussions, and chats. The first instance of clearly rude and/or inappropriate behavior will result in a warning. The second instance will result in a deduction of five percentage points from your overall grade. The third instance will result in a drop of a letter grade (A to B, A- to B-, and so on).

GETTING HELP:

For issues with technical difficulties for E-learning in Sakai, please contact the UF Help Desk at:

- learning-support@ufl.edu
- (352) 392-HELP - select option 2
- https://lss.at.ufl.edu/help.shtml

COURSE SCHEDULE (TENTATIVE):

The course schedule is subjected to change according to students' background and interests based on the survey conducted at the beginning of the class.

	Topic	**Notes**
1	Introduction and course overview: data science articulated, history and context, examples, technology landscape	Discussion of course project
2	Data science toolbox: ideas and tools 1) Ideas behind turning data into actionable knowledge 2) Tools that are commonly used in the realm of data science Topics in machine learning: basic concepts • Lecture on uncertainty	
3	Python tutorial 1	

4	Python tutorial 2	
5	Topics in machine learning: basic concepts • Lecture on probability inference	
6	Topics in machine learning • Lecture on supervised learning	
7	Supervised learning in Python (scikit learn)	
8	Topics in machine learning • Lecture on unsupervised learning	Project proposal and team presentations
9	Presenting and interpreting analytical results: visualization, data products, visual data analytics	
10	Database and data management • Pandas	
11	Introduction to Semantic Web technologies Linking data: longitudinal studies and linking different data sources;	
12	Big Data: cloud-computing, parallel computing paradigms, big data analytic	
13	Special Topics: graph-structured data (network analysis), and textural data (NLP)	
14	Social Media Analysis – 1	Discussion of course project
15	Social Media Analysis – 2	
16	Course project presentations	