TEXAS A&M UNIVERSITY HEALTH SCIENCE CENTER
SCHOOL OF PUBLIC HEALTH

PHPM 672 Data Science for Health Services Research
PHPM 677 Data Science in Public Health
Spring 2015

COURSE WEBSITE
http://pinformatics.tamhsc.edu/phpm672/
You are required to check the class website regularly (at least three times a week) as it will have important class announcements.

CLASS SESSION(S)
Course Time: Tuesdays, 9 am to 12 pm
Location: SPH Classroom Building 119 (Computer Lab)

INSTRUCTOR
Hye-chung Kum, PhD, MSW;
Building A, Rm 312; Office hours: schedule via email (kum dot tamhsc dot edu)

TEACHING ASSISTANT
Obioma Nwaiwu (Building A, Rm 308)
Office hours: Tues (5:30-7:30) & Friday (12-2) in the computer lab (Classroom Building 119)

COURSE DESCRIPTION
Data science is the systematic study of digital data using scientific techniques of observation, theory development, systematic analysis, hypothesis testing, and rigorous validation. Data scientists are those that can apply data science to continuously changing deluge of digital raw data that are often inconsistent and erroneous to extract and deliver actionable knowledge in a timely manner. Data scientists are interdisciplinary scientists who have a combination of skills in statistics and data management (to turn raw data into information), programming (to build the data pipeline and infrastructure for efficient processing of raw data into information), and domain expertise (to (a) correctly interpret the data, information, and results, (b) know what information is required, and (c) how accurate it needs to be).

The primary purpose of this course is to apply data science to health data for health services research in order to improve the three core dimensions in healthcare: (1) improve quality, (2) reduce costs, and (3) improve access.

PREREQUISITES
PHPM672: PHPM 601, PHPM 671, STAT 652, or approval by the instructor.
PHPM677: PHEB 602 Biostatics I, or approval by the instructor.

COURSE LEARNING OBJECTIVES
The following Table presents core/discipline specific course objectives relative to competencies**:**

SPH Competencies: "Upon completion of this course a student should be able to …"	Course Objectives (LINKED TO Competencies)
PHPM672: PhD Competencies	
Conceive of a research design that would allow these hypotheses to be tested in a manner that would stand up to peer review, including appropriate hypothesis-generating and hypothesis-testing research.	Conceive of and understand the different methods to build the required measures for the analysis using the raw data available, and conduct sensitivity analysis to understand the impact of the different methods for building the same measure, then select the most appropriate measure for the given research.

Execute quantitative and qualitative analytical techniques appropriate to the research design and data studied.	Execute the planned data intensive research in a tractable and replicable manner by building a data flow process for converting the raw data into the analytic data, and then to conduct the analysis into results. The process should be modular so that it is agile for quick and accurate modifications to the research plan to adapt as necessary during execution of the plan, including the revisions to respond to review comments from journal submissions.
Draw appropriate conclusions about the research undertaken.	Draw appropriate conclusions from the analysis using available observational (operational) data fully understanding the limitations which result in selection biases and measurement validity and reliability issues including appropriate interpretations of results from sensitivity analysis.
PHPM677: MPH Competencies	
Communications and Informatics Use information technology to access, evaluate, and interpret public health data. (6) Use informatics and communication methods to advocate for community public health programs and policies. (7) Use informatics methods and resources as strategic tools to promote public health. (5) Collaborate with communication and informatics specialists in the process of design, implementation, and evaluation of public health programs. (8) Describe how the public health information infrastructure is used to collect, process, maintain, and disseminate data. (3) **PHPM Core** Communicate health policy and management issues using appropriate channels and technologies. (5) **Professionalism and Ethics** Apply evidence-based principles and the scientific knowledge base to critical evaluation and decision-making in public health. (3)	Conceive of and describe the different informatics methods to: • build the required measures for the real world problem (e.g. design, implementation, evaluation, strategize, advocate) using real world raw data available • conduct sensitivity analysis to identify the impact of the different methods for building the same measures • select the most appropriate measure for the given research question Execute the planned data intensive analysis in a tractable and replicable manner: • by building a data flow process for converting the raw data into the analytic data, • to conduct the analysis into results in order to answer the real world question Develop and implement a modular, agile, accurate, and adaptable data analysis process Draw appropriate conclusions from the analysis. Apply available observational (operational) data and define and describe the limitations which result in selection biases , and measurement validity and reliability issues Identify and apply appropriate interpretations of results from sensitivity analysis.
Communications and Informatics Apply legal and ethical principles to the use of information technology and resources in public health settings. (9)	Describe the issues of privacy in sensitive health data Define and explain the different security technology available to protect the sensitive data.
Communications and Informatics Discuss the influences of social, organizational and individual factors on the use of information technology by end users. (2)	Examine the differential impact of the digital world by social and individual factors

COURSE REQUIREMENTS

This course requires substantial programming, submission of programming assignments bi-weekly, a midterm, and completion of a final project.

Required Text

The little SAS book (NOT the enterprise guide version). Online book available at the library.

Other Required Readings

Additional materials beyond that found in the texts will be required reading.
These materials will be available on the course website.

Additional Recommended, but NOT required texts/materials

https://idre.ucla.edu/stat/sas

Assessment and Grading Policy

Student grades will be based on:

[Homework Assignments (6 * 8 points each)] …… [48] points [48%]
[Midterm Exam] …….…….. [20] points [or 20%]
[Final Project (different requirements for PHPM672 & PHPM677)]… [32] points [or 32%]

Homework Assignments

There will be a homework assignment due roughly every other week. Homework assignments and related materials will be made available on the course website. **I strongly advise students to begin to work on their homework assignments soon after they are assigned, so that if they need help from the instructo r they have time to ask and receive assistance.** There will be ***six (6)*** regular HW assignments during the semester. The overall weight of the assignments on the final course grade is ***48%, HALF the grade.***

All assignments are due at 11:59pm the day before the class they are due.

Late Assignments. Each student will be allowed one late assignment, due 7 days from the due date. NO other late assignments or make up will be accepted.

Collaboration: Collaboration on assignments, in class labs and homeworks, IS encouraged. However, what you hand in must be in your own writing/typing. Good scholarship requires that all collaborations must be acknowledged. Thus, if you collaborate on the solution of the problem set, we expect that you list your collaborators at the top of the page. Collaboration on in-class evaluations (quizzes, mid-terms, and the final exams) is, of course, a violation. This includes a discussion of questions on a quiz, midterm, or final with students from sections that has not yet taken the evaluation.

Plagiarism: If you consult any outside sources when doing your work, you are expect to further document these sources. Give credit where credit is due. Plagiarism will not be tolerated.

All handed in homework should state at the top any assistance with debugging and programming, as well as citations of any program segments copied from a website.

Seeking Help: Language classes are cumulative; don't fall behind. Ask for assistance from your fellow students, TA, or from the instructor, especially if you find yourself struggling. But remember in the end, programming classes are like skill building math classes. Either you know how to do them or not. No amount of watching others or the instructor program or debug it will suffice for you to build these skills. Only your hard work to work through them will build your skills. It is a lot of work to build these skills, but data skills are highly valuable in the job market in the modern digital world, so it will be well worth your effort. Furthermore, once you learn to think in this manner, it's not something you forget.

Elegance: There is always more than one way to say something, but some ways will be more "elegant" than others. You will learn to recognize elegant expressions as you become more familiar with a language and use the elegant idioms as you become more skilled.

Assignments and Exam Schedule

	Date	Assignment Given	Assignment Due
1	1/20/2015	Assignment 1	
2	1/27/2015	Assignment 2	Assignment 1
3	2/3/2015		
4	2/10/2015	Assignment 3	Assignment 2
5	2/17/2015		
6	2/24/2015	Assignment 4	Assignment 3
7	3/3/2015		
8	3/10/2015	Midterm	
	3/17/2015	SPRING BREAK	
9	3/24/2015	Assignment 5	Assignment 4
10	3/31/2015	Assignment 6	Assignment 5
11	4/7/2015		
12	4/14/2015	Final Project (FP)	Assignment 6
13	4/21/2015		FP MileStone 1
14	4/28/2015		FP MileStone 2
15	5/5/2015		Final Project

COURSE TOPICS

The detailed schedule of classes is posted on the course website.

The following topics in data science will be covered

- What is data science? What is big data?
- How is big data different from traditional sources of data ?
- How do you effectively convert big data to useful information and knowledge?
- What are the privacy issues in using sensitive data for research, and how should we address these issues?

The following programming topics will be covered: variables, assignments, conditional logic, loops, control flow and program design, indirection (functions/macros), arrays, fie I/O, debugging, and reshaping tables.

OTHER RELEVANT MATERIAL

Academic Integrity Statement and Policy

Academic integrity is the pursuit of scholarly activity free from fraud and deception and is an educational objective of this institution. Students are expected to adhere to all TAMUS, TAMU, HSC, and School policies regarding academic integrity and classroom conduct. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used, or tampering with the academic work of another student. Individuals found guilty of academic dishonesty may be dismissed from the degree program, and at a minimum will receive an F for the course. It is the student's responsibility to have a clear understanding of how to reference other individuals' work, as well as having a clear understanding in general as to the various aspects of academic dishonesty. A tutorial on this issue is available at: http://srph.tamhsc.edu/academic-affairs/academic-integrity.html. A plagiarism tutorial can be found in Blackboard. Information on the Aggie Honor Code can be found at http://aggiehonor.tamu.edu .
Remember: *"An Aggie does not lie, cheat, or steal, or tolerate those who do."*

Americans with Disabilities Act (ADA)

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring an accommodation, please contact Disability Services, in Cain Hall, Room B118, or call 845-1637. For additional information visit http://disability.tamu.edu

Blackboard Statement

Blackboard (Bb): This course will be delivered via Blackboard, your online learning community, where you will interact with your classmates and with me. Within the course Blackboard site you will access the learning materials, tutorials, and syllabus; discuss issues; submit assignments; take quizzes; email other students and the instructor; participate in online activities; and display your projects.

In order to access the course material you will need to go to https://tamhsc.blackboard.com or look for Quick Links on the bottom of the School's homepage. Please do not contact your instructor with technical problems. If you are having a technical problem with the course, review the Blackboard Learn Tutorials (at the top-right of School's Office of Assessment and Instructional Technology website). Please note that the Blackboard emails and the SRPH emails are the same. If you continue to have trouble accessing the course web site please contact John Lingsweiler in the School's Office of Assessment and Instructional Technology. John may be reached at (979) 458-3032 or at lingsweiler@tamhsc.edu. You will need to possess the required computing technology to be successful in an online course. All computing problems or other technical issues can be routed to the TAMHSC Help Desk at helpdesk@tamhsc.edu via E-mail, or phoned to 979-862-8029 or 1-800-799-7472 Important!!! Save your work as you go along. Nothing is more discouraging than to lose an assignment due to a computer hang ups! You may want to also make hard copies of your work to have "proof" and save yourself time and trouble!

End of Course Evaluation

Constructive feedback from students on course evaluations is taken very seriously at the School of Rural Public Health. I am asking for your assistance in helping the School in its assessment of courses and faculty through your participation in the evaluation of your courses. As public health professionals you will one day have the responsibility to evaluate colleagues and health initiatives. The School views providing feedback on the School's courses as part of your professional responsibility.

FERPA Statement

The Federal Education Rights & Privacy Act requires that we advise students that by registering for this course, their HSC assigned e-mail address will be revealed to classmates and the instructor. By continuing your enrollment in the course you acknowledge your understanding of this policy. By enrolling in this course you agree to the following statement: "I understand that as a result of registering for this course, my HSC/Blackboard assigned e-mail address will be revealed to classmates and the instructor."

Equal Opportunity Statement

The Texas A&M Health Science Center is an Equal Opportunity/ Affirmative Action employer. Inquiries regarding nondiscrimination policies may be directed to the Human Resources Officer by phone at (979) 436-9208, email hr@tamhsc.edu, or by mail at 200 Technology Way, College Station, TX 77845.

Disclaimer

This syllabus is representative of materials that will be covered in this class; it is not a contract between the student and the institution. It is subject to change. These changes will be communicated via email or posted as announcements. If you have any problems related to this course, please feel free to discuss them with the instructor.

Plagiarism Virtual Course

Plagiarism is the leading form of academic dishonesty that the School of Rural Public Health has to address. As a SRPH student, you are responsible for knowing what plagiarism is and how to avoid it. All SRPH students are automatically enrolled in Plagiarism Virtual Course on Blackboard E-Learning. This virtual course provides you with information and examples related to plagiarism in an effort to reduce the number of reported incidents. Please find a tutorial and resources under "Content." In addition, please find Turnitin, a software package that allows you to check whether you may have plagiarized your document. Please see Phuong Huynh: phuong@srph.tamhsc.edu for additional information.

University Attendance Policy Statement

The University views class attendance as the responsibility of an individual student. Attendance is essential to complete the course successfully. University rules related to excused and unexcused absences are located on-line at http://student-rules.tamu.edu/rule07.