

---

**University of Florida**  
**College of Public Health & Health Professions Syllabus**  
**PHC 7065: Critical Skills in Data Manipulation for Population Science (2 credit hours)**  
Spring 2021  
Delivery Format: Online, M 12:50-14:45  
Canvas Course Website: <https://elearning.ufl.edu/>

---

### **Instructor Information**

Hui Hu, PhD  
Assistant Professor  
Department of Epidemiology  
Office: 2004 Mowry Road, Clinical and Translational Research Building (CTRB) #4224  
Phone: (352) 294-5944  
Email: [huihu@ufl.edu](mailto:huihu@ufl.edu)  
Office hours: By appointment  
Preferred course communications: Email, office hours

### **Prerequisites**

PHC 6052 and PHC 6000 (or equivalent), or instructor permission.

### **Purpose and Outcome**

*Course Overview.* This course focuses on providing basic knowledge and skills needed in data manipulation for population science. Included will be: data context and concepts; relational databases; data collection and extraction; parallel manipulation of massive datasets; NoSQL systems and concepts. The course is designed for advanced students to learn the “code of best practice” for data engineering in population science.

*Course Objectives.* The overall goal of this course is to develop competence in data engineering for population science. Specific outcome competencies that are among the goals of this course include, but are not limited to, competency in:

1. The understanding of different data - Students will become familiar with common structured, semi-structured, and unstructured data types used in population science.
2. The concept and implementation of relational databases - Students will learn and understand relational databases and relational algebra, and gain the experience and skills with SQL.
3. The understanding of NoSQL systems - Students will learn the context of NoSQL systems and be familiar with popular NoSQL systems such as MongoDB.
4. The development of skills in data collection and extraction - Students will acquire the skills in data collections and extractions from multiple sources such as querying APIs.
5. The skills to manipulate data at scale - Students will get experience in popular algorithms and tools in parallel manipulation of massive datasets such as MapReduce and Spark.

6. Ethical maintenance to human subject data – Students will learn the important issues regarding data storage and backup, data sharing and transferring, and meta-data maintenance.

*Instructional Methods.* We will meet for 13 weeks. Each week includes a recorded lecture and a live recitation session that lasts around 1 hour (start at 12:50pm on Monday). Students will be expected to run VirtualBox on their laptops for the labs and recitation sessions unless otherwise required. Recitations are weekly live sessions with your TA that will answer your questions and explain solutions to your homework and exam. For students in the Certificate in Psychiatric Epidemiology (CPE) program, if you are not able to attend the recitation, you can review a recording of the session, posted the day after. Grades will be based on attendance and participation (5%), homework assignments (25%), mid-term exam (35%), and final project (35%).

## Description of Course Content

### Course Schedule.

Date	Agenda (● = topic, ◇ = Lab, ◆ = Recitation)	Assignment Due (% of Final Grade*)	Project Due (% of Final Grade)
January 11, 2021	● Course overview (live session) ◇ Virtual machine setup	-	-
January 25, 2021	● Introduction to R ◇ R crash course ◆ VM setup troubleshoot, R crash course Q&A	-	-
February 1, 2021	● Basic Structured Query Language ◇ SQL 1 ◆ Assignment 1 solution	Assignment 1 (5%)	-
February 8, 2021	● Data models and relational SQL ◇ SQL 2 ◆ SQL Q&A	-	-
February 15, 2021	● Many-to-many relationships in SQL ◇ SQL 3 ◆ SQL Q&A	-	-
February 22, 2021	◇ In-class exercise: SQL ◆ Assignment 2 solution	Assignment 2 (5%)	Form a project team (0%)
March 1, 2021	◇ Mid-term exam: SQL ◆ Assignment 3 solution	Assignment 3 (5%)	-
March 8, 2021	● Review of mid-term exam and SQL	-	Project proposal (5%)
March 15, 2021	● Access web data ◇ http ◆ Mid-term exam solution	-	-
March 22, 2021	● In-class exercise: access web data ◆ Web data Q&A	-	-
March 29, 2021	● Spatial data ◇ PostGIS ◆ Spatial data Q&A	-	-
April 5, 2021	● NoSQL databases ◇ MongoDB ◆ Assignment 4 solution	Assignment 4 (5%)	-
April 12, 2021	● Student presentations (live session)	Assignment 5 (5%)	-

\* Only the highest 5 grades will count towards the final course grade.

*Course Materials and Technology.* Readings for the course will be assigned each week for the next lecture. There is no required textbook. For technical support for this class, please contact the UF Help Desk at: [Learning-support@ufl.edu](mailto:Learning-support@ufl.edu), (352) 392-HELP – select option 2, or <https://lss.at.ufl.edu/help.shtml>.

## Academic Requirements and Grading

### *Grade Composition.*

- Attendance and participation: 5%
- Homework assignments: 25%
- Mid-term exam: 35%
- Project proposal: 5%
- Final project presentation: 10%
- Final project report: 20%

*Attendance and participation.* Class attendance is mandatory (only students in the CPE program cannot attend the live recitation sessions). 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. Each unexcused absence results in a 1.5% point deduction from the final grade. Missing more than three scheduled sessions without excuse (each live recitation session is about 1 hour) will result in a failure. Students are responsible for all material presented in class and meeting the scheduled due dates for class assignments.

*Homework assignments.* A total of 6 assignments will be given, and the highest 5 grades will count towards the final grade. All assignments are small programming projects. You are required to compliant with the following assignment rules:

- Your assignment must be turned in no later than the due time.
- Late homework assignments will NOT be accepted, unless you have a formal proof of the exception (e.g., a written doctor note, a police ticket, etc.).
- 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 40% of the final 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 two (2) members. Exception can only be made with written explanation and subject to the instructor's approval. In addition, 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). Two options are available for the course project. **OPTION A:** Pick one article from a list of publications (which will be made available mid-February), and write codes to reproduce its data engineering and descriptive analyses. **OPTION B:** Come up your own ideas for the course project. It is required to include at least ONE non-traditional data source (e.g. spatial data, web data, etc.) other than the traditional survey data. Please note that although an analyses/modelling section is required for the project, it is not the focus of this course. Instead, you should focus more on the data accessing and engineering part. Please follow the detailed requirements below.

*Project proposal requirements:*

- Cover Page: Include title and list of team members.
- Project description: Up to one (1) page.
  - Specific Aims/Objectives: for those choosing option A, please cite the article you'd like to reproduce and briefly summarize the specific aims/objectives of the article. For those choosing option B, please state your aims/objectives.
  - Data Source: please provide details about the data and how it can be accessed
  - Preliminary Data Pipelines: please briefly describe the data engineering steps involved in this project
  - Timeline
- Literature cited (no page limit); please follow the Vancouver 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; 0.5 inch margins.

*Project report requirements:*

- For those choosing option A, the project report should be structured as an R Markdown Notebook, with all the codes, explanations to the codes, and outputs.
- For those choosing option B, 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 fifteen (15) slides and no more than 15 minutes of presentation with 5 minutes Q&A.

- Please send the slides to the instructor at least three (3) days in advance.

*Point system.*

Points earned	93-100	90-92	87-89	83-86	80-82	77-79	73-76	70-72	67-69	63-66	60-62	<60
Letter grade	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	E
Grade points	4.0	3.67	3.33	3.0	2.67	2.33	2.0	1.67	1.33	1.0	0.67	0.0

For greater detail on the meaning of letter grades and university policies related to them, see the Registrar’s Grade Policy regulations at: <http://catalog.ufl.edu/ugrad/current/regulations/info/grades.aspx>

**Student Expectations, Roles, and Opportunities for Input**

*Expectations Regarding Course Behavior.* Please come to class on time and be prepared to stay until the time scheduled as the end of class. We think your investment in the degree is worth maximizing your in-class experience, and we expect to provide materials that utilize the full, scheduled class times. The use of cell phones is not permitted. Please turn them off or, if you expect urgent calls, set them to “vibrate.” Please do not engage in “side conversations” while the instructor or a presenter is leading the class. If the material is unclear, other students are likely to have a similar question; you are strongly encouraged to ask in-class questions so that all students may benefit from the discussion.

*Communication Guidelines.* Assistance with course material is available during scheduled office hours or by appointment. Emailed questions are also welcome, and we aim to address all such inquiries within 24 hours of receipt (or on Monday if the email was sent on Friday). Please do not re-send the same question until the appropriate time frame has elapsed (24 hours or end of day Monday for emails sent on Friday). Student success and understanding is of the utmost importance, so each email receives careful consideration.

*Academic Integrity.* Students are expected to act in accordance with the University of Florida policy on academic integrity. As a student at the University of Florida, you have committed yourself to uphold the Honor Code, which includes the following pledge:

“We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity.”

You are expected to exhibit behavior consistent with this commitment to the UF academic community, and on all work submitted for credit at the University of Florida, the following pledge is either required or implied:

“On my honor, I have neither given nor received unauthorized aid in doing this assignment.”

It is your individual responsibility to know and comply with all university policies and procedures regarding academic integrity and the Student Honor Code. Violations of the Honor Code at the University of Florida will not be tolerated. Violations will be reported to the Dean of Students Office for consideration of disciplinary action. For additional information regarding

Academic Integrity, please see Student Conduct and Honor Code or the Graduate Student Website for additional details:

- <https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/>
- <http://gradschool.ufl.edu/students/introduction.html>

Please remember cheating, lying, misrepresentation, or plagiarism in any form is unacceptable and inexcusable behavior.

*Online Faculty Course Evaluation Process.* Students are expected to provide feedback on the quality of instruction in this course by completing online evaluations at <https://evaluation.ufl.edu>. Students will be given specific times when the evaluations can be made. Summary results of these assessments are available to students at <https://evaluations.ufl.edu/results/>.

It is very important to me that you are able to develop a set of data engineering tools in this course that you will find useful in your career. Naturally, the presentation of some tools may be stronger than others. Your feedback on this issue is extremely valuable to me; please feel free to comment on what strategies worked and which might be improved, as I will modify future versions of this course to leverage such knowledge of strengths and weaknesses. As an additional consideration, these evaluations are also useful at the University level, as they are examined in the context of faculty tenure and promotion procedures.

### **Online Synchronous Sessions**

Our class sessions may be audio visually recorded for students in the class to refer back and for enrolled students who are unable to attend live. Students who participate with their camera engaged or utilize a profile image are agreeing to have their video or image recorded. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate exclusively using the "chat" feature, which allows students to type questions and comments live. The chat will not be recorded or shared. As in all courses, unauthorized recording and unauthorized sharing of recorded materials is prohibited.

### **Support Services**

*Accommodations for Students with Disabilities.* If you require classroom accommodation because of a disability, you must register with the Dean of Students Office <http://www.dso.ufl.edu> within the first week of class. The Dean of Students Office will provide documentation of accommodations to you, which you must then give to me as the instructor of the course to receive accommodations. Please make sure you provide this letter to me by the end of the second week of the course. The College is committed to providing reasonable accommodations to assist students in their coursework.

*Counseling and Student Health.* Students sometimes experience stress from academic expectations and/or personal and interpersonal issues that may interfere with their academic performance. If you find yourself facing issues that have the potential to or are already

negatively affecting your coursework, you are encouraged to talk with an instructor and/or seek help through University resources available to you.

- The Counseling and Wellness Center (352) 392-1575 offers a variety of support services such as psychological assessment and intervention and assistance for math and test anxiety. Visit their web site for more information: <http://www.counseling.ufl.edu>. Online and in person assistance is available.
- You Matter We Care website: <http://www.umatter.ufl.edu/>. Your well-being is important to the University of Florida. The U Matter, We Care initiative is committed to creating a culture of care on our campus by encouraging members of our community to look out for one another and to reach out for help if a member of our community is in need. If you or a friend is in distress, please contact [umatter@ufl.edu](mailto:umatter@ufl.edu) so that the U Matter, We Care Team can reach out to the student in distress. A nighttime and weekend crisis counselor is available by phone at (352) 392-1575. The U Matter, We Care Team can help connect students to the many other helping resources available including, but not limited to, Victim Advocates, Housing staff, and the Counseling and Wellness Center. Please remember that asking for help is a sign of strength. In case of emergency, call 9-1-1.
- The Student Health Care Center at Shands is a satellite clinic of the main Student Health Care Center located on Fletcher Drive on Campus. Student Health at Shands offers a variety of clinical services. The clinic is located on the second floor of the Dental Tower in the Health Science Center. For more information, contact the clinic at 392-0628 or check out the website at: <https://shcc.ufl.edu/>.
- Crisis intervention is always available 24/7 from: Alachua County Crisis Center, (352) 264-6789, <http://www.alachuacounty.us/DEPTS/CSS/CRISISCENTER/Pages/CrisisCenter.aspx>. Do not wait until you reach a crisis to come in and talk with us. We have helped many students through stressful situations impacting their academic performance. You are not alone so do not be afraid to ask for assistance.