

**UNIVERSITY OF MICHIGAN
SCHOOL OF MEDICINE
DEPARTMENT OF LEARNING HEALTH SCIENCES**

**Exploratory Data Analysis for Health
LHS 610
Winter 2024**

CREDIT HOURS	3
PRE-REQUISITES	Graduate standing or permission of the instructor. Advisory pre-requisites: Students should have taken one course in statistics and have experience with either a statistical or general-purpose programming language.
CLASS SCHEDULE	Mondays: 10:00 am – 11:20 am Fridays: 10:00 am – 11:20 am
LOCATION	2903 THSL
FACULTY	Course Instructors Andrew Krumm, PhD (aekrumm@umich.edu) Karandeep Singh, MD, MMSc (karandeep@ucsd.edu) Graduate Student Instructor Zhe Zhao (zzhaozhe@med.umich.edu)
COURSE WEBSITE	Canvas

COURSE DESCRIPTION

This course introduces students to fundamental skills needed to explore health data. The course focuses on two large themes: (a) exploring health data in the context of the US healthcare system using the statistical programming language R, and (b) evaluating the potential role of machine learning to support health-related decisions. Learning from health data requires a solid grasp of data operations, data visualization, statistics, and machine learning, as well as an understanding of ethical and legal frameworks guiding health data privacy and security. The course focuses on giving students an introductory understanding of health data and its limitations, while developing knowledge, skills, tools, and techniques that are critical for data analysis. Special emphasis is placed on learning a tidyverse-based approach, which is a popular data analysis framework in R.

COURSE OBJECTIVES

By completion of the course students will be able to:

- Analyze the challenges of working with health data, including legal and methodological issues pertaining to the collection and secondary use of clinical data.
- Examine the features of a health-related question that make it important and answerable, including data sources that can be leveraged to address it.
- Display relationships between variables in health data, both visually and statistically.
- Apply machine learning methods to train and evaluate predictive models in health data.
- Apply common approaches to working with missing data.
- Describe how machine learning models can be implemented in healthcare settings.
- Demonstrate using R for research and analysis.

DROP-ADD Deadlines:

Tuesday, January 30 – full tuition refund – HILS doctoral students wishing to drop after this date must have permission from the HILS Program Director.

Tuesday, February 20 – half tuition refund.

CONTENT TOPICS

- Module 1: Introduction
- Module 2: Data Frames – 2 weeks
- Module 3: Tidy Data – 2 weeks
- Module 4: Formulating a Health-Related Question
- Module 5: Telling Stories with Plots
- Module 6: Hypothesis Testing
- Module 7: Interactive Data Analysis
- Module 8: Introduction to Machine Learning
- Module 9: Supervised Learning Algorithms
- Module 10: Handling Missing Data
- Module 11: Analyzing Health Text Data
- Module 12: Final Project

REQUIRED TEXTS AND OTHER MATERIALS

We will be using these free e-books in the course:

1. R for Data Science (<https://r4ds.had.co.nz/>)
2. Introduction to Statistical Learning, 2nd Ed. (<https://www.statlearning.com/>)
3. ISLR tidymodels book (<https://emilhvitefeldt.github.io/ISLR-tidymodels-labs/>)

TEACHING METHODS

This course has 2 sections: a **residential** section and an **online** section. The residential section is only available to residential students. The online section is only available to eligible online degree program students. Course content will be available in Canvas, and active discussion via our course's Slack page will be encouraged. Within the Canvas site, you will access learning materials, submit assignments, and view class announcements. This course will include both collaborative and individual assignments, described in further detail under the grading and evaluation section.

Residential section: The residential section will take place in-person, and a recording will be made available to students unable to attend. In-person classes will alternate between lecture and live coding sections.

Online section: The online section will take place asynchronously, though students are highly encouraged to communicate via Slack and attend office hours.

GRADING AND EVALUATION

Grading will be based on 3 evaluation methods. Sandboxes are based on content covered in the live coding class sections. Students are expected to follow along with code snippets that are presented and then apply what they have learned on a dataset. These are intended to be exploratory, collaborative, hands-on assessments. You will be graded on completion. In addition to sandboxes, there are homework assignments, which must be completed individually, and a final project. The weighting is detailed below:

1. 10 Sandboxes
 - a. 2 points each x 10 = 20 points
2. 6 Homework Assignments
 - a. 10 points each x 6 = 60 points
3. Final project – can be completed individually or in groups of up to 3
 - a. 20 points

NOTE: YOU MAY ONLY COLLABORATE WITH OTHER STUDENTS ON SANDBOXES. ANY OTHER COLLABORATION ON ASSIGNMENTS WILL BE CONSIDERED PLAGIARISM.

GRADING POLICIES

No late assignments will be accepted without penalty. Unless an agreement has been reached in advance with the instructor, the student will lose two points for every day that the assignment is late. For example, if the assignment is worth 10 points and is turned in one day late, the maximum number of points for that assignment will be 8/10. Two days late will result in the maximum number of points being 6/10, and so on. Graded assignments will be returned through Canvas. Final grades will be accessible through Wolverine Access; grades for each assignment will be posted through the Canvas course site. Overall course grades are posted within 72 hours of when the final course assignment component (typically the final course project report) is due. Letter grades will be assigned using this conversion scale:

A+	97-100	B+	87-89	C+	77-79	D+	67-69	F	<60
A	93-96	B	83-86	C	73-76	D	63-66		
A-	90-92	B-	80-82	C-	70-72	D-	60-62		

COURSE SCHEDULE

Module	Date	Class	Assigned	Due
	1/10	<i>Semester Begins</i>		
1	1/12	Lecture	Homework 1	1/24 11:59 pm
	1/15	<i>No Class, MLK holiday</i>		
1	1/19	Live Coding ^{*(KS)}	Sandbox 1	1/21 11:59 pm
2	1/22	Lecture ^{*(AK)}	Homework 2	2/04 11:59 pm
2	1/26	Live Coding	Sandbox 2	1/31 11:59 pm
2	1/29	Live Coding	Sandbox 2 (continued)	
3	2/02	Lecture	Homework 3	2/14 11:59 pm
3	2/05	Live Coding	Sandbox 3	2/11 11:59 pm
3	2/09	Live Coding ^{*(KS)}	Sandbox 3 (continued)	
4	2/12	Lecture	Homework 4	2/21 11:59 pm
4	2/16	In-Class Work on Homework 4		
5	2/19	Lecture		
5	2/23	Live Coding	Sandbox 4	3/04 11:59 pm
	2/24-3/03	<i>Spring Break</i>		
6	3/04	Lecture	Homework 5	3/14 11:59 pm
6	3/08	Live Coding ^{*(AK)}	Sandbox 5	3/11 11:59 pm
7	3/11	Lecture	Final Project	4/26 11:59 pm
7	3/15	Live Coding	Sandbox 6	3/17 11:59 pm
8	3/18	Lecture	Homework 6	4/07 11:59 pm
8	3/22	Live Coding	Sandbox 7	3/27 11:59 pm
8	3/25	Live Coding	Sandbox 7 (continued)	
9	3/29	Pre-Recorded Lecture		
9	4/01	Live Coding	Sandbox 8	4/03 11:59 pm
10	4/05	Lecture		
10	4/08	Live Coding	Sandbox 9	4/10 11:59 pm
11	4/12	Lecture		
11	4/15	Live Coding	Sandbox 10	4/17 11:59 pm
12	4/19	In-Class Work on Final Project		
12	4/22	In-Class Work on Final Project		
	4/23	<i>Classes End</i>		
	4/26	Final Project due (11:59 pm)		

* Instructor out of town

Please note that the syllabus is a progressive document which will be updated frequently throughout the semester. Please plan to check in at least weekly to review any changes in upcoming work.

DLHS Grade Grievance Policy

Regular feedback on graded assignments and timely resolution of any grading concerns is helpful for student learning and overall positive experience. Students are encouraged to discuss any assignment grade concerns with the instructor(s) within two weeks of the grades being posted on Canvas. Similarly, any concerns and grievances related to the overall course grades should be raised by contacting the faculty of record within two weeks of the course grades being posted on Wolverine Access.

For this course, the grade dispute arbitration will happen within the Department of Learning Health Sciences (DLHS). After initial discussion with the faculty of record for the course, there is an opportunity for a second look by another DLHS faculty member with appropriate expertise for resolving grade disputes. If the student and faculty are not able to come to an agreement about the grade dispute, the Associate Chair for Educational Programs or assigned designee will adjudicate the dispute. Per Rackham's Academic Dispute Resolution Process, if the concern is not resolved, the student may seek a formal resolution conference with the Medical School's Rackham Resolution Officer (see <https://rackham.umich.edu/academic-policies/section9/> for full policy and the contact information for the current resolution officer).

COURSE POLICIES

U-M Health Response

We each have a responsibility for protecting the collective health of our community. Applicable safety measures are described on the U-M Health Response website: <https://healthresponse.umich.edu/> Please bookmark this site for policies, guidance, and FAQs.

Attendance

This only applies to the residential section.

Students who are feeling ill should not come to class in-person. Class recordings will be made available to all students. While students are generally expected to attend class, we will not be taking attendance.

Expectations

This class aims to teach both a skill (data analysis in R) and a way of thinking about data. The inclusion of both aspects in the course can make this course challenging for students who are experiencing both aspects for the first time. Students are expected to participate actively with the material and to ask for help when needed. For Collaborative Sandboxes, students may ask help from any source (for example: the internet, peers, and instructors), while for all other assignments, students may only ask for assistance from instructors. Students are expected to provide feedback to colleagues respectfully and constructively. Students are expected to draw from their own experience as they develop an idea and analysis for their final project.

Digital Etiquette

When communicating with other students or instructors, please share (either directly or through your name signature) how you prefer to be addressed, including your name (or nickname) and pronouns. Please respect other students' preferences with regards to how they prefer to be addressed.

Students should turn off ringtones and other audible alerts on their phones/tablets/laptops before class sessions begin, whether in-person or online. The use of phones/tablets/laptops for non-course related activities, such as texting, emailing, and social media is distracting to your classmates and instructors. If there is an emergency, please excuse yourself from the classroom and/or the Zoom meeting to attend to it.

Issues related to hybrid or flipped courses

In this course, students will be expected to use the statistical programming language R for their analysis. We will experiment with different platforms for deploying R, including Google Colab as well as locally to one's own computer running Windows, Mac, or Linux operating systems. For students who do not own a computer or do not have the ability to install software on a computer (for example, a library computer), you may still access R from a web browser by going to <https://rstudio.cloud>, which has a free tier that allows 25 hours of computing time per month. If you have issues accessing RStudio or R, please contact the instructors as early as possible so we can look for alternative options.

INSTITUTIONAL POLICIES

Academic Integrity of Students

The academic community, like all communities, functions best when its members treat one another with honesty, fairness, respect, and trust. Your instructors expect students to work and study together to foster learning and understanding of the material. However, direct copying of homework, copying of homework from existing solutions, cheating on an exam, and other conduct that violates the academic integrity and ethical standards of the Rackham Graduate School community cannot be tolerated and will result in serious consequences and disciplinary action.

All written submissions must be your own, original work. Original work for narrative questions is not mere paraphrasing of someone else's completed answer: you must not share written answers with each other at all. At most, you should be working from notes you took while participating in a study session. Largely duplicate copies of the same assignment will receive an equal division of the total point score from the one piece of work.

You may incorporate selected excerpts, statements, or phrases from publications by other authors, but they must be clearly marked as quotations and must be attributed. If you build on the ideas of prior authors, you must cite their work. You may obtain copy editing assistance, and you may discuss your ideas with others, but all substantive writing and ideas must be your own or be explicitly attributed to another. If we suspect you have cheated (including plagiarism), at the very least you will receive a zero on the assignment. Rackham policy dictates that we must report every instance of academic dishonesty, no matter how small. Suspected academic misconduct will be handled by Drs. Piatt and Donnelly. <https://rackham.umich.edu/academic-policies/section8/>

Student Academic Dispute Procedures

Rackham's Academic Dispute Resolution Policy and Procedures are available to Rackham students who have a dispute or disagreement with faculty or staff about the equity and fairness of decisions or procedures that affect their academic standing, the conduct of their research, and progress toward the degree. Such issues may arise regarding fair and equal treatment in the conduct of a class, in the pursuit of the student's research, and in the grading or evaluation of academic work and research. Other issues may concern the equity and fairness of program, department, or Rackham policies. <https://rackham.umich.edu/academic-policies/section9/>

Diversity, Equity, and Inclusion

The Department of Learning Health Sciences (DLHS) is committed to developing the institutional mechanisms and norms necessary to promote the values of diversity, equity, and inclusion, both inside and outside our classrooms. To this end, DLHS upholds the expectations that all courses will: (1) be inclusive, (2) promote honest & respectful discussions, (3) follow multicultural ground rules, and (4) abide by UM policies and procedures. Inclusive courses are those in which teachers and learners co-create and co-sustain environments that support and encourage all members to participate equitably. <https://rackham.umich.edu/rackham-life/diversity-equity-and-inclusion/>

Accessibility and Accommodations

Students should speak with their instructors before or during the first week of classes regarding any special needs. Students seeking academic accommodations should register with Services for Students with Disabilities (SSD). SSD

arranges reasonable and appropriate academic accommodations for students with disabilities. Please visit <https://ssd.umich.edu/accommodations> for more information on student accommodations.

Student Parents and Caregivers

If you are a pregnant, parenting student, or primarily responsible for providing care for a loved one or family member, and you need any accommodations, please let the instructor know at your earliest convenience. You may also reach out to mcasep.org and CEW+ for resources and community support.

Student Mental Health and Well-Being

The University of Michigan is committed to advancing the mental health and wellbeing of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and <https://caps.umich.edu/> during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at (734) 764-8320 and <https://uhs.umich.edu/mentalhealthsvcs>, or for alcohol or drug concerns, see <https://uhs.umich.edu/aodprograms>.

For a listing of other mental health resources available on and off campus, visit: <https://caps.umich.edu/article/finding-your-community-provider-database-included>.

Sexual Misconduct/Sexual Harassment Reporting

Title IX prohibits sex discrimination to include sexual misconduct: harassment, domestic and dating violence, sexual assault, and stalking. If you or someone you know has been harassed or assaulted, you can receive confidential support and academic advocacy at the Sexual Assault Prevention and Awareness Center (SAPAC). SAPAC can be contacted on their 24-hour crisis line, 734-936-3333 and online at sapac.umich.edu. Alleged violations can be reported non-confidentially to the Office for Institutional Equity (OIE) at institutional.equity@umich.edu. Reports to law enforcement can be made to University of Michigan Police Department at 734-763-3434.

Support for Food Insecurity

Students across the country experience food insecurity at alarming rates. The Maize and Blue Cupboard at the University of Michigan provides food, kitchen and cooking supplies, personal and household items, and support services. For information about accessing their services, please visit <https://mbc.studentlife.umich.edu/>

Confidentiality and Mandatory Reporting

As instructors, one of our responsibilities is to help create a safe learning environment on our campus. Dr. Piatt also has a mandatory reporting responsibility related to her role as the Director of the Health Infrastructures and Learning Systems degree program. Instructors are required to share information regarding sexual misconduct or information about a crime that may have occurred on U-M's campus with the University. Students may speak to someone confidentially by contacting SAPAC's Crisis Line at (734) 936-3333.