

PUBHLTH 460 (3 credits)
Telling Stories with Data: Statistics, Modeling, and Data Visualization
Spring 2021 :: T/Th 11:30-12:45am

INSTRUCTORS

Nicholas G Reich, nick [at] umass.edu, [Reich Lab website](#), [@reichlab](#)

Lead teaching assistant: Anna Barry

Course assistants: Johanna Copeland, Sarah Gerstein, Dhruvi Vora, Nutch Wattanachit, Yilin Xu

Instructor and TA Office Hours: See Moodle

MATERIALS

Required Textbook

Kaplan, Daniel T. 2011. *Statistical Modeling: A Fresh Approach, 2nd Ed.*

Recommended Textbook (freely available online)

Diez D, Barr C, and Çetinkaya-Rundel M. 2015. *OpenIntro Statistics, 3rd Ed.*

Wickham H and Golemund G. 2017. *R for Data Science*

Software

R :: [r-project.org](#) (or just Google "r")

RStudio :: [rstudio.org](#)

PREREQUISITES

One of any of the following introductory stats courses taught at UMass: BIOSTAT 223, STAT 111, STAT 240, STAT 501, ResEcon 212, PSYCH 240. If you have not taken an intro stats course at UMass but still want to enroll in this course, you are encouraged to petition the instructor for permission, especially if any of the following apply: (a) you have taken AP Stats in high school, (b) you have taken a college-level intro stats course just not one of the ones listed above, or (c) you are confident in your quantitative skills and your ability to succeed in a fast-paced, advanced introductory course. Additionally, prior programming experience with R or concurrent enrollment in PUBHLTH 497D (Introduction to Statistical Computing with R) is required.

COURSE DESCRIPTION

The aim of this course is to provide students with the skills necessary to tell interesting and useful stories in real-world encounters with data. Specifically, they will develop the statistical and programming expertise necessary to analyze datasets with complex relationships between variables. Students will gain hands-on experience summarizing, visualizing, modeling, and analyzing data. Students will learn how to build statistical models that can be used to describe and evaluate multidimensional relationships that exist in the real world. Specific methods covered will include linear and logistic regression. Students will work with the R statistical computing language and by the end of the course will require substantial independent programming. The course will not provide introductory training in R programming. To the extent possible, the course will draw on real datasets from biological and biomedical applications. This course is designed for students who are looking for a second course in applied statistics/biostatistics (e.g. beyond BIOSTATS 223 or STAT 240), or an accelerated introduction to statistics and modern statistical computing.

LEARNING GOALS (*By the end of the course students will be able to...*)

- design data-driven experiments to answer specific questions,
- use data to identify and distinguish patterns of randomness vs. non-randomness,
- create powerful data visualizations that reveal and highlight important features of data or models,
- understand and critique statistical model equations as representations of a given real-world setting,
- formulate, fit, and interpret statistical models designed to answer specific scientific questions,
- weigh evidence for/against hypotheses about associations between variables,
- diagnose the appropriateness or "goodness-of-fit" of a given model,
- write concise, professional, and reproducible statistical analysis reports.

EXPECTATIONS

This course will require you to work thoughtfully, carefully, and independently and will require substantial work outside of class time. Because we will be using a more project-driven approach in this course, with assignments that will build upon one another into a final product, it is vital that you do not fall behind. If you feel as though you are falling behind or starting to lose a handle on the content, I expect you to talk to me during office hours or during a separate video conference appointment so that I can help as much as I can to set you back on track. Please do not wait to talk to me if you start to fall behind.

I also expect you to devote substantial outside-of-class time to your work for this course, typically involving 5-10 hours per week. I anticipate that this work will be divided among:

- finishing in-class activities
- reading assigned articles and chapters
- reviewing your notes
- working on assignments
- conducting project work
- preparing for exams

Things you should expect from me:

- timely feedback on assignments and quizzes
- response to questions via Moodle in < 2 working days (often sooner)
- attention to your questions related to coursework during office hours
- instruction in how to write, research, and debug R code

Things you should not expect from me:

- time for frequent non-office hour drop-in questions
- comments on a research project that is unrelated to your coursework
- writing your code for you or *extensive* debugging of your code

TYPES OF ASSIGNMENTS AND ACTIVITIES, WITH GRADE CONTRIBUTIONS

Problem sets (30%): There will be approximately five lab assignments and five coding challenges that you will complete over the course of the semester. Each problem set will have components that you will hand in for grading. Assignments, total possible point values, and due dates will be posted in advance on the course website. Some assignments will require you to submit a digital file with reproducible solutions, i.e. a knitr file that reproduces your answers. Late assignments will not be accepted under any circumstances. If a problem set (either a lab or a coding challenge) is not handed in on time, it will receive a grade of zero. I will drop your lowest two grades when calculating your final problem set grade.

Midterm exam (25%): There will be a mid-term exam in this course. Exact format and timing is still being determined.

Final Project (25%): In the second half of the course, you will develop and write your own data story as part of a small team. This project will be presented to your classmates. A separate handout will provide details.

Participation (10%) : Class participation is critical to succeeding in this class (see class policies section). Attending class, working with your assigned breakout group during class, participating in discussions, and submitting in-class assignments at the end of class will earn you a strong participation score. This is about showing up, being present, and doing the work.

Citizenship (10%) : Being a good class “citizen” also plays a large role in your final grade. A partial list 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, not interrupting others, helping to foster a non-judgmental and inclusive classroom environment, and participating on the Moodle forum (both asking and answering questions). Citizenship, unlike participation, is more a function of quality than quantity and can't just be earned by “showing up”. The default citizenship score is 5 out of 10.¹

¹Acknowledgments to Aaron Swoboda for introducing me to the concept of course citizenship and for some of this text.

COURSE POLICIES

Working together: Collaboration on all assignments is expected and encouraged, although you must write up your own assignment. **No copying or cutting and pasting.** Each project will contain an independent component which must be your own work. You may discuss your project with others and even solicit ideas and advice, but at the end of the day, you must complete all the analysis and write-up on your own. Any explicitly borrowed ideas or language must be cited or acknowledged appropriately. Copying of assignments will result in a zero grade for the assignment and appropriate consequences will be assessed in line with the campus-wide academic honesty policy (see below).

Remote learning: Being “fully present” for classtime is very important to keeping a respectful, cohesive, and strong learning environment for all students. This is challenging in an in-person setting, and even more challenging in a remote-learning setting. Please read below for additional guidelines about how to be a good course citizen in this remote course.

Cameras and online distraction: To the extent possible, I ask that all students keep their cameras on during class-time. I have found that this increases engagement with the activities and discussions going on and minimizes (although does not remove entirely!) the temptation to be digitally distracted. The expectation for this class is that you will attend each class from a location that is as distraction-free as possible. I routinely fall into the bad habit of distracting myself with random online things when in meetings, so I understand the impulse to multitask. When taking classes and having meetings remotely, there is an inclination to not take your presence at these meetings as seriously as if it were an in-person event. For example, do not attend class while walking back from the Mullins Center having just received a COVID test, or while driving or being driven somewhere.

Asking questions remotely: During remote discussions and or lectures, please refrain from verbally or visually interrupting the presenter (faculty, guest speakers, or fellow students). Instead, use either the “raise hand” feature on Zoom or type your question into the Chat window.

(A)synchronous learning: This course will have both asynchronous materials (videos, readings, and assignments you will do on your own) and synchronous activities (we will meet each week, Tues and Thurs, and there will be lots of break-out sessions with small group discussions.) In addition, we will use classtime to explain present additional materials to you, have group discussions, and explain the expectations for assignments.

To be successful in this course, you must do both the asynchronous and synchronous work. We will take attendance in each class and will evaluate your participation through in-class activities and group discussions. At the end of each class, students will be required to submit materials that can be graded as a part of their participation grade. Late assignments will not be accepted. Participation and class citizenship are worth 10% of your grade. We know that you may occasionally be absent from class due to minor illness, job interviews, or family obligations. You do not need to inform us of your occasional absences. You can miss up to 3 of class days without any impact on your participation grade.

Some students may not be able to attend the scheduled class (because of extremes in time zones, or long-term illness). If this is the case, you must inform Dr. Reich at the beginning of the semester, or as soon as long-term illness is diagnosed. If you are unable to attend and participate in the scheduled classes, to get “credit” for the participation portion of the class, and to ensure that you are continuing to engage with the assigned material, you must do the following:

1. OBTAIN PERMISSION FROM DR. REICH TO PARTICIPATE ASYNCHRONOUSLY.
2. Watch videos posted from the synchronous class discussion.
3. Complete the assigned ‘asynchronous activity’ which will be posted/discussed in the class video. The assignment must be completed by the day/time indicated on Moodle. Late assignments will not be accepted.

Digital devices: Students can be easily distracted from their work by mobile devices. Mobile devices will be distracting to you or others during class. Therefore, the policy for this class is that all mobile devices must be stored and turned off or switched into “do not disturb” mode at the start of class. Small mobile devices may not be used during class time for any reason except if this is your only means of connecting to the class Zoom meeting. You may use laptops and tablets during classtime for class purposes only. In the past, this has been a very hard policy for students (and myself) to follow. To help me, I have set an automatic and repeating “do not disturb” rule on my phone so that it never releases any notifications during my classtimes, even when I forget to turn my phone off before class. I also put my phone somewhere where I cannot easily access it or turn it off

during class so I am not tempted to check it. I encourage you to consider taking actions such as these to avoid feeling the urge to use your phone during class. If you send a direct message via email to the TA Anna Barry with the text "I read the syllabus" by the beginning of the second class, you will receive two points of extra credit on your final grade.

Attendance: Attendance is required. Absences (excused or not) will impact your participation grade. Some accommodations may be made for students who are living in different time zones during remote learning, however, these must be negotiated individually with the instructor before or in the first weeks of the semester.

Respect: Due to the team-based learning setup of this course, there will be times throughout the semester when you will be listening to presentations from other students within your team or from other teams. Act respectfully during these presentations by listening carefully, not interrupting, and waiting until presentations are over to ask questions or excuse yourself for a water or bathroom break.

GRADING SCALE

Grade	Percentage
A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D+	67-69
D	63-66
D-	60-62
F	0-59

COURSE SCHEDULE

This is a tentative course schedule and is subject to change with little or no notice.

Unit 1 Getting started

Unit 2 What is "data"?

Unit 3 Data exploration and visualization

Unit 4 Modeling

Midterm review and exam

Unit 5 Understanding uncertainty

Final projects

COUNCIL ON EDUCATION FOR PUBLIC HEALTH (CEPH) COURSE COMPETENCIES

- Distinguish among the different measurement scales and the implications for selection of statistical methods to be used based on these distinctions.
- 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.
- Describe appropriate methodological alternatives to commonly used statistical methods when assumptions are violated.
- Integrate analysis strategies in biostatistics with principles and issues in epidemiology. literature
- Develop written and oral presentations based on statistical analyses for both public health professionals and educated lay audiences.
- Apply statistical methods to solve problems in the health sciences and carry out theoretical research in statistical methodology.

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](#).