DACSS 601
Data Science Fundamentals
Spring 2024

Instructor: Dr. Rosemary Pang
Email: mrpang@umass.edu
Office Location: Bartlett 263 or Zoom

Classroom Assistant: TBD
Email: TBD
Office Hours: TBD

Course Time and Location:

Tuesday & Thursday 5:30 - 6:45 PM   Multimodal   Machmer W-13 or Zoom

Office Hours:

Regular office hours will be held 2:00 PM - 5:00 PM Tu&Th in Bartlett 263 or on Zoom. Please book in advance through Calendly. Make sure you summarize the question you have. If this time does not work, please send me an email for appointment.

Google Classroom and Piazza

All classroom material will be posted in Google Classroom. Class recordings, slides, tutorials, and assignment descriptions will all be found here. Be sure you are logged into your UMass Google account.

You will be invited to join the Piazza platform for this course. Asking or answering questions on Piazza counts as your class participation. Students are expected to help each other and answer others' questions. If your question is not answered after three working days, please @ the tutor or the instructor on Piazza.

Attendance Policy:

All class sessions will be recorded and be made available to all students. Students who are taking this course synchronously (both in-person or online) are required to attend lectures and participate in discussions either in-person or on Piazza. Students who are taking this course asynchronously are required to watch lecture recordings and posted materials, and also participate in discussions on Piazza.

Course Description:

This 3-credit course serves as an essential introduction to the R programming language, catering especially to students who lack prior programming and statistical experience. R's significance spans across all core courses and a plethora of technical electives within the DACSS program. R serves as

a foundational tool utilized across all core courses and several technical electives. The contemporary landscape demands individuals skilled in versatile data science languages like R, distinguishing them from those proficient in narrower platforms like Excel or specialized statistical packages such as SPSS or Stata.

This course also establishes students with a robust foundation in fundamental skills encompassing general data management, data wrangling, and data visualization – which are crucial for advanced quantitative and data analysis courses. Beyond these skills, the course delves into the realm of data-driven storytelling by demonstrating the importance of technological and data literacy for purposes of analysis, argument, and understanding. It aims to equip students with the prowess to intuitively interpret and convey insights derived from data. This extends to arming students with the ability to construct appropriate tables and figures that fortify their arguments, thus empowering them to wield data as a persuasive tool in communication.

Learning Objectives:

By the end of the semester, you will be able to:

- Acquire skills necessary to conduct data analysis in R, capable of understanding and implementing data science research designs across a variety of settings.

- Use tools such as R, RStudio, Quarto, and GitHub to design and complete basic data science tasks, both individually and in group collaborations.

- Engage with research and adeptly identify both the strengths and weaknesses of increasingly common arguments grounded in empirical evidence.

- Communicate the results or limitations of data-centered research clearly and appropriately through both oral and written formats.

Textbook:

There are no textbooks to buy. Articles, coding manuals, book chapters, and other documentations and videos will be available online though Google Classroom.

There is a very good introductory source for the topics of this course and it is freely available online. I strongly recommend that you review existing documentation prior to and after the weekly meeting each week.
Wickham, H., & Grolemund, G. (2016). *R for data science* (DATA).

And for those of you who do not have much prior experience in R programming, this classic book will definitely help you survive the steep learning curve in the beginning. It provides guidelines to each step from installing R to more advanced analysis.
Long, J. D., & Teetor, P. (2019). *R cookbook* (COOKBOOK).

Feedback and Questions:

Students use Google Form to provide feedback and ask questions about course material every week. The instructor will address these questions in the following week.

Course Structure and Grading:
Final grades will be based on:

- **Participation (10%):**
  We have found that students who actively participate in class discussion get a lot more out of the course and learn new skills at an accelerated rate. Students should participate regularly (for at least 8 weeks). Participation does not need to reflect expertise; rather, students should seek to both ask and answer questions regularly and in equal proportion.

  Participation includes the following:

  - engaging in class discussions, which includes asking or answering questions in synchronous meetings
  - asking or answering questions on Piazza
  - emailing me with non-logistics questions
  - meeting with me or tutors during office hours.

- **Tutorials (20%):**
  There will be an R tutorial in most weeks to help you with mastering the R functions and concepts necessary to complete the upcoming challenges. Tutorials are not themselves graded but each tutorial has a related short answer question on Google Classroom. Completing the tutorials will make it a lot easier for you to complete the challenges.

- **Challenges (70%):**
  There will be a challenge *approximately* every two weeks. These challenges help you practice specific skills, and build your ability to capably and efficiently accomplish data science tasks in R. Feedback is provided to help individual students "stretch" and improve their data management and programming skills regardless of their background.

  Students can choose to resubmit up to three challenges to boost their grades. In week 11 (the week of Apr 16 & 18), students will inform the instructor which previous challenges they want to resubmit. Resubmitted challenges are due on May 10.

- **Grade Scale:**
  A: 94-100; A-: 90-93; B+: 86-89; B: 81-85; B-: 77-80; C+: 74-76; C: 70-73; FAIL: Below 70

Software:

We will primarily use R and RStudio for this course. Both software applications are freely available online. You can find a comprehensive installation guide here. No prior knowledge of the R programming language is assumed.

Students who have difficulties downloading RStudio can use alternative platforms:

- <u>Posit Cloud</u>: 25 hours compute time per month for Free Plan.

- <u>Google Colab</u>: not available to run Quarto for publishing files.

Incomplete and Retake Policy:

You will need to pass DACSS-601 before you can effectively take other DACSS courses. In special circumstances where an incomplete is needed, a student needs to have completed 60% of their work for the term.

For your reference, a copy of the essential sections of the UMass policy regarding Incomplete grades has been provided below. (More information can be found on page 28 of the following document: https://www.umass.edu/registrar/sites/default/files/academicregs.pdf):

"Students who are unable to complete course requirements within the allotted time because of severe medical or personal problems may request a grade of Incomplete from the instructor of the course. Normally, incomplete grades are warranted only if a student is passing the course at the time of the request and if the course requirements can be completed by the end of the following semester. Instructors who turn in a grade of "INC" are required to leave a written record of the following information with the departmental office of the academic department under which the course is offered: (1) the percentage of work completed, (2) the grade earned by the student on the completed work, (3) a description of the work that remains to be completed, (4) a description of the method by which the student is to complete the unfinished work, and (5) the date by which the work is to be completed. In the case of an independent study where the entire grade is determined by one paper or project, the instructor should leave with the department information pertaining to the paper or project, which will complete the course. To avoid subsequent misunderstanding it is recommended that the student also be provided with a copy of this information. Grades of Incomplete will be counted as F's until resolved."

Extensions and Late Submissions:

You have THREE 24-hour extensions that you may use on any challenges. You do not need advanced permission to use this extension, but it is your responsibility to tell me if you wish to apply the extension to the specific assignment. Further extensions will not be granted unless they are properly communicated in advance and supported by me or the related offices of the UMass (such as in a situation of a protracted illness or personal/family emergency that causes you to miss a week or more of coursework). You should contact DACSS student advising, which will coordinate with me to determine possible extensions or accommodation for missed work.

Late submission will result in a penalty of 10% of that challenge grade (or 1% of the final course grade) for every 24 hour the paper is late.

University Policies:

- Academic Honesty 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. 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 (http://www.umass.edu/dean_students/codeofconduct/acadhonesty/).

- Accommodation Statement:
  The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For further information, please visit Disability Services (https://www.umass.edu/disability/).

- Title IX Statement:
  In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: https://www.umass.edu/titleix/resources. You do not need to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day / 7 days a week / 365 days a year at the SASA Hotline 413-545-0800.

Class Schedule and Readings

The schedule is tentative and subject to change. We may adjust the schedule due to time or interest.

Feb 1 Class Introduction
　　Syllabus
　　[COOKBOOK] Ch 1.1-1.10
　　[COOKBOOK] Ch 3.1-3.10

Challenge 1 Posted. Due on Feb 16

Feb 6 & 8 Basic R Commands & Data Structure
　　[DATA] Ch 4
　　[COOKBOOK] Ch 2.1-2.14
　　[DATA] Ch 20 & Ch 21

Feb 13 & 15 Tidyverse & Data Transformation I
　　[DATA] Introduction
　　[DATA] Ch 3
　　Tidyverse

Challenge 2 Posted. Due on Mar 1

Feb 20 & 27 Data Transformation II & Time Data
　　[DATA] Ch 5
　　[DATA] Ch 17
　　lubridate

No class on Feb 22. Follow Monday Schedule

Feb 29 Joins & Strings
　　[DATA] Ch 19
　　[DATA] Ch 14

Challenge 3 Posted. Due on Mar 15

Mar 5 & 7 Function & Iteration
　　[DATA] Ch 25
　　[DATA] Ch 26
　　purrr

Mar 12 & 14 Introduction to Visualization
　　A grammar for graphics Ch 3
　　The R Graph Gallery

Challenge 4 Posted. Due on Mar 29

Mar 19 & 21 No Class: Spring Recess

Mar 26 & 28 Visualizing Time-series and Spatial Data
    <u>Time-series</u>
    <u>Draw maps</u>

Challenge 5 Posted. Due on Apr 12

Apr 2 & 4 Graph Customization
    [DATA] Ch 11
    <u>Dealing with colors</u>

Apr 9 & 11 Compelling Narratives Based on Data
    [Miller2005] Ch 2 & Ch 5

Challenge 6 Posted. Due on Apr 26

Apr 16 & 18 Designing Data Science Projects

Inform the instructor if you want to resubmit at most three previous challenges. Re-submitted challenges are due on May 10

Apr 23 & 25 Topics in Data Science & Data Science Ethics
    <u>Data Science Ethics</u>

Challenge 7 Posted. Due on May 10

Apr 30 & May 2 Data Project Management & Github
    <u>Happy Git</u> Ch 1, Ch 4, & Ch 9 - Ch 12

May 7 & 9 Review/Wrap Up or Additional Office Hours