

Binghamton University  
**MATH 488P/575A/590S: Principles of Data Science**  
Summer 2022 Syllabus

## Instructor Information



**Xingye Qiao**  
[qiao@math.binghamton.edu](mailto:qiao@math.binghamton.edu)

**Office:** WH-134

**Research Interests:** Statistics, Machine Learning, High-dimensional Inference, Big Data Analytics, Precision Medicine

**Office hours:** After class, or by appointments.

## **Communication Policy**

*Preferably all communication will be done on Piazza. Questions of common interest to all can be addressed on Piazza so that both the questions and the answers are visible to all. Questions may be posted anonymously. Private messages can also be sent on Piazza, but only if the question does not concern anybody else. Most messages should be sent either publicly or anonymously. Please follow the [Netiquette](#) rules shown below in the syllabus when you communicate on Piazza.*

## Course Information

### **Description**

Computational data analysis is an essential part of modern statistics and data science. This course provides a practical foundation for students to think with data by participating in the entire data analysis cycle. Students will generate statistical questions and then address them through data acquisition, cleaning, transforming, modeling, and interpretation. This course will introduce students to tools for data management and wrangling that are common in data science and will apply those tools to real-world applications. Students will undertake practical analyses of large, complex, and messy data sets leveraging modern computing tools. 488P-02 has a writing component and will fulfill the Harpur W requirement.

### **Course Delivery Mode**

This course will be an online-synchronous course.

## Course Objectives

- Understand and describe the life cycle of data science and explain how the components of data science interact with each other.
- Recognize main data science tasks and techniques from statistics and machine learning and identify the most suitable techniques in different tasks.
- Master and apply core programming skills in R to conduct data collection, data retrieval, data cleaning, data wrangling, exploratory data analysis, modeling, and data visualization.
- Recognize, quantify, and communicate the uncertainty in the data science process.
- Criticize unethical use of statistical analysis and misinterpretations of data, and examine ethical issues that arise in the data science process; describe the professional ethic code for data scientists.
- Communicate effectively the procedures and results in both written and verbal forms.

## Prerequisite(s)

Prerequisites: Knowledge of elementary statistics, calculus, linear algebra, and experience with programming, or consent of the instructor.

## Topic Outline/Schedule

| Week | Lecture | Content                                                                            | MDSR |
|------|---------|------------------------------------------------------------------------------------|------|
| 1    | 1       | Introduction. Install R, basic commands, R Markdown, working directory, R projects | 1    |
| 1    | 2       | Data Visualization: grammar                                                        | 2    |
| 1    | 3       | Data Visualization: ggplot2                                                        | 3    |
| 2    | 4       | Data Wrangling (dplyr and tidyr): one table                                        | 4    |
| 2    | 5       | Data Wrangling (dplyr and tidyr): two tables                                       | 5    |
| 2    | 6       | Data Wrangling (dplyr and tidyr): tidy data                                        | 6    |
| 2    | 7       | Iteration                                                                          | 7    |
| 3    | 8       | Statistical Foundations                                                            | 9    |
| 3    | 9       | Predictive Modeling                                                                | 10   |
| 3    | 10      | Supervised Learning                                                                | 11   |

|   |    |                       |        |
|---|----|-----------------------|--------|
| 3 | 11 | Midterm               |        |
| 4 | 12 | Unsupervised Learning | 11     |
| 4 | 13 | Database Query        | 15     |
| 4 | 14 | Geospatial Data       | 17, 18 |
| 4 | 15 | Text mining           | 19     |
| 5 | 16 | Ethics                | 8      |
| 5 | 17 | Network Science       | 20     |
| 5 | 18 | Big Data              | 21     |
| 5 | 19 | Final Exam            |        |

## Course Requirements

### Required Text

- **(MDSR)** Benjamin S. Baumer, Daniel T. Kaplan, Nicholas J. Horton (2021). *Modern Data Science with R, 2nd edition*. Chapman and Hall/CRC.
  - The book can be accessed for free online at: [Modern Data Science with R \(mdsr-book.github.io\)](https://mdsr-book.github.io)
  - You can purchase a hard copy from [CRC Press](#) or from [Amazon](#).
- **(DF)** D'Ignazio, C. and Klein, L.F., (2020). *Data Feminism*. MIT Press.
  - Free copy online: <https://data-feminism.mitpress.mit.edu/>
  - Chapter 1 and Chapter 2 are required reading.

### Companion Books

- **(R4DS)** Hadley Wickham and Garrett Golemund (2016). *R for Data Science*. O'Reilly Media, Inc.
  - This book can be accessed for free online at [R for Data Science \(had.co.nz\)](https://r4ds.had.co.nz)
  - You can purchase a hard copy from [O'Reilly](#) or from [Amazon](#).
- **(ISLR)** James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: springer. An excellent introduction to statistical learning. In terms of regression methods, it goes beyond linear models.
  - A free copy is available on the authors' website: [An Introduction to Statistical Learning \(statlearning.com\)](https://statlearning.com)

## **Other Requirements**

- A working computer with a web-camera and microphone.
- A second device (smartphone or tablet) with camera and microphone.
- Internet access to Zoom, brightspace, Piazza, GradeScope, and Google Drive.
- Software: R and RStudio

## **GradeScope and Piazza**

- We will use GradeScope (<https://www.gradescope.com/courses/400015>) to collect homework assignments, grade, and provide feedback.
- We will use Piazza (<https://piazza.com/binghamton/summer2022/math488p575a590s/home>) for all the discussions and Q & A.

## **Credit Hours and Expectations**

*This course is a 4-credit course, which means that in addition to the scheduled meeting times, students are expected to do at least 25 hours of course-related work each week of the 5-week Summer session. This includes work done completing assigned readings, participating in lab sessions, studying for tests and examinations, preparing written assignments, and other courses related tasks.*

## **Assignments and Grading**

The following assignments are required for the course.

- Homework and lab assignments
- Quizzes
- Midterm exam
- Final exam
- Group project
- Classroom and Piazza participation

## **Final Grade Breakdown**

| Assignment Name           | Percent of Total |
|---------------------------|------------------|
| Homework & Lab            | 10%              |
| (2) Quizzes               | 20%              |
| Midterm Exam              | 20%              |
| Final Exam                | 30%              |
| Group Project             | 15%              |
| Participation             | 5%               |
| Extra credits (see below) | up to 2%         |

## Homework Assignments

The homework assignments receive weights proportional to the points assigned to each assignment. Together the homework assignments are 10% of the final grade.

## Group Project

This project is a comprehensive analysis of a moderately large dataset using the tools we have learned throughout the course. The project should be done by a group of no more than 3 students. The instructor will provide a few candidate datasets for the groups to choose from, or they can use their own dataset after seeking approval from the instructor. There are no right or wrong answers for many questions involved in the project. The goal of the project is to try to mimic the analysis of a real data set that students might come across. As a final deliverable of the project, the student will make a blog post about their project which will be accessible from the internet.

## Writing Designation

Section 02 of Math 488P has a Harpur W designation. Successful completion of this section will fulfill the W requirement. Students in 488P-02 will be graded at a higher standard than students in the other sections, with more scrutiny on the writing and presentation aspect of the work. Students in Section 02 should pay extra attention in their writing; otherwise their grades will take a hit. Students who do not need a W course and do not wish to go through this elevated scrutiny should talk to the course instructor to be switched to a different section.

## Other Participation

- 5% will be rewarded to students who actively participate in lectures, answer questions, actively participate in the class or on piazza, including asking and answering questions.

## Grading Scheme

| <b>Grade</b> | <b>Percent</b> |
|--------------|----------------|
| A            | 93% – 100%     |
| A-           | 90% – <93%     |
| B+           | 87% – <90%     |
| B            | 83% – <87%     |
| B-           | 80% – <83%     |
| C+           | 77% – <80%     |
| C            | 73% – <77%     |
| C-           | 70% – <73%     |
| D            | 60% – <70%     |
| F            | 0% – <60%      |

## Accessing Grades

Students can assess their grades using the Brightspace.

## Extra Credit (2 points maximum)

1. Your feedback will make this course better for you and future students. Give 5+ pieces of feedback/corrections to course material.
  - typos/spelling/grammar (there are likely many)
  - rewording awkward/unclear language
  - improve visualizations
  - point out explanations you found effective/could be better

Send the feedback (whenever it becomes available) to the course instructors publicly on Piazza (using the *feedback* tag; you can hide your identity, but the content should be visible to all). We will mark the feedback down and make adjustments to your grade in the end if you contribute 5 pieces or more.

Please keep an eye on the feedback that other people contributed, as sometimes it helps to correct an error or clarify a muddy point that you may have. If a similar feedback has been made, then a duplicate feedback will not be counted for the extra credit purpose.

*Last Updated 05/31/2022*

2. For each class meeting, I will ask for volunteers to turn on their camera. This helps me to get a better sense of the “vibe” of the class. I will give extra credits (up to 0.5 points) to students who consistently volunteer to turn on the camera.

## **Course Policies**

### **Penalties for Late Work and Requests for Extensions**

- Late homework and project submissions are allowed for up to one week. For each day the submission is late, the grade will be multiplied by a factor of 90% (a 10% discount). For example, a submission that is 3 days late will only receive  $0.9^3 = 72.9\%$  of the real grade. *A submission that is late by 1 minute is late by 1 day.*
- Students may request an extension of the work before the due time.

### **Make-up Exam Policy**

Students who cannot take their exams as scheduled because of religious conflict, documented serious illness, or compelling, unexpected circumstances may appeal for a make-up exam. The following are examples that are not considered compelling reasons to grant a make-up exam: lack of preparation, negligence, misinformation, or planned vacations and other events.

### **Absences Due to Religious Holidays**

If you anticipate being absent because of any religious observance, please notify me in writing at least one week in advance. We will work together to accommodate.

### **Attendance & Participation**

*Class attendance and participation points are given to encourage your active class participation and discussion. You will be rewarded with a perfect score as long as you frequently come to class and actively contribute to the class discussion during recitations and lectures.*

Although it is not required, most students send their professor a brief e-mail to explain their absence in advance. Students who repeatedly arrive late to the lecture or recitation will have their Class Participation grade lowered. Please sign the attendance sheet (or e-attendance sheet) when you come to the class. Any false signatures will result in zero participation grades for all parties involved.

Class participation is a very important part of the learning process in this course. You will be evaluated on the *quality* of your contributions and insights. Quality comments possess one or more of the following properties:

- Offers a different and unique, but relevant, perspective;
- Contributes to moving the discussion and analysis forward;
- Builds on other comments;

*Last Updated 05/31/2022*

- Transcends the “I feel” syndrome. That is, it includes some evidence, argumentation, or recognition of inherent trade-offs. In other words, the comment demonstrates some reflective thinking.

We will use our assessment of your participation to manage borderline grades. While your participation grade is subjective, it will not be random or arbitrary. And, clearly, more frequent quality comments are better than less frequent quality comments.

Activities on the online forum are an important part of the participation points.

## **Incomplete**

- *It is possible to enter an incomplete contract with the instructors if you cannot complete the course in time. Consult the instructor.*

## **Academic Integrity**

*Binghamton University provides explicit guidelines in the Student Academic Honesty Code (see the [University Bulletin - Academic Policies and Procedures for All Students](#)). Unless specified otherwise in the syllabus, I expect the work you submit for grading to be yours and yours alone. Not acknowledging another's work with proper references, taking credit for someone else's work, letting your work appear in another student's paper, or fabricating "results" are grounds for failing the assignment and/or the course. If you have any questions about what constitutes plagiarism or cheating, please ask me.*

## **Netiquette**

Unfortunately, there are instances where disrespectful and inappropriate behavior can occur in a course, even in an online setting. It is important to remember to be respectful of your classmates and their ideas. Here are the Netiquette Rules students are expected to follow in this course. Based on “15 Rules of Netiquette for Online Discussion Boards” by Touro’s College Online Education Department.

1. Before posting your question to a discussion board, check if anyone has asked it already and received a reply. Just as you wouldn’t repeat a topic of discussion right after it happened in real life, don’t do that in discussion boards either.
2. Stay on topic – Don’t post irrelevant links, comments, thoughts, or pictures.
3. Don’t type in ALL CAPS! If you do, it will look like you’re screaming.
4. Don’t write anything that sounds angry or sarcastic, even as a joke, because without hearing your tone of voice, your peers might not realize you’re joking.
5. Always remember to say “Please” and “Thank you” when soliciting help from your classmates.
6. Respect the opinions of your classmates. If you feel the need to disagree, do so respectfully and acknowledge the valid points in your classmate’s argument. Acknowledge that others are entitled to have their own perspective on the issue.



*Last Updated 05/31/2022*

7. If you reply to a question from a classmate, make sure your answer is accurate! If you're not 100% sure when the paper is due, DO NOT GUESS! Otherwise, you could really mess things up for your classmates, and they will not appreciate it.
8. If you ask a question and many people respond, summarize all answers and post that summary to benefit your whole class.
9. Be brief. If you write a long dissertation in response to a simple question, it's unlikely that anyone will spend the time to read through it all.
10. Don't badmouth others or call them stupid. You may disagree with their ideas, but don't mock the person.
11. If you refer to something your classmate said earlier in the discussion, quote just a few key lines from their post so that others won't have to go back and figure out which post you're referring to.
12. Before asking a question, check the class FAQs or search the internet to see if the answer is obvious or easy to find.
13. Check the most recent comments before you reply to an older comment since the issue might have already been resolved or opinions may have changed.
14. Be forgiving. If your classmate makes a mistake, don't badger him or her for it. Just let it go – it happens to the best of us.
15. Run a spelling and grammar check before posting anything to the discussion board. It only takes a minute and can make the difference between sounding like a fool and sounding knowledgeable.

## **Campus Help for Students**

### **Disability-Related Equal Access Accommodations Statement**

Students wishing to request academic accommodations to ensure their equitable access and participation in this course should notify the instructor as soon as they are aware of their need for such arrangements. Authorizations from Services for Students with Disabilities (SSD) are generally required. We encourage you to contact SSD at (607) 777-2686 to schedule an appointment with the Director or Learning Disabilities Specialist. The [SSD website](http://www.binghamton.edu/ssd/) ([www.binghamton.edu/ssd/](http://www.binghamton.edu/ssd/)) includes information regarding their Disability Documentation Guidelines. The office is located in UU – 119.

### **University Tutoring Services**

UTS offers free tutoring for undergraduate students at Binghamton University. All UTS tutoring appointments must be scheduled online through the [my.binghamton.edu](http://my.binghamton.edu) portal. Walk-in tutoring is also available for select courses. If you have any questions about UTS, call 607-777-9235, email [uts@binghamton.edu](mailto:uts@binghamton.edu), or visit the website: <http://www.binghamton.edu/tutoring>.

## **ITS Helpdesk/myCourses Support**

Walk-in: Located in the Computer Center's first floor lobby.

Call: 607-777-6420; E-mail: [helpdesk@binghamton.edu](mailto:helpdesk@binghamton.edu).

<https://www.binghamton.edu/its/>

## **Libraries**

The Libraries offer a wide variety and range of services including research assistance, instruction, user-friendly interfaces, digital preservation, digital scanners, and resource sharing.

Text: 607-205-8173; Call: 607-777-2345; Email: [refquest@binghamton.edu](mailto:refquest@binghamton.edu)

<http://www.binghamton.edu/libraries>

## **Dean of Students**

If you are experiencing undue personal or academic stress at any time during the semester or need to talk with someone about a personal problem or situation, I encourage you to seek support as soon as possible. I am available to talk with you about stresses related to your work in my class. Additionally, I can assist you in reaching out to any one of a wide range of campus resources, including:

1. Dean of Students Office: 607-777-2804
2. Decker Student Health Services Center: 607-777-2221
3. University Police: On-campus emergency, 911
4. University Counseling Center: 607-777-2772
5. Interpersonal Violence Prevention: 607-777-3062
6. Harpur Advising: 607-777-6305
7. Office of International Student & Scholar Services: 607-777-2510

## **University Counseling Center**

At some point during their college experience, students may encounter personal, social, or developmental issues that call for assistance beyond the advice provided by friends and family. That's where the University Counseling Center (UCC) can help. The UCC provides a variety of free and confidential counseling services delivered by professional counselors. All currently enrolled Binghamton University undergraduate students, graduate students and affiliated entities are eligible to receive these services free of charge. Services and programs available through the center include individual and group counseling, consultation, referral, and psychoeducational programs. For more information or to make an appointment, visit <https://www.binghamton.edu/counseling>.