Data Science I **Fall 2021**
STAT/CS 287
Time: Tu/Th 4:25–5:40 PM **Place**: Innovation E432

Professor James Bagrow

Office: Innovation E426
Email: james.bagrow@uvm.edu
Office Hours: We/Th 2:00–3:00 PM or by appointment. Held in MS Teams.

Teaching Assistant: Damin Zhu (damin.zhu@uvm.edu)
TA Office Hours: Tu 2:00–3:00 PM or by appointment. Held in MS Teams.

Course website: https://bagrow.com/ds1/
Textbook: *None*

Extracting meaning from data remains one of the most important tasks of science and industry. The Internet and modern computers have given us vast amounts of data, so it is more important than ever to understand how to collect, process, and analyze these data. A picture is worth a thousand words, so visualizations, from scientific plots and infographics to interactive data explorers, are crucial to summarize and communicate new discoveries.

Learning objectives In this course students will learn:

1. data harvesting, storage, and "munging" or cleaning to process data,
2. analyzing data with existing methods such as descriptive statistics and visualizations,
3. developing new, problem-specific measures to explore trends and features in data,
4. automating data science procedures with computer programs, and
5. communicating data-driven results.

Particular emphasis will be placed on nontraditional (non-numeric) data such as networks, text corpora, etc. and on developing good habits for rigorous and reproducible computational science.

Programming This is a programming-intensive course taught using **Python**, and homework and projects will use Python (version 3). Python is very popular in industry and is free, easy to learn, and has many useful third-party packages. To support Windows, Mac, and Linux, please use:

- Anaconda. A free, scientifically-focused "bundle" of Python and important Python libraries. It provides a text editor, enhanced interactive prompt called IPython, and a graphical package manager.

You should download and install the Python 3.8 version of Anaconda. I assume you have a personal computer to work from. If this is not the case, please see me so we can make accommodations.

While you are expected to have prior programming experience (such as UVM's CS021), experience with Python is not necessary. You will be given an opportunity at the start of the course to set up your Python working environment and to familiarize yourself with Python programming.

Homework Regular assignments will be given to assess progress. The first few lectures will have take-home programming assignments. These are mostly to ensure you are ready to tackle the bigger assignments and projects later in the course by showing that you have a working Python environment and you understand the example programs being shown.

Quizzes Approximately 6–7 short (≈15-minute) in-class quizzes will be used to track progress during the semester.

Projects Students will work together in small assigned teams on two midterm projects and one final project. The **mid-term projects** will be instructor-driven. The **final project** will be student-driven and students are free to choose their own topics to investigate. You should consider the final project to be assigned at the start of the semester, with further details forthcoming. All students should come to office hours or make an appointment with me to **discuss their final project topics** before they begin any work on their project, to make sure the project is acceptable for the course. Written report and code are to be handed in for all projects. Code will be graded on clarity and reproducibility, so you are expected to have simple, readable, and well commented code.

Grades 30% for homework, 20% for quizzes, 15% for each of two midterm projects, 20% for final project and presentation.

- **Graduate students and students taking the course for graduate credit** will be held to a higher grading standard on projects and homework, requiring more thorough analyses and very well documented code. Any ***bonus problems*** given as part of assignments are required for students receiving graduate credit. (These problems will be labeled "Bonus/Required".) Further, projects are required to merge data across disparate sources whenever possible, which is typically challenging. Written reports are expected to conform more closely to standards regarding technical and scientific writing.
- Unfortunately, **late assignments** cannot be accepted (see also the remarks below on Blackboard). To accommodate issues that may occur during the course, the *lowest homework grade* (all students) and *lowest quiz grade* (undergraduates) will be *dropped* when computing final grades.
- We reserve the right to **deduct points** for unorganized/illegible assignments or failure to follow submission and/or formatting criteria. Automating data science using computer programs is a crucial learning objective of this course and requires being ***exact*** with respect to filenames, code formats, and so forth.

Important remarks:

- The course website and Blackboard page will be updated often with lecture summaries, homework, and other information. You should check them regularly. Consult Blackboard for all due dates, announcements of quizzes, and so forth. A 24-hour clock will be used, so noon is 12:00 and midnight is 00:00. For any material or assignments collected through Blackboard, be aware of the time of day when it is due as well as the due date.
- I may convey important information to you via your UVM email account. If you do not use your UVM account, please have mail from this account forwarded to an account you check frequently. When emailing me, please include **Data Science I** within the subject line.
- **University Attendance Policy**: Classroom attendance is a necessary part of this course. The lecture notes will form the bulk of materials, so attendance is important. Please refer to the most recent UVM Catalog: *"Students are expected to attend all regularly scheduled classes. The instructor has the final authority to excuse absences."* Lecture attendance is mandatory as is attendance to the **final exam period**.

Since every student is entitled to full participation in class without interruption, ***all students are expected to come to class prepared and on time, and remain for the full class period***. Class starts promptly at 4:25. You should be in your seat and ready to begin class at this time. Arriving late is *disruptive* to others around you and to myself. Class ends at 5:40 PM. Packing up your things early is disruptive to others around you and to myself.

UVM expects students, faculty, and staff to remain compliant with all ***COVID-19 recommendations and measures*** in place for UVM, the State of Vermont, and the City of Burlington. This includes following all rules regarding facial coverings when attending class and generally in indoor spaces. If you do not follow these guidelines, I will ask you to leave the class. If you forget your mask, you cannot enter the class and should go back and retrieve your mask. The Code of Student Conduct outlines policies related to violations of University policies that protect health and safety on campus.

If a student will not be able to attend in-person classes for qualifying health reasons, Student Health Services (SHS) will send a notification to the appropriate student services office or designated staff member informing them of this along with the dates the student is unable to attend. The SHS notification will specify whether the request for flexibility is only around in-person class attendance or includes additional flexibility for assignments and tests because the student is too ill to participate. Students are responsible for working with their faculty to make up class content and work they miss due to a documented illness.

- Personal emergencies and other events, planned or otherwise, can disrupt your work in the course. If such an event occurs, please work with ***CEMS Student Services*** in a prompt manner to minimize the impact of the event on your coursework and keep your learning on track. Note that the instructor makes the final determination on any accommodations and that it is ultimately your responsibility to keep up with the course.

- Spreadsheet files and applications such as Microsoft Excel are not allowed for any purposes in the course, including as a data analysis tool or a data storage format! See also: *Microsoft's Excel Might Be The Most Dangerous Software On The Planet*. We will discuss why in class and we will explore alternatives to spreadsheets for working with tabular data.

- Ensure that you have ***suitable backups*** of all electronic submissions and other class materials. Losing data due to technological mishaps is not a sufficient excuse for missing work, and it is your responsibility to prevent such mishaps from interfering with your coursework. Likewise, unless the university Blackboard service is unreachable, mistakes on Blackboard, such as uploading incorrect or incomplete files, uploading to the wrong assignment, and so forth, are not sufficient excuses for missing work. If submitting material to a Blackboard assignment multiple times, only the latest submission will be considered and all prior submissions will be discarded. Therefore, please ensure each submission is complete and self-contained. ***It is your responsibility to ensure Blackboard is used correctly at all times***.

- **Classroom Disruption Offenses**. Students who disrupt a classroom, laboratory, or other environment in which educational or research activity takes place may be subject to action under the Code of Student Rights and Responsibilities. Disruptive classroom conduct means engaging in behavior that substantially or repeatedly interrupts either the instructor's ability to teach or student learning. As a guide to properly engaging with all members of the university community, consult the values laid out in the ***Our Common Ground*** statement. *It is your responsibility to know and understand the Our*

Common Ground statement and the University Code of Student Rights and Responsibilities.

- Please keep all activities in class and in assignments as "safe for work".
- It is possible for multiple assignments to be assigned at the same time. For example, the next homework may be assigned before the current homework is due.
- The course begins with a review of statistics and of Python programming. I recommend taking extra time early on to familiarize yourself with the provided supplemental review materials. Some assignments in the course may take an *inordinate amount of time and effort* to complete if you are unfamiliar with these supplemental materials.
- Please do not make course materials, assignments, your own work for the course, etc. publicly available without my prior permission.
- Please do not copy someone else's work. It does not help with the learning process and violates the UVM student Code of Academic Integrity. This includes **plagiarism** of online and offline sources. Ignorance of the definition of plagiarism is not an acceptable excuse.

  Offenses against **academic integrity** are any acts which would have the effect of unfairly promoting or enhancing one's academic standing within the entire community of learners. Such acts are serious offenses, which insult the integrity of the entire academic community of the University. The UVM policy on academic integrity is a useful guide. ***It is your responsibility to know and understand the university policy on academic integrity and plagiarism.*** Any suspected violations of the policy will not be tolerated and all allegations will be forwarded to the Center for Student Ethics & Standards.
- In keeping with University policy, any student with a documented disability interested in utilizing accommodations should contact SAS, the office of Disability Services on campus. SAS works with students and faculty in an interactive process to explore reasonable and appropriate accommodations, which are communicated to faculty in an accommodation letter. All students are strongly encouraged to meet with their faculty to discuss the accommodations they plan to use in each course. Contact SAS: A170 Living/Learning Center; 802-656-7753; access@uvm.edu; or https://www.uvm.edu/access.
- **UVM Religious Holidays Policy**: Please *submit in writing by the end of the second full week of classes* your documented religious holiday schedule for the semester. Students who miss work for the purpose of religious observance will be permitted make up work within a mutually agreed-upon time.