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. basic data harvesting and storage with automated computer programs,
2. data “munging” or cleaning to process data,
3. analyzing data with existing methods such as descriptive statistics and visualizations,
4. developing new, problem-specific measures to explore trends and features in data, 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 3.6 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 should have prior programming experience (such as UVM’s CS021), experience with Python is not necessary. Early lectures include reviews of Python programming and how to set up your working environment.

**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 8-10 short (≤ 10-minute) in-class quizzes will be used to track progress during the semester.

Projects  Students will work on two projects during the semester. The mid-term project will be an instructor-driven assignment (think of it as a multi-week homework). The final project will be self-driven and students are free to choose their own topics to investigate. All students should come to office hours or make an appointment with me to discuss their final project topics before they get started, to make sure the project is acceptable for the course. A written report and code are to be handed in. Code will be graded on clarity and reproducibility, so you are expected to have simple, readable, and well commented code.

Grades  35% for homework, 15% for quizzes, 20% for midterm project, 30% for final project and presentation.

- **Graduate students** taking this course will be held to a higher grading standard on projects and homework, requiring more thorough analyses and very well documented code. Projects are required to merge data across disparate sources whenever possible, which is typically challenging. Written reports are expected to more closely conform to standards regarding technical and scientific writing.

- **Late assignments** will be marked off 10% for every calendar day past due.

- I reserve the right to deduct points for unorganized/illegible assignments or failure to follow submission and/or formatting criteria.

Important remarks:

- The course website and Blackboard page will be updated often with lecture summaries, homework, and other information. You should check them regularly. A 24-hour clock will be used, so noon is 12:00 and midnight is 00:00.

- I may convey important information to you via your UVM email account. If you do not use your uvm.edu account, please have mail from this account forwarded to an account you check frequently. When emailing me, please include Data Science 1 in the subject line.

- **University Attendance Policy:** 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 considered mandatory as is attendance to the final exam period.

- 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. Any suspected violations of the policy will not be tolerated and all allegations will be forwarded to the Center for Student Ethics & Standards.

- Please do not share or post online course materials or your submissions without prior permission.

- 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.

- 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. A student’s accommodation letter lists those accommodations that will not be implemented until the student meets with their faculty to create a plan. Contact SAS: A170 Living/Learning Center; 802-656-7753; access@uvm.edu; or https://www.uvm.edu/access. If you need specific accommodations in this class, please bring your accommodation letter as early as possible (at least 2 weeks prior to any quiz or exam) so that we can make appropriate arrangements.

- **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.

- **Extra help**: Do not hesitate to come to my office during office hours or by appointment to discuss a homework problem or any aspect of the course.