

Introduction to Data Science and Visualization (DSV)
MATH/CS 195/295

Time: Tu/Th 8:30–9:45

Spring 2014

Place: Perkins 102

Professor James Bagrow

Office: 212 Farrell Hall, Trinity Campus, 210 Colchester Ave. ([Map](#))

Telephone: (802) 656-2825

Email: james.bagrow@uvm.edu

Office Hours: TBA, or by appointment.

Course website: <http://bagrow.com/dsv/>

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.

Goals In this introductory 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. Python is free, easy to learn, and has many useful third-party packages. To support Windows, Mac, and Linux, we will use a bundled python working environment called **Enthought Canopy**. Canopy is free for academic purposes (you must [request an academic license](#) using your .edu email). It provides a text editor, enhanced interactive prompt called **IPython**, and a graphical package manager.

Canopy has been installed on the classroom computers, but I am assuming 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 CS021), experience with python is not necessary. The first few lectures will review python programming and how to set up your working environment.

Homework The first few lectures will have take-home programming assignments. These are mostly to ensure you are ready to tackle the bigger projects later in the course by showing that you have a working python environment and understand the example programs being shown. Homework assignments will taper off as the semester progresses, being replaced with the larger mid-term and final projects.

Projects Two long-term projects will make up the majority of graded work. Students are free to choose topics they are interested in, and some overlap between the two projects is OK. All students should come to office hours or make an appointment with me to **discuss their 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. Canopy provides a facility to make reports that contain python code; We will discuss this in class. Code will be graded on clarity and reproducibility, so you are expected to have simple, readable, and well commented code.

Grades 20% for homework, 30% for midterm project, 50% 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. Additionally, written reports are expected to more closely conform to standards regarding technical and scientific writing.

Remarks:

- The course website will be updated often with lecture summaries, homework and other information. You should check it quite regularly.
- Since this is a 195/295 pilot course, details may change as the semester progresses. Expect things to break.
- Spreadsheets are not allowed!
- 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. Also, when emailing me, please include **Data science** in the subject line.
- 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. Any suspected violations of the policy will not be tolerated and all allegations will be forwarded to the Center for Student Ethics & Standards.
- UVM, through its ACCESS office, provides accommodation, consultation, collaboration and education support services to **students with disabilities**. To contact the ACCESS office, go to: <http://www.uvm.edu/access/>; email them at access@uvm.edu; or call 656-7753. If you need specific accommodations in this class, please bring a letter from ACCESS as early as possible so that we can make appropriate arrangements (at least 2 weeks before any homework or exam).
- **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 this work.